Learnability of the mora-counting alternation of /g/ nasalization in Japanese compounds

Hailang Jiang[*]

Abstract: Tokyo (Yamanote) Japanese exhibits an unnatural and typologically bizarre alternation of /ɡ/ nasalization in compounds. The probability of each compliant word undergoing nasalization is reported to be significantly conditioned by the mora count of the compound and the nasality of the preceding segment (Breiss et al. 2022). The former factor is unnatural and uncommon because phonology cannot count the exact number of moras in a word if it is more than 2, and patterns counting to more than 2 are typologically rare. I conducted a wug test to examine whether native speakers could productively extend the lexical trend on these two factors to nonce words. The results show that while the factor of nasality is perfectly internalized by native speakers, the factor of mora count is not, exhibiting a 'surfeit of the stimulus' effect which can be explained by a learning bias against counting. Finally, a potential alternative explanation about token frequency is proposed.

Keywords: wug test, experiment, learnability, nasalization, Japanese

1 Introduction

1.1 Background: Japanese /g/ nasalization in general and in compound words

In Tokyo (Yamanote) Japanese, there is an allophonic alternation of /ɡ/: the voiced velar [ɡ] and the nasal [ŋ] surface word-initially and word-medially, respectively,[1] and the process of /ɡ/ becoming [ŋ] is commonly referred to as Voiced Velar Nasalization (VVN) in the literature (Ito & Mester 1996). The nasalization rule is formalized in (1) with examples.

(1) /ɡ/ nasalization in Tokyo (Yamanote) Japanese

/ɡ/ → [ɡ] / $[_{\text{word}}$ __	e.g., /ɡeki/ [ɡeki] 'drama'
/ɡ/ → [ŋ] / $[_{\text{word}}\ X_0$__	e.g., /kaɡami/ [kaŋami] 'mirror'

[*] I would like to express my sincere gratitude to Andrew Nevins for supervising this project and to James White, Canaan Breiss, Andrew Lamont, Michael Becker, anonymous reviewers at AMP2023, and audiences at both AMP2023 and 2023 Linguistic Institute for their valuable comments on this project. Thanks to Teru Konishi, Saki Stait, Chihiro Taguchi, Rina Furusawa and Yasutada Sudo for support with the Japanese stimuli used in the experiment. Thanks to Canaan Breiss for providing the data from Breiss et al. (2022). Any remaining errors would be mine.

[1] Yamanote Japanese is a traditional type of Japanese spoken by Yamanote 'uptown' people in Tokyo. In recent years, change has continued in the direction of replacing the word-internal [ŋ] with [ɡ] (Hibiya 1995), and thus, some Tokyo residents no longer use the allophone [ŋ] in their daily speech. The word-internal [ŋ] allophone is mainly used by elder people and in formal speech. Many speakers, however, still know about this alternation and have some implicit grammatical knowledge of it. I will discuss this and the participants in my test in detail in Section 2, especially in footnote 10.

When we concatenate two members[2] together to form a new compound word, if the second member begins with the phoneme /g/, according to the rule in (1), it should surface as [ŋ]. However, there are some variations in the data (Ito & Mester 1996): there are times when /g/ can only surface as [g], and other times when both [g] and [ŋ] are acceptable, which appears to be lexically-specific. In other words, for compound words where /g/ serves as the initial phoneme in the second member, some words optionally[3] undergo VVN, surfacing as either [g] or [ŋ], and some words never undergo VVN, remaining unchanged as [g]. The examples in (2) illustrate such two types of compound words.

(2) Japanese two-member compound words (from Breiss et al. 2022)

i. that optionally undergo VVN
/doku + ga/ → [doku-ŋa]~[doku-ga] 'poison moth'

ii. that never undergo VVN
/nou + geka/ → [nou-geka] *[nou-ŋeka] 'brain surgery'

The pivotal distinction between the two types is whether a form with the nasal variant [ŋ] is acceptable or not. Breiss et al.'s (2022) corpus study quantitively summarizes this morpho-phonological observation and proposes three factors[4] that can condition the optional application of VVN in compound words in the data: (i.) the mora length of the compound word, (ii.) relative frequency of each member and (iii.) nasality of the preceding segment.

The first factor, the mora length[5] of the entire compound word, is the central concern of this essay. According to their findings, the mora length of a word has a significant effect on a compound word's probability of undergoing VVN. The following Figure 1,[6] adapted from Figure 3 in Breiss et al. (2022, p. 4), depicts the proportion of compliant words undergoing VVN across two-member compound words of different mora lengths. To put it simply, the longer the compound word, the lower the proportion of compound words that can undergo VVN. For example, shorter /ki-ga/ 'hunger' can nasalize, [ki-ŋa], but longer /touʃi-gahou/ 'perspective drawing' cannot, *[touʃi-ŋahou].

[2] Each member is either a single morpheme (e.g., bou-gai ぼう-がい/妨-害 'hindrance', both members are monomorphemic) or a combination of a few morphemes (e.g., himitsu-gaikou ひみつ-がいこう/秘密-外交 'secret diplomacy', both members are bimorphemic). In this essay, hyphens stand for member boundary, rather than morpheme boundary.

[3] Note that in the context of Japanese VVN, the term 'optional' implies that for certain specific words such as the example in (2.i), the voiced velar /g/ has two options of free variants to choose from, namely [g] and [ŋ].

[4] Some anonymous reviewers of AMP2023 suggest that there may be some overlooked factors that can explain the pattern discussed in this study and thus need examining, e.g., foot structure, lexicon stratum (Sino vs. Yamato origin), the number of phonological domains of two-member compound words (1 vs. 2), etc. I acknowledge that these factors may play a role, but they are beyond the scope of this essay. The main focus here is to examine the effects of the factors proposed by Breiss et al. (2022), rather than conducting a new corpus study on Japanese VVN.

[5] In Japanese, a mora-timing language, each mora takes almost an identical amount of time (Warner & Arai 2001).

[6] Both Figure 1 in this essay and Breiss et al. (2022)'s Figure 3 depict the proportion (or average 'probability') of compound words of different mora lengths undergoing VVN. The former only includes compound words whose second member is free (i.e., it can also be used as an independent word), while Breiss et al. (2022)'s Figure 3 includes all words. The trends reflected in the two datasets look similar. The final paragraph in this subsection explains why I would use this dataset with a smaller range.

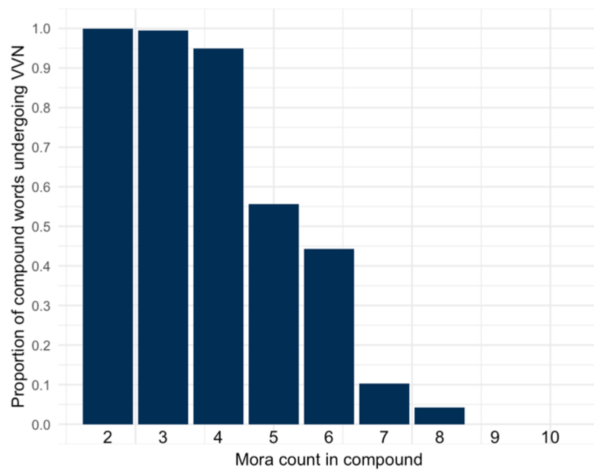

Figure 1. Probability of undergoing VVN across words of different mora lengths
(Excluding words with free members)

Specifically, the figure above shows that at mora lengths 2 through 4, almost all words have the option to undergo VVN; at lengths 5 through 6, nearly half of the words optionally undergo VVN; at lengths 7 and above, almost no words undergo VVN. There are two dramatic drops between 4 and 5, and 6 and 7. This is a theoretically strange phenomenon because speakers should at least have information about the mora length of a given word if they are really aware of the generalization here and thus able to apply VVN based on mora length. The unnaturalness and typological rarity of the effect of this factor is explained in Subsections 1.2 and 1.3, and before them, I will briefly discuss the two other factors that Breiss et al. (2022) report.

In addition to mora length, the second factor is the nasality of the segment preceding of /g/ (being a nasal vs. a vowel). A preceding nasal can give rise to a slightly higher proportion of VVN than a preceding vowel, as shown in Figure 2, which is adapted from Figure 3 in Breiss et al. (2022, p. 4) in the same way as Figure 1.

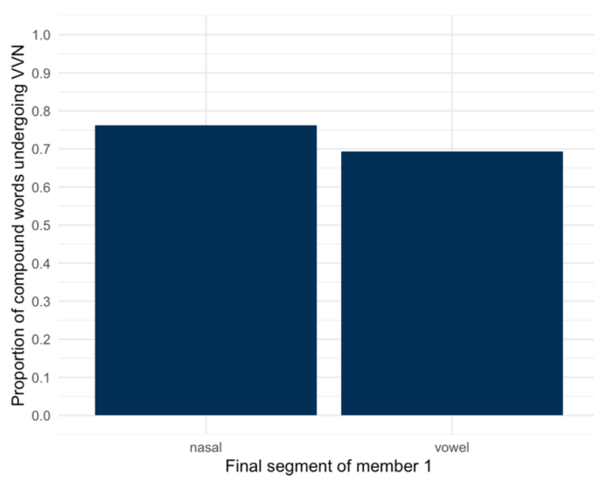

Figure 2. Probability of undergoing VVN across words with a nasal or a vowel before /g/
(Excluding words with free members)

The third factor which can significantly affect the probability of undergoing VVN is the relative frequency of the first member of the compound word, M1, or the second member, M2.

Relative frequency of members (cf. Hay 2003 and Hay & Baayen 2005, which demonstrate the utility of this measure) is defined as the log-frequency of a member (M1 or M2) minus the log-frequency of the compound word. This variable reflects the tendency to maintain identity between different types of members (free vs. bound) and the compound word. It is explained by Breiss et al. (2021) as a means to enforce probabilistic paradigm uniformity (or output-output identity).

When M2 is bound and has zero frequency on its own, it has a relative frequency of negative infinity, and according to the corpus data from Breiss et al. (2022), all the compound words with such a bound M2 never undergo VVN and thus exhibit no variation in the relative frequency of M2, so they are irrelevant to the factor of mora length, and thus excluded from the analysis of this essay. This is why I choose not to present the data with all words in Figure 1 and Figure 2 in this essay and only focus on a subset of it, i.e., the compound words with free members.

The rest of Section 1 discusses the details of the three factors. Section 2 discusses the design of an experiment related to the factors; Section 3 presents the results obtained. Section 4 explores related topics and raises potential explanations of the results. Section 5 concludes.

1.2 Explanations of the three factors

This subsection first discusses the factor of relative frequency for its ground in phonological theories, and then the other two factors using the term 'naturalness.'

Relative frequency can be a part of lexical characteristics that can explain the variability of a process involving reference to surface forms in the paradigm of a word (see Steriade & Stanton 2020, Breiss 2021a, b) and this is a valid phonological ground to account for the relative-frequency-based VVN variations in Japanese compound words; for the detailed account, interested readers may refer to Breiss et al. (2021).

Then, I use the term 'naturalness' for the other two factors, the nasality of the segment preceding of /g/ and the mora length of the compound word, in the sense of Peperkamp, Skoruppa & Dupoux (2006): a rule is natural when it satisfies the standard of (i.) Phonetic Proximity, (ii.) Contextual Relevance and (iii.) Markedness Reduction. The nasality of the preceding segment is a natural factor since it satisfies all three standards; mora length is not a natural one as it greatly violates (ii.) Contextual Relevance which means 'the set of triggers[context] … is relevant to the change in that it tends to be homogeneous with respect to the feature(s) that undergo(es) the change' (p. 1) because the target changed from [g] to [ŋ] becomes neither more similar nor dissimilar to its context in shorter words than in longer words, i.e., the change is equally irrelevant to contexts across words of different mora lengths. For standard (i.) and (iii.), namely whether the change from [g] to [ŋ] results in greater phonetical harmony or less phonological markedness in shorter words than in longer words, I leave these for future research.

Naturalness of a phonological pattern can enhance its learnability, as shown by many experiments (e.g., Pater and Tessier 2003, Wilson 2003, Peperkamp, Skoruppa & Dupoux 2006, Hayes et al. 2009). In the case of unnatural phonological generalizations, very often there is a learning bias against them. For a review of the learnability of unnatural patterns, see Hayes & White (2013), Moreton & Pater (2012).

Additionally, from another perspective, the effect of mora length on whether to nasalize a stop like [g] is unnatural because it involves second-order phonotactics (Walker & Dell 2006, Becker et al. 2011), in that the alternation is dependent on some aspect of the syllable or the word (here, the number of moras in the entire compound word) other than the position it fills. There is evidence suggesting that second-order phonotactics, if not unlearnable at all, are acquired at a slower pace and with less robustness in certain instances (for a review, see Walker & Dell 2006).

To conclude this subsection, the factor of relative frequency and nasality of the preceding segment is well grounded by phonological theories. The factor of mora length, however, is unnatural. Additionally, there are reasons to hypothesize that this unnatural pattern is underlearned in Japanese, but before finally presenting the hypothesis in Subsection 1.4, in the next subsection 1.3, I will further discuss why the factor of mora length is a bizarre predictor of /g/'s nasalization from a typological perspective in terms of counting in phonology. Progressive assimilation spotted in the pattern of Japanese VVN is not typologically rare, but counting the exact number of moras is. In short, very few languages exhibit a pattern that involves counting the phonological units in a domain beyond two as a condition of the rule under any analysis, especially when it conditions segmental features like [nasal] in /g/.

1.3 Counting in phonology

How can the grammar have the information on the exact length of a word? That is to say, how does phonology get access to the exact number of phonological units (here, moras) in a domain?

If we assume that the number of moras is not a part of the lexical information and that the mora length is the factor that is relevant and indeed plays a role in determining a word's probability of undergoing VVN, one possibility is that we can only access it by *counting*. Counting has been an unusual thing in phonology. There are many phonologists claiming in various ways the idea that phonological generalizations (rules or constraints) cannot 'count' to three or more (see Paster 2019 for a full list). For example, McCarthy & Prince (1999) argues that considering locality, a count in grammar can run up to two: a rule may fix on one specified element and examine a structurally adjacent element and no other; Rose (1999, p.397) argues that the structure of binary branching is not a sort of counting in that 'the number 2 is an artefact of the hypothesis that languages select binary versus unary constituents … if arithmetic counting were allowed in representations, we would also expect ternary or quaternary constituents, a possibility which is arguably not required'.

In this study, 'counting'[7] is used in the sense that an exact number (larger than two) of a phonological unit in a domain is stated in the context of a phonological rule. For example, a hypothetical rule where /A/ becomes [B] only when the word that contains /A/ has 5 or more syllables, is a counting rule.

There are some languages in which phonological generalizations are analyzed as counting (to 3 or more), but under re-analyses, they can be perfectly explained by binary structures (i.e., counting within 2) (see Paster 2019 for a review). Paster (2019) also proposes the idea that some phenomena, such as H tone assignment in Kuria, cannot be analyzed with any approach where counting is limited to 2. Whether or not currently we are able to formalize the counting pattern in UG at all, it is certain that the pattern of counting to 3 or more is typologically rare (Paster 2019).

The Japanese VVN pattern is a phonological alternation which at least 'counts' to 7. To the best of my knowledge, no one has reported any alternation which counts to at least 7 in any other language. This renders counting-based VVN in Japanese typologically uncommon.

Moreover, the reported counting phenomena all focus on prosody (i.e., stress or tone), as an example of a counting phenomenon involving segmental features is missing (Paster 2019), thus it is even more unusual that counting can regulate nasality (e.g., [ɡ] vs. [ŋ]) in Japanese VVN.

---

[7] Counting has different senses in different literatures. Another common meaning of counting in phonology exists in the term 'counting effect' of cumulative models such as Harmonic Grammar, which means that the exact number of violations of a constraint can influence the harmonic score of a candidate, e.g., in Breiss (2020).

1.4 Research question and an overview of the experiment

Based on the discussion in Sections 1.2 and 1.3, the factor of mora length affecting the probability of a compound word undergoing VVN is both unnatural and typologically uncommon, while the factor of the nasality of the preceding segment is natural. Thus, I hypothesized the factor of mora length is not internalized into native speakers' grammatical knowledge, i.e., the trend observed in the lexicon cannot be productively extended to nonce words.

To test the hypothesis, I conducted a wug test to examine the productivity of the mora-counting alternation in Japanese. To allow comparison, I also examined the factor of nasality of the preceding segment. The relative frequency of members was not examined and was controlled in the experiment, with each member appearing only once. All stimuli were nonce words[8] in the form of M1-M2, where M1 and M2 were free words[9] and M2 started with the phoneme /g/. Forty-five compound words in total were created by manipulating the mora length of the compound word and the nasality of the segment preceding of /g/.

Each nonce compound word had two potential forms, one with [g] and the other with [ŋ]. Participants were asked to rate the relative naturalness of these two forms.

2 Method

The experiment adopts the experimental design originally introduced by Berko (1958), commonly known as the 'wug-test.' The objective of a wug-test is to assess whether participants of the test have internalized the observed generalizations observed in their native language by having them extend the generalizations to nonce words, which they have never encountered before.

2.1 Participants

I recruited 30 participants aged between 18 and 65 from Prolific, without any history of linguistics learning or linguistic disorder. To be eligible, they had to be native speakers of Tokyo Japanese and self-report that they knew the [ŋ] variant and the [g]/[ŋ] alternation.[10]

To ensure data quality, I incorporated an attention check randomly intermixed with the test trials. Participants who failed the attention check were excluded from the data (n=9). Additionally,

---

[8] For the details of stimuli, see Section 2.2. Note that for monomoraic members like *ga* and *ta*, I cannot require them to be novel, so I only require the compound word containing any of them to be novel. For monomoraic M1, this is tolerable, since I arbitrarily required M1 to be novel, and the status of M1 being free or bound is not reported to affect the probability of undergoing VVN; for monomoraic M2, which must begin with /g/, this is logically unavoidable, if not tolerable, since *ga*, *gi*, *gu*, *ge*, *go* can never be novel.

[9] Given that only compound words with a free M2 exhibit VVN, all the stimuli of M2 in the test were free. For the sake of simplification, M1s were all free as well.

[10] It is difficult to recruit the exact group of Yamanote Japanese speakers online. Because some speakers living in a wider area in Tokyo are also able to produce the [g]/[ŋ] alternation in their speech, I defined participation criteria as being native speakers of Tokyo Japanese. To further screen for eligible participants, they should self-report to be able to discern differences between two audio samples [ga, gi, gu, ge, go] and [ŋa, ŋi, ŋu, ŋe, ŋo] (produced by a Tokyo Japanese native speaker) and naturally employ such distinctions in some contexts in their daily speech. Individuals not meeting these criteria were kindly asked to withdraw from participation. Canaan Breiss raised a potential concern, suggesting that ineligible participants may falsely claim their ability to produce the [g]/[ŋ] alternation in their speech in order to participate and get the reward. I acknowledge that this aspect was not taken into consideration prior to the experiment. However, my screening was likely to be successful for (i.) the results on these factors which are not null (see Section 3), (ii.) the ABX test (see Section 2.3), and (iii.) the final survey where participants were asked what their strategies were and a typical answer was 'intuition/observations of speech by self and others.' Strictly speaking, these three justifications are post hoc. It is indeed necessary to ensure the participants genuinely meet the criteria prior to the experiment, e.g., Kawahara et al. (2021) conducted interviews to screen participants. For any future rerun of this experiment, this must be considered.

I excluded participants who did not pass the final ABX tasks (n=3) (see the penultimate paragraph of Section 2.3). Finally, 18 of them met the qualification criteria.

2.2 Materials

All compound words used as stimuli consisted of two members M1 and M2, and M2 always began with /g/. For simplification, the member boundary was fixed in the middle of the word,[11] since only mora length of the compound word, rather than that of each member, has been reported to be a factor at play.

All five vowel phonemes in Japanese /a, e, o, i, u/ were utilized and all vowels were short in the test. All eleven simple (i.e., no clusters like *Cj* or *Cw*) consonants which could be possibly combined with all of these five vowels as onsets to form CV-shaped moras (i.e., /k, s, z, t, d, n, h, p, b, m, r/, excluding /g/ at issue), were chosen to create stimuli. Additionally, there are two moraic codas in Japanese, /N/ and /Q/,[12] which stand for a nasal coda and an obstruent coda in gemination with the following obstruent. All the M1s and M2s were assembled from the aforementioned moras in random order. To summarize, all generated compound words can fit in the following template: μ ... μ - $gV\mu$ ... μ with 3 restrictions: the mora /gV/ in the middle; no /N/ word-initially; no /Q/ at any member boundary.[13]

All the compound stimuli were created by manipulating mora length and nasality of the segment preceding of /g/. Specifically, as for the factor of mora length, every free member M1 or M2 consisted of 1 to 5 moras, so each compound word consisted of 2 to 10 moras. Each compound length corresponded to 5 items so 5*9 = 45 compound stimuli were created. As for nasality, around half of the stimuli contained an M1 that ended with a vowel while the other half contained an M1 that ended with a moraic nasal coda /N/.

Furthermore, I asked a native speaker of Tokyo Japanese to remove all actual Japanese words and members (if they were not tolerable, according to footnote 8) and regenerated a similar one to the removed one. Finally, there were n = 45 compound words in total, each containing a unique M1 and M2. The list of words used as stimuli can be found in Appendix.

I created 45*4 = 180 audio files. For each of the 45 compound words, 4 recordings were needed, which were separate M1, separate M2, M1-M2 compound with a [g] variant and M1-M2 compound with an [ŋ] variant. All stand-alone M2 as a free word should start with the surface form [g]; the /g/ in compound words may surface as [g] or [ŋ], so recordings of both versions were needed for the test. All stimuli were pronounced by a native speaker of Tokyo Japanese and recorded in the recording room at UCL. All recordings had their peak intensity and pitch equalized using Praat, and then all the processed recordings were inspected by another native speaker of

[11] Specifically, for a compound word with *n* moras, if *n* is even, both member should be 0.5*n moras long (e.g. *taki-gara* 2+2, not *takira-ga* 3+1); if *n* is odd, the first member should be 0.5*(n-1) moras long and the second member should be 0.5*(n+1) moras long (e.g. *ta-gaki* 1+2 not *taki-ga* 2+1).

[12] Moraic /N/ and /Q/ can only stand in coda position, i.e., they cannot be member-initial. /Q/ cannot be word-final because it requires a following consonant. Additionally, for the current purpose, /Q/ cannot be placed at the end of M1, creating a geminate /gg/ which is not discussed here.

[13] In so far as pitch accent, I randomly put accent on any mora for any given M1 or M2 words, i.e., the accent pattern is not constant for members of each mora length. Then the accent of the compound M1-M2 is predictable and can be derived according to the default compound accent rule in Nishimura (2013): '(i.) When the head component is bimoraic or shorter, the accent falls on the final syllable of the first component; (ii.) When the head component is trimoraic or quadrimoraic, the accent falls on the first syllable of the second component; (iii.) When the head component is quinquimoraic or longer, the lexical accent is retained.' Additionally, in the nonce word judgement test, the head of the compound word is randomly assigned to M1 or M2 because the head status of a member is determined by outside non-phonological factors, which is irrelevant here.

Tokyo Japanese and judged to be acceptable Japanese words; moreover, each compound audio pair, one with [ɡ] and the other with [ŋ], was judged to be identical in every perceivable aspect except the [ɡ]/[ŋ] allophony.

2.3 Procedure

The experiment was carried out on the Gorilla Experiment Platform, where participants were expected to spend around 15 to 20 minutes to finish the experiment.

Before the experiment, participants were required to sign a consent form, which explained the general procedure of the experiment. After that, participants were taken to a page to test their devices. Participants heard a Japanese sentence and they were asked to type the sentence into the textbox. This was to make sure the audio worked properly.

They were also told to write down 'yes' when they were asked what day it was today during the test, as a method to examine the participants' attention to the instruction.

Then they were told that sometimes *ga*, *gi*, *gu*, *ge*, *go* sounds like *ŋa*, *ŋi*, *ŋu*, *ŋe*, *ŋo* and this may be lexicon-specific. These two sequences of sounds, pronounced by the same pronouncer of stimuli, were played to participants. Participants were encouraged to try to pronounce these two variants of these sounds, and were required to quit the test if they were unable to distinguish them or if they did not use both variants in their daily speech. Following that, three real words with /ɡ/, as well as two possible forms (one with [ɡ] and the other with [ŋ]) in audio forms, were given as examples.

Then they were all directed to the instruction page where they were told the procedure of the experiment with a concrete example. At the bottom of this page, there was a scale with the recording of the word with [ɡ] on one end and the recording of the word with [ŋ] on the other end. The placement of each was in random order. Participants were told that, in each page of a trial, the task was to rate the relative naturalness using the 7-value scale between two forms. They were also told the meaning of 'relative naturalness' here: for a scale with numbers from 1 to 7, '1'/'7' indicated that only the left/right word was acceptable as a natural form; '4' meant both forms were equally natural, and '2, 3, 5, 6' were intermediate options.

Then they were directed to the test pages. Each page corresponded to a unique compound word, so each participant encountered 45 pages in total. The stimuli were presented in a self-paced forced-choice task. For a quick preview of all the elements in the task, see Figure 3.

One of the most important things in this task was to indicate the status of M1 and M2 as free words. I presented on the upper left part of each page three sentences which all contained M1 as a free word, written in Japanese orthography:[14] 'This is X.'; 'That was my X.'; 'The X I saw yesterday is good.' *X* was to be replaced by the word M1. It was bold and underlined to draw participants' attention. The three sentences in different actual use provided an assurance of the wordhood of M1. The recording of the word M1 was played automatically one second after the participant was directed to this page. One second later, the same for M2 was presented in the upper right part of the page.

After the recording of M2 was played, the identical three sentences in Japanese with the M1/M2 replaced by their concatenated compound word M1-M2 appeared at the lower part of the page. No audio was provided. After 2 seconds, two recordings of compound words, one with [ɡ] and the other with [ŋ], were played sequentially in random order with a short pause of 500ms in between. While the first/second recording was being played, there appeared a temporary sign of a

[14] Japanese orthography, i.e., *hiragana* here, does not distinguish between the allophonic alternations between [ɡ] and [ŋ].

loudspeaker at the left/right part of the blank area, and disappeared when the audio was finished. The two loudspeakers which had disappeared appeared again in the same place, by clicking which participants could hear the word again for unlimited times, and 7 buttons between the two loudspeaker signs appeared for clicking, with the numbers from 1 to 7 on them. After clicking any number button, there appeared a ‘next’ button, which led to the next page. An example page with all elements demonstrated simultaneously is shown in Figure 3.

<table>
<tr><td>This is temi.
That was my temi.
The temi I saw yesterday is good.</td><td>This is gemo.
That was my gemo.
The gemo I saw yesterday is good.</td></tr>
<tr><td colspan="2" align="center">This is temigemo.
That was my temigemo.
The temigemo I saw yesterday is good.

[temigemo] 1 2 3 4 5 6 7 [temiŋemo]
[Next]</td></tr>
</table>

Figure 3. An example page of the stimulus compound *temi-gemo* in the test
(English translation, originally written in Japanese orthography)

The attention check might appear randomly at any stage of the test, occupying a single page. There was a line saying ‘Please write down what day it is today’ in English, but according to the previous instruction, participants should write down the word ‘yes.’

Then, there were ten ABX tasks. Participants were asked whether a word in the audio form, chosen from the word list of stimuli, contained the sound of a non-nasal [ɡV] or a nasal [ŋV]. The audio form in the question could only be played once manually by clicking, but there were recordings provided at the bottom of the page to remind participants of what are (non-)nasals, in the form of a sequence of either [ɡa, ɡi, ɡu, ɡe, ɡo] or [ŋa, ŋi, ŋu, ŋe, ŋo] with labels ‘non-nasals’ or ‘nasals’, respectively. These two reminder audios could be played multiple times. The ABX tasks were to make sure participants had reliably perceived the difference between the allophones [ɡ] and [ŋ]. Participants with an accuracy rate below 80% were excluded from the analysis.

Finally, a survey was made on the demographics and language background of participants and how they made a decision on the relative naturalness in the test.

3 Results

3.1 Transformation from the 7-scale bar to a binary variable

In the quantitative analysis by Breiss et al. (2022), whether a compound word can undergo VVN is a binary variable: each word is described to have only two choices, yes or no.[15] This is because the corpus of that study only provides attested forms (attested or unattested), rather than a fine-grained scaled preference for either form.

To align with it and to allow potential comparison between the corpus results and the experimental results in this study, I transformed the 7-value variable of participants’ responses on the relative naturalness of each compound pair into a binary variable. Whenever the nasalized form

[15] Again, ‘yes’ means that a word can ‘optionally’ undergo VVN, i.e., both [ɡ] and [ŋ] are attested; ‘no’ means the nasalized form is never attested. See footnote 3 for the meaning of ‘option’ here.

with [ŋ] can be accepted to *any* degree, the new binary variable 'nasalized response' takes the value of 1. In other words, for a given word on a test page, if [g] was on the left and [ŋ] on the right, the original response of '1' corresponds to '0' for the new binary variable; original '2, 3, 4, 5, 6 or 7' all correspond to binary '1' because any of these numbers means nasalized form is acceptable. Figure 4 below shows the merger of values to form a new variable where '|' marks the boundary. If [g] was on the right and thus the rightmost option '7' meant unacceptable nasalized forms, the scale should be reversed first and then processed as above.

Original 7-value variable 'relative naturalness': [temigemo]	1	|	2	3	4	5	6	7	[temiŋemo]
New binary variable 'nasalized response':	0	|		1					

Figure 4. A transformation from the 7-scale bar into a binary variable

This is an attempt for the responses in the test to align with the meanings of variables in the quantitative study by Breiss et al. (2022). In doing so I lost plenty of information on the speakers' judgements and mental grammar, but this is necessary to allow comparison. Comparison between the trend in the corpus lexicon and the results in the experiment data is important because the main concern of this study is to find out whether an observed pattern is extended to novel words or not.

Table 1 demonstrates the original response of 'relative naturalness.' For those cases where [g] was on the right and, the scale is reversed and thus, for all the responses, '1' means unacceptable nasalized forms and '2' to '7' means acceptable nasalized forms (to different degrees). There were 810 responses in total; each of the 18 participants had 45 words.

Table 1. The distribution of numbers of each response of the original variable 'relative naturalness'

1	2	3	4	5	6	7	2 to 7	sum
Only [g]	[ŋ] is acceptable. The higher score, the more acceptable.							
244	108	114	291	23	19	11	566	810

Thus, for the new binary variable 'nasalized response', 244 samples take the value of '0' and 566 take the value of '1', comprising 69.9% of the total.

There are two reasons why I provided a 7-option scale to participants at first, rather than two straightforward options 'can undergo nasalization' vs. 'cannot undergo nasalization.' Firstly, this was to impartially present two forms of the compound forms without implying that '[g] is the underlying form'. If two options were about whether a compound word should nasalize, they implied the underlying status of the non-nasalized form with [g], which might distort participants' judgements on the nasalized form with [ŋ]. Secondly, this was also to prevent the risk of categorizing less probable [ŋ] as entirely impossible when the scale is too coarse. If there were only 3 options on the scale: (i.) 'only [g] is acceptable'; (ii.) '[g] and [ŋ] are equally acceptable'; (iii.) 'only [ŋ] is acceptable', then, participants who perceived [ŋ] as acceptable to a very low extent, say, only 10%, might choose the option (i.) 'only [g] is acceptable,' which was not what they intended and was not faithful to participants' mental grammar.

3.2 Quantitative results

The following Figure 5 illustrates the proportion of nasalized responses classified by the factor of nasality of the preceding segment of [g] in the experiment. Figure 6 is copied from Figure 2, showing the trend in real lexicon conditioned by the same factor.

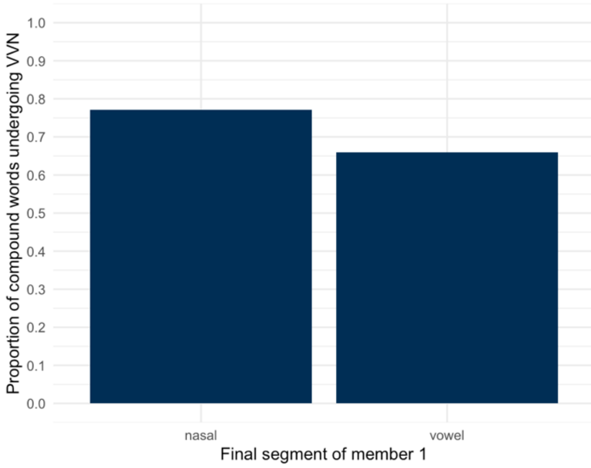

Figure 5. Trend in nonce words

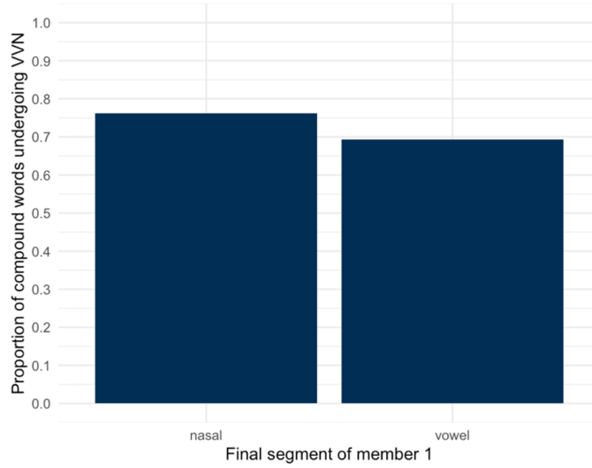

Figure 6. Trend in real lexicon (copied from Figure 2)

The two figures indicate a high degree of overlap in trends. In both figures, the proportion of nasalized responses following nasals is slightly over 75%, and it is consistently slightly higher than that of vowels. This result can indicate that the participants have perfectly internalized the effect of the nasality of the preceding segment.

As for the factor of mora length, Figure 7 shows the proportion of nasalized responses in the test classified by mora length from 2 to 10 moras. In contrast, Figure 8, copied from Figure 1, shows the trend in the real lexicon.

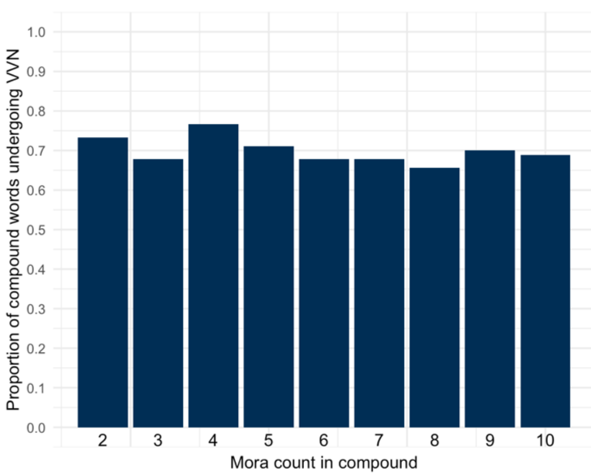

Figure 7. Trend in nonce words

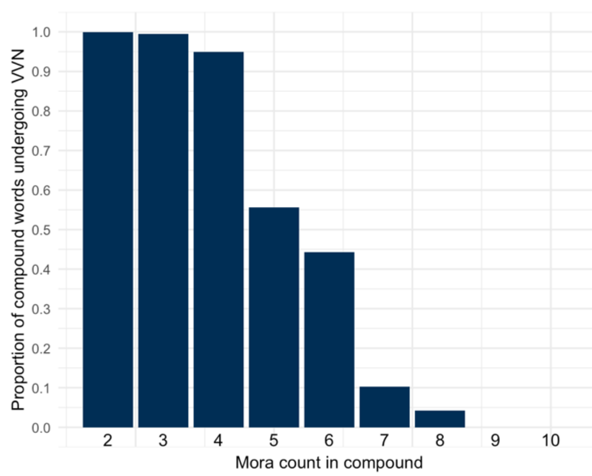

Figure 8. Trend in real lexicon (copied from Figure 1)

In Figure 8, there is evidently a sigmoid curve, while in Figure 7, the proportion for each mora length is consistently around 70%. Therefore, it can be tentatively concluded that mora length does not play a role in the experimental results.

Although mora length appears to be non-influential, the proportion of nasalized responses remains consistently around 70% for any given mora length. This closely aligns with the observed proportion in the real lexicon, which is 70.5%. Therefore, the Law of Frequency Matching (Hayes et al., 2009) is observed here. It is implied that while participants might not have internalized the factor of mora length, they had internalized the proportion of nasalized words within the overall vocabulary. They uniformly applied this proportion to words of varying lengths.

3.3 Statistical analysis of the experimental results

To confirm the observation above through statistical analysis, I ran a mixed-effect logistic model with maximal random effects in R, following Barr et al. (2013). I started by a full model with both factors (mora length and nasality of the preceding segment) and their interaction. The two factors play a role in the lexicon in the quantitative study (Breiss et al. 2022); their interaction was not reported by the quantitative study but incorporated here for analysis of the experimental data. Then I compared the initial model with simpler ones, using backward stepwise comparison implemented in R with *anova (model_1, model_2, test="Chisq")*.

In the initial full model, the dependent variable was the binary[16] 'nasalized response,' and the independent variables were the nasality of the preceding segment, mora length and their interaction. Each participant had maximal random effects for all slopes and an intercept, and each word item has a random intercept. The results are presented in Table 2, along with the R code that I used.

Table 2. Fixed effects of the initial model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.3265	1.0449	2.227	0.026 *
nasality	1.4725	1.4260	1.033	0.302
length	-0.1400	0.1034	-1.354	0.176
nasality:length	-0.0362	0.1916	-0.189	0.850

The R code I used: glmer(Nasalized_Response ~ nasality + length + (1 + nasality + length | participant) + (1 | word), data = data, family = binomial)

No factor was significant in the initial model. The interaction had the largest p-value (Pr(>|z|) = 0.850), so I excluded it first in model 2. After this exclusion, the new model was not significantly different from the initial one (Pr(>Chisq) = 0.7977784), so the excluded factor was unnecessary to be contained.

In model 2 without the interaction, the effect of nasality was weakly significant (Pr(>|z| = 0.0121*) while the effect of mora length was not (Pr(>|z|) = 0.1173). Therefore, I excluded the factor of mora length during the second step, to get model 3.

Exclusion of mora length in Model 3 exhibited a significant change (Pr(>Chisq)=0.000), showing that the factor of mora length indeed plays some role. After the exclusion, the AIC increased from 608.7 to 622.0; the BIC, however, decrease from 655.7 to 650.2. Note that the p-value in the regression model Model 2 is not significant (Pr(>|z|) = 0.1173) but relative low. We can tentatively conclude that mora length appears to play a very subtle role, and there is no evidence to suggest that it should be excluded from Model 2. Therefore, backward stepwise comparison should terminate at this point, and the optimal model obtained is Model 2. The fixed effects are reported in the table below, along with the R code.

[16] In this subsection, the response is a binary variable (nasalized or not) to allow the comparison between the lexicon and the experimental results, according to 4.1. To better model the mental grammar of the speakers shown in the wug test, the original 7-value responses should be made use of with an ordinal logistic model. For current purposes, this modeling is not discussed here.

Table 3. Fixed effects of Model 2

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.36468	1.05536	2.241	0.0250 *
nasality	1.28077	0.51019	2.510	0.0121 *
length	-0.14906	0.09516	-1.566	0.1173

The R code I used: glmer(Nasalized_Response ~ nasality + length + (1 + nasality + length | participant) + (1 | word), data = data, family = binomial)

In summary, nasality of the preceding segment plays a significant role in the experimental results, while mora length, although not statistically significant, should not be ignored. It is worth noting that even though mora length can predict experimental results to a limited extent, this trend diverges significantly from the trend observed in the corpus (Figures 7 and 8). The next subsection will provide a more detailed comparison between the experimental data and lexicon trend.

3.4 The mismatch between lexicon and experimental results

This subsection focuses on the factor of mora length.

Model 2 treated mora length as a numeric variable. While the term 'length' does indicate that it should be numerical, the mora length was a result of counting in my test. The nature of counting essentially involves listing in an ordered manner from small to large. In other words, mora length is not interval- or ratio-scaled but merely ordinal. Increases in mora length at each level are not necessarily homogeneous in terms of the process of counting (i.e., in the experiment, the computational burden of counting from 2 to 3 may not necessarily be the same as counting from 7 to 8, although they both seem to be a 1-unit increment in *length*). Thus, I converted the factor of mora length to an ordinal variable in the analysis. A more appropriate name for this variable is 'mora count'; however, for consistency, I used the name 'mora length' in the model.

To compare each level of length with the level to its left, rather than a fixed level (i.e., compare length = 5 to length = 4, length = 6 to length = 5, instead of comparing every length to length = 2), I coded the variable of mora length using *contrasts(data$length) <- contr.sdif(nlevels(data$length))* in R. Then I ran the code of Model 2 again. The finer-grained results are demonstrated in Table 4 below.

Table 4. Fixed effects of the model with ordinal mora length

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.4858	0.8964	1.658	0.0974 .
nasality	1.4734	0.6091	2.419	0.0156 *
length:3-2	-0.7175	0.8345	-0.860	0.3899
length:4-3	0.6497	0.9228	0.704	0.4814
length:5-4	-0.9267	0.9268	-1.000	0.3174
length:6-5	-0.7836	0.7789	-1.006	0.3143
length:7-6	1.4846	1.3030	1.139	0.2545
length:8-7	-1.4866	1.1651	-1.276	0.2020
length:9-8	0.4998	0.7167	0.697	0.4855
length:10-9	-0.0550	0.6548	-0.084	0.9331

The R code I used: glmer(Nasalized_Response ~ nasality + length + (1 + nasality + length | participant) + (1 | word), data = data, family = binomial)

In this model, the difference between each two adjacent values of mora length was considered separately. It can be observed that none of the factors was significant, and all p-values were greater than 0.20. Comparing this model with ordinal mora length and a model without this factor using *anova()*, there is no significant difference (Pr(>Chisq) = 0.869; AIC: 694.1 to 622.0; BIC: 1004.1 to 650.2), indicating the simpler model[17] can better fit the experimental data. In this model, nasality cannot be further excluded (Pr(>Chisq) = 0.000). The final model was reported in Table 5 below.

Table 5. Fixed effects of the final model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.4210	0.7110	1.999	0.0457 *
nasality	1.1354	0.4575	2.482	0.0131 *

The R code I used: glmer(Nasalized_Response ~ nasality + (1 + nasality | participant) + (1 | word), data = data, family = binomial)

Additionally, I treated the experimental data as real words from the real lexicon in Japanese and used the corpus-based quantitative model proposed by Breiss et al. (2022)[18] to predict the probability of each nonce word undergoing VVN as if they were Japanese lexicon, using *predict()*. I calculated the mean value for every item of each mora length across participants, for example, all nonce words with 4 moras had a lexicon-probability of undergoing VVN of 0.969, and all 8-mora nonce words had a rate of 0.108. Then, for nonce word items, the correlation coefficient between the lexicon probability (predicted by the corpus-based model) and proportion of undergoing VVN in the experiment was only 0.161 (see Figure 9). If we average the lexicon-probability using both factors (nasality and mora length), the correlation coefficient will increase to 0.217.

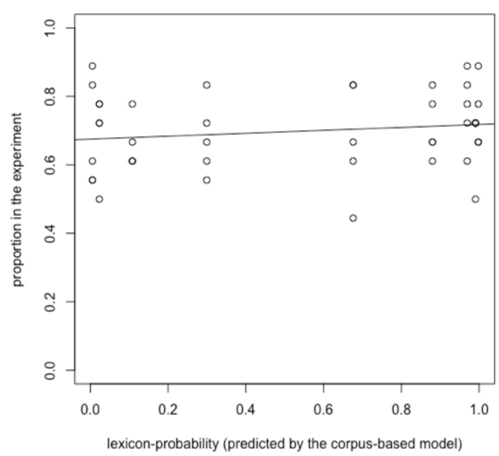

Figure 9. The correlation between the lexicon-probability of words undergoing VVN and proportion of nasalized responses, for all nonce words; each plot represents a unique nonce word

[17] In Subsection 3.3, the interaction term was excluded when mora length served as a numerical factor. When mora length is an ordinal factor here, the interaction term can be excluded as well (Pr(>Chisq) ≈ 1; AIC decreases from 903.4 to 694.1; BIC decreases from 1786.5 to 1004.1).

[18] I ran the *glm()* logistic model containing all the factors proposed by Breiss et al. (2022). Relative frequencies of both members of the compound words were always equal to 0 because all novel members and nonce words appeared only once. I did not run the *brm()* model as in Breiss et al. (2022) due to hardware limitations.

In summary, even though Subsection 3.3 indicates a subtle overall linear effect of mora length on the experimental results, this effect, upon closer scrutiny, is by no means the trend observed in the lexicon. The factor of mora length was not internalized by native speakers during the wug test.

4 Discussion

4.1 A learning bias against counting

The results above are consistent with my hypothesis that the factor of the mora length, though a significant predictor in the lexicon, cannot be productively extended to nonce words. To account for it, I argue that there exists a learning bias against counting-involved alternations. This argument is in line with some prior findings, such as Warker & Dell (2006) who discovers that second-order phonotactics are learned more slowly and with greater difficulty, and Hayes et al. (2009) who argues that unnatural patterns are disfavored and tend to be underlearned.

Since the pattern of mora length is not productive at all, what is the cause of the mora-counting pattern? A typical answer to such questions is that this non-productive pattern does not require a synchronic explanation because it is merely an accidental generalization. Becker et al. (2011, 2012) term this phenomenon 'the surfeit of the stimulus.' In their research on Turkish and English, they also observed instances where certain patterns were not internalized by native speakers as part of their grammar. The cause of the Japanese VVN pattern may potentially be a now inactive diachronic process, but I leave this question open for future research.

4.2 The number of nasals in a compound[19]

If we posit that there are synchronic reasons behind such a pattern, the number of nasals in compounds might be a cause. The more moras a compound word contains, the more nasals it may contain, and thus, there may be a higher likelihood of nasal disharmony targeting the phoneme /ɡ/. Mora-length-based VVN in Japanese can potentially make the surface form of /ɡ/ dissimilar to its context (i.e., a context with only a few nasals or plenty of nasals). I examined the relationship between the number of nasals in a compound word and mora length of it using the data of Breiss et al. (2022): the correlation coefficient between them was only 0.031. Additionally, in the corpus data, the predictor of the number of nasals did not have a significant effect ($Pr(>|z|)=0.666$) on the dependent variable, the probability of a compound word undergoing VVN, in a model[20] with this factor and all three factors in Breiss et al. (2022). Therefore, it is not a likely cause.

4.3 The effect of token frequency

Another possible explanation of the cause of the non-internalized pattern on mora length is the effect of token frequency. The longer a word is, the less frequently it appears (Zipf 2013). Therefore, the words with many moras in them are less frequent. If a word has a low token frequency, there will be only a limited number of examples available to observe the rule governing that word.

Extra token frequency is necessary for learning 'exceptions' in such cases (Endress & Hauser 2011). They argue 'for the exceptions to be learned, they have to occur sufficiently often so that participants can memorize them; this can be achieved by a high token frequency.' (p. 77) In the Japanese VVN case, I hypothesize that for compound words with /ɡ/ where the second member is a free word, [ŋ] is an exceptional variant because paradigm uniformity requires this /ɡ/ to surface

[19] Thanks to Andrew Lamont for suggesting this possibility.

[20] The code I ran in R: model <- glm(undergoVVN ~ nasality + length + relativefreq_member1 + relativefreq_member2 + number_of_nasal_in_compound, data = corpus_data, family = binomial).

as [g] as in other stand-alone second members which are free words (Breiss et al. 2021). For long words, to apply the VVN rule (i.e., to have the exceptional variant [ŋ]), a high token frequency is required. However, long words tend to have a low token frequency, and thus the exceptional variant [ŋ] is hard to be learned. This hypothesis needs verification through another experiment which includes the effect of token frequency.

A good explanation here should account for not only the alternating /ɡ/ allophony in compound words, as in (2), but also the regular /ɡ/ allophony in general context, as in (1). The explanation above in this subsection is applicable to both situations. For compound words, [ŋ] is an exception to [g] due to the correspondence between compound words with /ɡ/ and /ɡ/-initial M2 as free words. The exceptional variant /ɡ/ is harder to be learned in longer, less frequent words. For regular non-compound words, however, [ŋ] is not an exception to anything, because in word-internal positions, /ɡ/ should always be realized as [ŋ], and [g] is never observed. Without identifying an exception, the above frequency-based hypothesis is not valid.

To conclude this subsection, it is possible that mora length can indirectly affect the learnability of the Japanese VVN pattern through the channel of token frequency. The finding of the current study is that, if we only consider the direct effect of mora length on learnability, this is not likely. The following figure shows the indirect and direct effects of the factors of mora count and nasality.

mora count	→ token frequency → learnability	(not covered, probable)
	×→ learnability	unlikely
nasality	→ learnability	likely

Figure 10. The indirect and direct effects

4.4 The implication of experimental results for counting in phonology

In the preceding subsection, it is demonstrated that mora length does not directly contribute to the learnability of the Japanese VVN pattern. The non-internalized factor of mora length provides no evidence that native speakers are indeed counting moras in the case of Japanese VVN; otherwise, counting should be reflected in the experimental results. This aligns with the view of many phonologists (refer to Subsection 1.3) who argue that phonology does not allow a structure counting to 3 or more. In a case where counting could be an effective clue to condition the [g/ŋ] alternation in the lexicon, native speakers tend not to take advantage of it because they are unable to – phonology cannot count.[21] A bias against counting neutralizes all counting-based clues in the lexicon, filtering lexical statistics.

Even if Paster (2019) identifies some patterns that can only be analyzed to be counting to 3 or 4, the Japanese VVN case, which 'seems to be counting' to at least 7 in the lexicon, is not really counting. Paster (2019) supports a phonological theory that allows counting, and this essay demonstrates that even if counting can exist in the theory, there should be some restriction on the number of units phonology can count to – perhaps phonology can never count to a number as large as 7. More evidence is needed to develop a more detailed theory on phonological counting.

[21] Even if in a future study, token frequency turns out to play a role and can explain the variation in the Japanese VVN pattern to a certain extent, my argument still stands. This is because the experimental results demonstrate that the participants were insensitive to the factor of mora length; they did not utilize the valid clues related to counting at all.

4.5 A follow-up study of artificial language learning

In another study of mine (Jiang 2023), I extended the nonce word test from Japanese to an artificial language, with prior training in a structure akin to Japanese VVN in compound words on English-speaking participants, to examine whether a nasalization rule counting to any number of syllables (from 2 to 7, depending on participant groups) is underlearned. The conclusion is that there is no evidence supporting that counting-involved pattern can be learned to any extent, no matter what length (number of units) is stated in the context of a rule.

5 Conclusion

The evidence from a wug test shows that Japanese speakers internalized the natural factor of nasality of the preceding segment conditioning the tendency of /ɡ/ undergoing nasalization in compound words; however, they failed to directly internalize the unnatural factor of mora count, although it significantly conditions the trend in the lexicon. I argue that a learning bias against counting in phonology can account for the non-internalized pattern.

While the cause of the mora-counting pattern in the lexicon remains open for future research (whether it is diachronic or synchronic and what the mechanism is), this essay shows that the participants' judgements do not directly match existing patterns in the lexicon; corpus-based statistical models lacking information on learning biases cannot directly model the judgements on nonce words in an experiment, which reflect speakers' mental grammar.

References:

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 255-278.

Becker, M., Ketrez, N., & Nevins, A. (2011). The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language*, 84-125.

Becker, M., Nevins, A., & Levine, J. (2012). Asymmetries in generalizing alternations to and from initial syllables. *Language*, 231-268.

Berko, J. (1958). The child's learning of English morphology. *Word, 14*(2-3), 150-177.

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

Breiss, C. (2021a). Base effects are probabilistic: a case study in lexical conservatism. In *Talk at the 95th Annual Meeting of the LSA*.

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

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

Breiss, C., Katsuda, H., & Kawahara, S. (2022). A quantitative study of voiced velar nasalization in Japanese. *University of Pennsylvania Working Papers in Linguistics, 28*(1), 4.

Endress, A. D., & Hauser, M. D. (2011). The influence of type and token frequency on the acquisition of affixation patterns: Implications for language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 77–95.

Hay, J. (2004). *Causes and consequences of word structure*. Routledge.

Hay, J. B., & Baayen, R. H. (2005). Shifting paradigms: Gradient structure in morphology. *Trends in cognitive sciences*, 9(7), 342-348.

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

Hayes, B., & White, J. (2013). Phonological naturalness and phonotactic learning. *Linguistic inquiry, 44*(1), 45-75.

Hibiya, J. (1995). The velar nasal in Tokyo Japanese: A case of diffusion from above. *Language Variation and Change, 7*(2), 139-152.

Itô, J., & Mester, A. (1996). *Correspondence and compositionality: The ga-gyō variation in Japanese phonology*.

Jiang, H. (2023). Learnability of length-referencing alternations. *M.A. dissertation at University College London*. [Available here.]

Kawahara, S., Breiss, C., & Katsuda, H. (2021). Token frequency in the grammar: a wug-test of Japanese voiced velar nasalization. *Talk at AMP 2021*.

McCarthy, J. J., & Prince, A. (1999). Prosodic morphology 1986. *Phonological theory: the essential readings*, 238-288.

Moreton, E., & Pater, J. (2012). Structure and substance in artificial - phonology learning, part II: Substance. *Language and linguistics compass*, 6(11), 702-718.

Nishimura, K. (2013). Morphophonology in Japanese compounding. *Tokyo: University of Tokyo dissertation*.

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

Pater, J., & Tessier, A. M. (2003). Phonotactic knowledge and the acquisition of alternations. In *Proceedings of the 15th International Congress on Phonetic Sciences* (Vol. 1180, pp. 1177-1180).

Peperkamp, S., Skoruppa, K., & Dupoux, E. (2006). *The role of phonetic naturalness in phonological rule acquisition*.

Rose, Y. (1999) 'A structural account of Root node deletion in loanword phonology', *Canadian Journal of Linguistics*, 44, pp. 359–404.

Steriade, D., & Stanton, J. (2020). Productive pseudo-cyclicity and its significance. *Talk at LabPhon 17*. [Available here.]

Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(2), 387.

Warner, N., & Arai, T. (2001). Japanese mora-timing: A review. *Phonetica*, 58(1-2), 1-25.

Wilson, C. (2003, March). Experimental investigation of phonological naturalness. In *Proceedings of the 22nd west coast conference on formal linguistics* (Vol. 22, pp. 533-546). Somerville, MA: Cascadilla Press.

Zipf, G. K. (2013). *The psycho-biology of language: An introduction to dynamic philology*. Routledge.

Appendix: a list of stimuli (two-member compounds) used in the test

Note that in the test, all Japanese words were written in *kana* without the member boundary hyphen '-'. The underline indicates the head of the compound; the apostrophe indicates that the mora preceding it bears accent.

3 practice words:

housyasei-genso	ほうしゃせいげんそ(放射性-元素)
himitsu-gaikou	ひみつがいこう(秘密-外交)
bou-gai	ぼうがい(妨-害)

45 nonce words (5 words for each mora length; around half of them have a vowel-final M1 and the other half have a nasal-final M1):

2-mora words (2=1+1):

ta'-ga'	た-が
da'-gu'	だ-ぐ
to'-go'	と-ご
ze'-ge'	ぜ-げ
do'-gi'	ど-ぎ

3-mora words (3=1+2):

pe'-gi'de	ぺ-ぎで
ro'-go'ru	ろ-ごる
ma'-gi'ke	ま-ぎけ
hi'-guha'	ひ-ぐは
shi'-gibi'	し-ぎび

4-mora words (4=2+2):

te'mi-ge'mo	てみ-げも
ru'N-gi'pa	るん-ぎぱ
suta'-ga'hi	すた-がひ
mapu'-gapu'	まぺ-がぷ
kyaN-gosu	きゃん-ごす

5-mora words (5=2+3):

pu'ka-ga'seku	ぷか-がせく
bome'-giza'me	ぼめ-ぎざめ
dotsu'-guko'se	どつ-ぐこせ
noN-gamehi	のん-がめひ
tsuN-gozaku	つん-ござく

6-mora words (6=3+3):

shi'piN-gepiza'	しぴん-げぴざ

<u>zapu’N</u>-giNmu’ ざぷん-ぎんむ
<u>hada’pe</u>-gobo’ki はだぺ-ごぼき
ririN’-<u>gopo’N</u> りりん-ごぽん
nobaN’-<u>gi’tsuzo</u> のばん-ぎつぞ

7-mora words (7=3+4):
<u>me’deN</u>-ganichiN’ めでん-がにちん
<u>se’mehu</u>-go’huroho せめふ-ごふろほ
zote’ko-<u>gibe’neru</u> ぞてこ-ぎべねる
tariku’-<u>gimohu’pa</u> たりく-ぎもふぱ
pabupa’-<u>goNtaku’</u> ぱぶぱ-ごんたく

8-mora words (8=4+4):
<u>ji’nachiN</u>-genikare’ じなちん-げにかれ
<u>pehe’kiN</u>-goro’doki ぺへきん-ごろどき
<u>kasaka’so</u>-gosoki’shi かさかそ-ごそきし
renuchiN’-<u>ga’Nnobe</u> れぬちん-がんのべ
zosoNha’-<u>gipiba’mi</u> ぞそんは-ぎぴばみ

9-mora words (9=4+5):
<u>pe’koQta</u>-gaQtoho’mi ぺこった-がっとほみ
<u>rope’jiba</u>-go’kinachima ろぺじば-ごきなちま
soka’naN-<u>guN’kosene</u> そかなん-ぐんこせね
suzuka’N-<u>gubuNhehu’</u> すずかん-ぐぶんへふ
hazonabe’-<u>gusezo’moma</u> はぞなべ-ぐせぞもま

10-mora words (10=5+5):
<u>ho’benuraN</u>-gimanuhaN’ ほべぬらん-ぎまぬはん
<u>shiru’Ntomo</u>-go’nimerena しるんとも-ごにめれな
<u>hedohi’tara</u>-gabako’Qse へどひたら-がばこっせ
kutatoN’to-<u>gohepiri’nu</u> くたとんと-ごへぴりぬ
redobunibe’-<u>guba’hikere</u> れどぶにべ-ぐばひけれ