Frequency-conditioned variation constrained by the grammar in Japanese nasalization

Canaan Breiss[1], Hironori Katsuda[2], Shigeto Kawahara[3]
[1]University of Southern California, [2]University of Toronto – Scarborough, [3]Keio University

Summary: We present an experiment examining the influence of token frequency and length on voiced velar nasalization (hence "nasalization") in compounds in conservative dialects of Japanese [1]. We find that, controlling for morphological structure, the length of the compound itself is irrelevant, counter to the corpus. However, the frequency effect found in the corpus study of [2] and in a production task reported in [3] holds in novel compounds. We also observed an effect of length on the second noun (N2).

Japanese nasalization: In many conservative dialects of Japanese, [g] and [ŋ] stand in an allophonic relationship, with [ŋ] occurring in prosodic-word-medial position, and [g] occurring elsewhere (e.g. [gama] "toad" vs. [kaŋami] "mirror"). The relevance of this pattern to theoretical phonology was made clear by [1] and explored in quantitative detail in a corpus study by [2]. The basic pattern is shown in Table 1. (1) shows that nasalization is not merely a static generalization about surface allophonic patterns, but also drives alternations. [1] propose that the optionality in (1) arises from the conflict between a phonotactic requiring nasalization, and output-output correspondence between the free [g]-initial form of N2 and its status in the compound. A consequence is that when N2 is bound, as in (2), output-output correspondence is irrelevant because there is no free [g]-initial form of N2, and thus word-level phonotactic markedness is obeyed, resulting in obligatory nasalization.

[2]'s corpus study: The corpus study documented several predictors of the optional application of nasalization in compounds with free N2s, including the higher frequency of N2 accompanying less nasalization (Figure 1, right), and the length of the compound in mora as a whole, with shorter compounds undergoing nasalization more than longer ones (Figure 1, left). While the frequency effects have straightforward functional explanations in terms of lexical access ([3]), the pattern of length-dependence is typologically extremely unusual ([4]), appearing to contradict the axiom "phonology doesn't count" [5]. The authors' intuition suggests polymorphemic N2s undergo nasalization at a much lower rate, likely because they possess internal morphological structure and are typically longer, thus potentially explaining the apparent length-dependence in the corpus.

The current experiment: To test whether speakers of Japanese internalize the effects of frequency and length on nasalization, we carried out a production task involving novel compounds. We parametrically varied the length of N1 and N2 (one or two mora each), as well as frequency of N2. We recruited 12 speakers of conservative dialects of the Tōhoku region, and presented each speaker with the 168 stimuli via PowerPoint, recording whether they produced the compound with [ŋ] or [g]. Results are plotted in Figure 2. Using a Bayesian mixed effects regression model, we found that compounds with higher-frequency N2s are credibly less likely to undergo nasalization relative to those with low-frequency N2s and the length of N2 affects nasalization. However, the lengths of N1 do not independently matter.

Discussion: Our finding of frequency-dependent behavior in novel compounds poses a challenge to existing phonological models [6, 7] These models assume that speakers maintain multiple representations (e.g., the compound and N2 individually) at varying degrees in their lexicon, and thus have difficulty accounting for the frequency-conditioned behavior of morphemes *when placed in novel contexts*. We also speculate that the N2 length effect could stem from (i) prosodic word boundaries inhibiting nasalization, and (ii) monomoraic N2s being more prone to prosodic integration with the preceding N1 than bimoraic N2s.

(1) In compounds when N2 begins with [g] and is a free morpheme, nasalization targeting that [g] is optional: N2=/ga/ “moth”, [doku-ga]~[doku-ŋa] “poison moth”

(2) When N2 is a bound morpheme, nasalization is obligatory: N2=/ga/ “fang”, [doku-ŋa], *[doku-ga] “poison fang”

Table 1: Examples of Japanese nasalization.

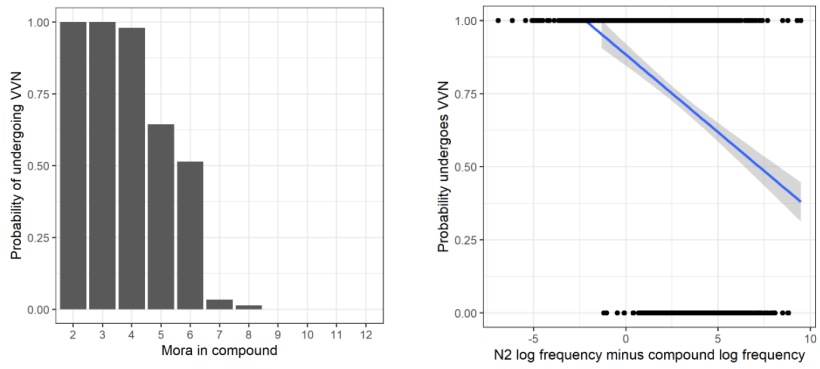

Figure 1: Left: left side of Figure 3 from [2], showing the length-dependence of nasalization (on the vertical axis abbreviated as VVN, Voiced Velar Nasalization). Right: left side of Figure 2 from [2], showing the N2 frequency effect on nasalization.

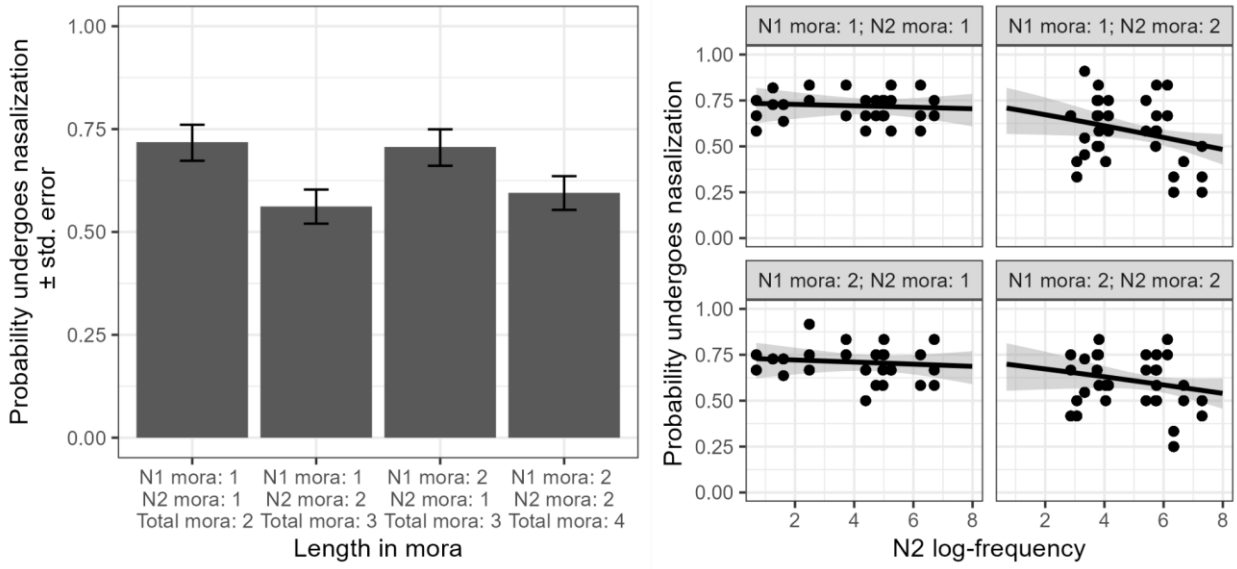

Figure 2: Right: mean proportion undergoing nasalization by total mora in compound. Left: effect of N2 frequency by mora in N1 and N2.

References: [2] 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. [1] Ito, J., & Mester, A. (1996). Correspondence and compositionality: The ga-gyō variation in Japanese phonology. [4] Jiang, H. (2023). Learnability of the mora-counting alternation of /g/ nasalization in Japanese compounds. Talk presented at AMP 2023. [3] Kawahara, S., Breiss, C., & Katsuda, H. (2021). Token frequency in the grammar: A *wug*-test of Voiced Velar Nasalization in Japanese. Talk presented at AMP 2021. [5] McCarthy, J. J., & Prince, A. (1999). Prosodic morphology 1986. *Phonological theory: the essential readings*, 238-288. [7] Moore-Cantwell, C., & Pater, J. (2016). Gradient exceptionality in Maximum Entropy Grammar with lexically specific constraints. Catalan Journal of Linguistics, 15, 53-66. [6] Zuraw, K. (2000). Patterned exceptions in phonology. UCLA Dissertation.