Department of Linguistics

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

Hailang Jiang
hailang.jiang.22@ucl.ac.uk
hailangjiang.github.io
University College London

AMP 2023

Background

Japanese /g/ nasalization (VVN)

- In Tokyo[1] Japanese,

/g/ → [g] word-initially /geki/ [geki] ‘drama’

→ [ŋ] word-internally /kagami/ [kaŋami] ‘mirror’

(Ito & Mester 1996)

[1] Traditional, and by Yamanote ‘uptown’ people; now mainly used by elder people. Change is continuing in the direction of replacing word-internal [ŋ] with [g] (Hibiya 1995).

Japanese /g/ nasalization in compounds

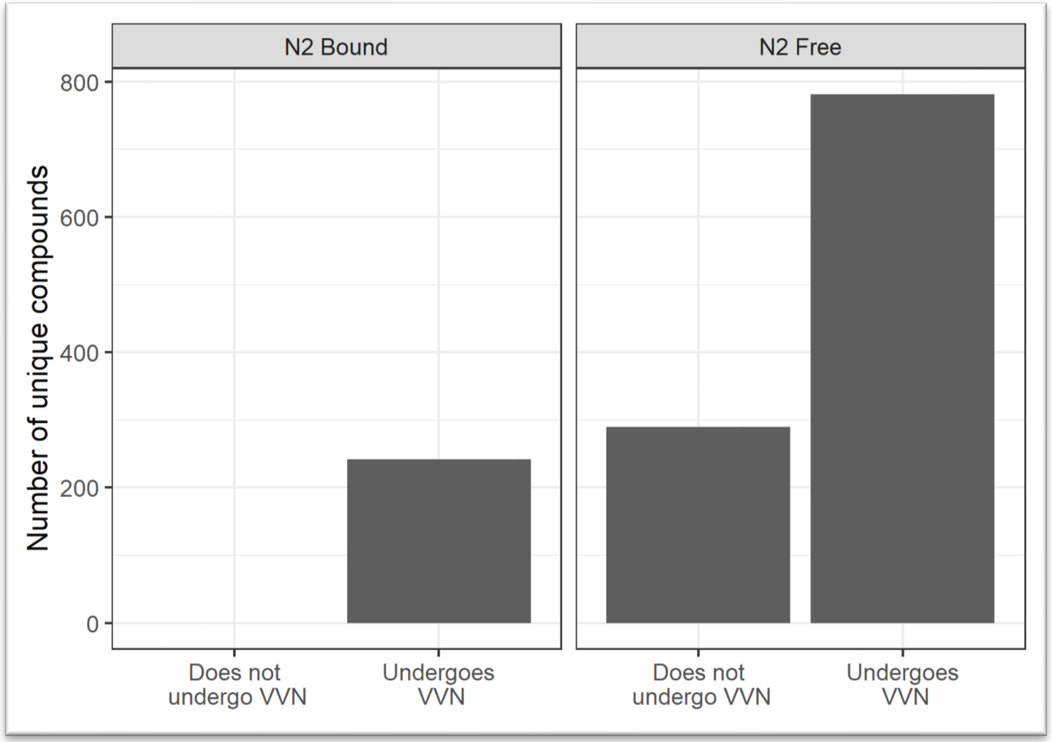
- For two-member compounds, not all /X-gY/ surfaces with /ŋ/.

/doku + ga/ → [doku-ŋa]~[doku-ga]
'poison moth' free variation [g~ŋ]

/noo + geka/ → [noo-geka] *[noo-ŋeka]
'brain surgery' one legal form [g]

(Ito & Mester 1996; Breiss et al. 2022)

- [ŋ] is always acceptable for words with a bound second member.
- This study only considers free members.



from Breiss et al. (2022)

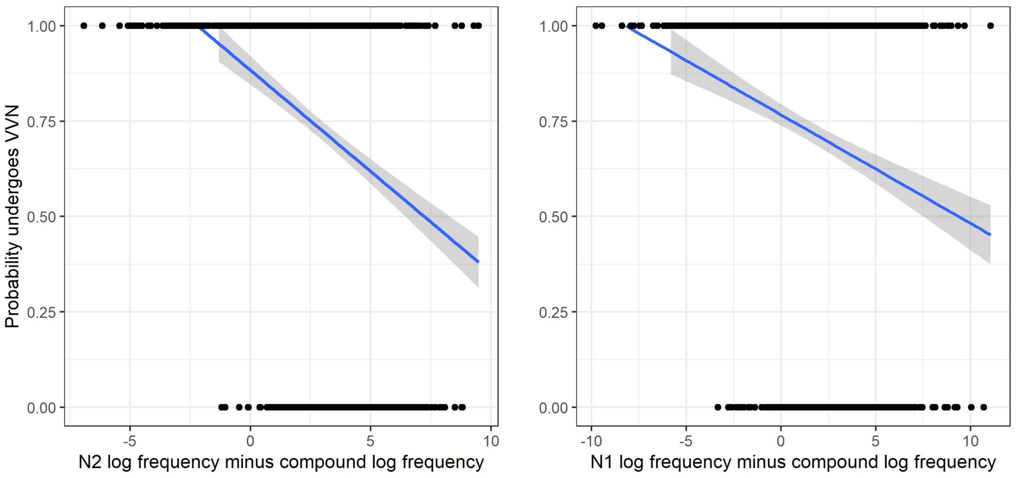
A corpus study on /g/ nasalization

- In a corpus study by Breiss et al. (2022), three factors that *significantly* affect whether such a /g/ undergoes nasalization:

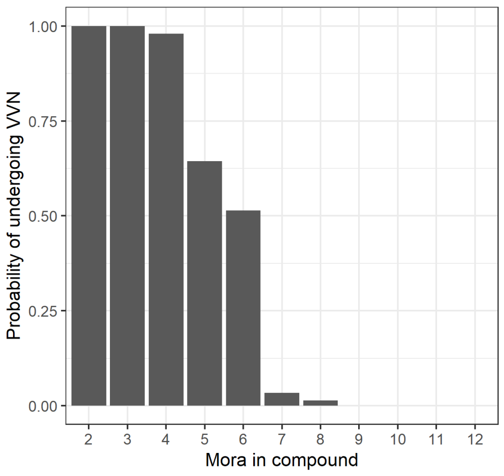
i. Relative frequency of both members	sin-ŋakki (新-学期)
log(freq(member))-log(freq(compound))	*RF(/gakki/)=-0.712*
	but *bijyutsu-ŋakkou (美術-学校)
	RF(/gakkou/)=5.734
ii. Nasality of the preceding segment	gin-ŋa but *noo-ŋeka
iii. Mora length of the entire compound	ki-ŋa but *toushi-ŋahou

A corpus study on /g/ nasalization (cont.)

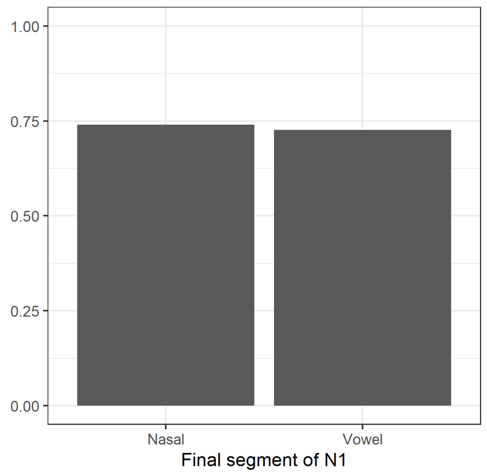
- Breiss et al. (2022) reports three factors quantitatively:



i. Relative frequency of both members (member 2 & member 1)



iii. Mora length of the entire compound



ii. Nasality of the preceding segment

all figures from Breiss et al. (2022)

The (un)naturalness of the three factors

i. Relative frequency of both members

→ Paradigm uniformity (Breiss et al. 2021)

ii. Nasality of the preceding segment

→ Progressive local assimilation in [nasal]

iii. Mora length of the entire compound

→ ?

(un)naturalness in the sense of Peperkamp, Skoruppa & Dupoux (2006)

second-order phonotactics (Warker & Dell 2006)

The mystery of the ‘counting pattern’

- Here, ‘counting’ is used in the sense that a exact number of a phonological units is stated in the context of a phonological rule.

  e.g., a hypothetical rule: [g] → [ŋ] / $|X_0 \_ Y_0| > 5\mu$

- Binary structure in phonology: phonology does not count to a number larger than 2 (e.g., McCarthy & Prince 1999)
- This may stem from difficulties in accessing precise information regarding the number of specific phonological units

- Counting units: syllables, moras … but never segments ☺

Counting in phonology

- Paster (2019):

  · Phonological generalization counting to more than 2 is ***almost*** unattested

  · There are some patterns that can only be ***analyzed*** as counting to more than 2 (e.g. grammatical tone assignment in Kuria)

  · Even so, no pattern counts past ***4***; no similar patterns to Japanese /g/ nasalization (involving counting to 5, 6, 7…) is attested

  · Counting patterns never involve segmental features (i.e., they only condition stress & tone but never [nasal])

Counting in phonology (cont.)
An example of ‘counting’ pattern (Paster 2019)

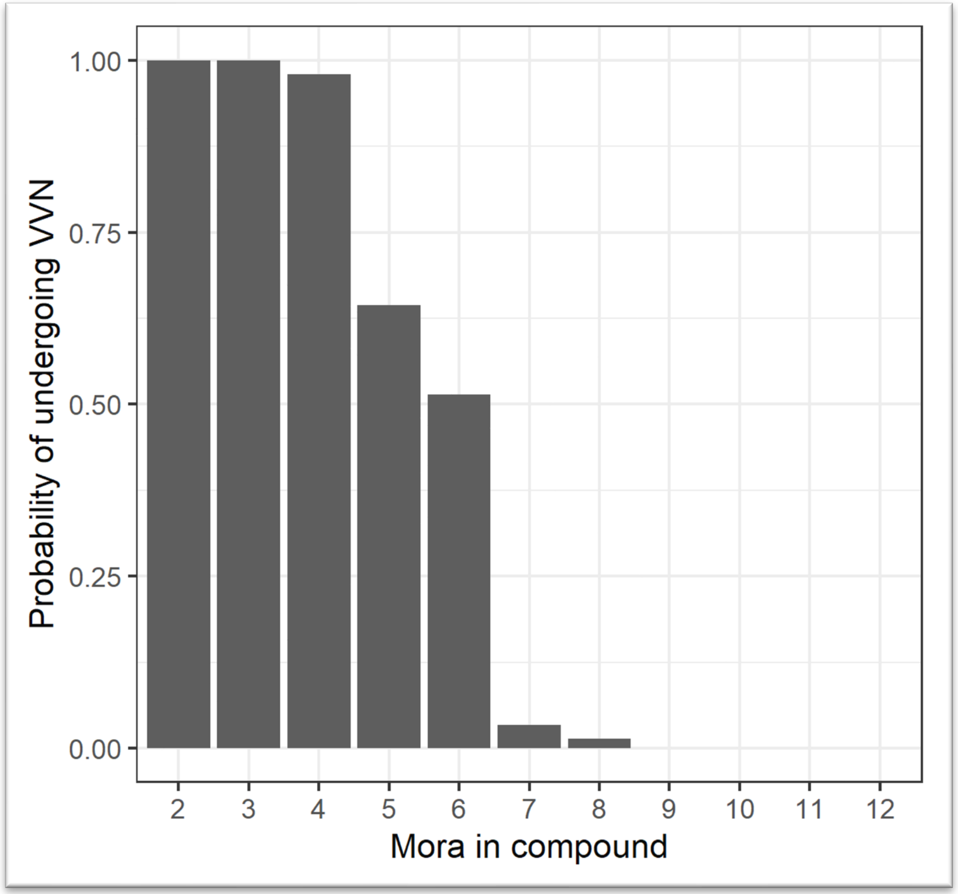
- Kuria (a Bantu language): grammatical tone assignment
- H is assigned to 1st, 2nd, 3rd, 4th mora of the verb stem to mark different tense/aspect/mood (TAM) categories

- Paster (2019) argues that this ‘counting’ pattern cannot be reduced to binary structures.
- This is a pattern that ‘counts’ to 4.

a.	μ1	n-to-o-hɔ̱́ɔ́tɔ́ɔ́t-ɛ́r-a	Past
		FOC-1PL-PAST-reassure-APPL-FV	
		‘we reassured’	
b.	μ2	n-tɔ-ɔka-hoó̱tóót-ééy-e	Past progressive
		FOC-1PL-PAST.HAB-reassure-APPL.PERF-FV	
		‘we have just been reassuring’	
c.	μ3	n-to-re-hɔɔtɔ̱́ɔ́t-ɛ́r-a	Remote future
		FOC-1PL-FUT-reassure-APPL-FV	
		‘we will reassure’	
d.	μ4	to-ra-hɔɔtɔɔ̱́t-ɛ́r-a	Inceptive
		1PL-INCEPT-reassure-APPL-FV	
		‘we are about to reassure’	

Research question

- Given counting is (almost) unattested in conditioning nasality, which is an unnatural pattern,
- Is it a part of grammar of those Japanese speakers? (i.e., can it be productively extended to novel words?)



from Breiss et al. (2022)

Experiment method

A wug test
design

- Wug test (Berko 1958): nonce words, to examine if speakers can productively extend attested patterns to non-existent words
- Choose a more natural form between [X-gY] and [X-ŋY]

- Participants: 30 *Tokyo* Japanese speakers from Prolific, aged between 18-65.
  - Self-report that they know[2] the [ŋ] variant
  - 18 of them were eligible (passed the attention check, the ABX test)

[2] I had them listen to the audios of [ga gi gu ge go] and [ŋa ŋi ŋu ŋe ŋo] and make sure they knew both variants and were able to distinguish them. I also conducted an ABX test to ensure this.

A wug test (cont.)
stimuli

- 45 trials, 4 forms per trial
- Two separate members [X] & [gY] and two potential compounds [XgY] [XŋY]

  e.g., [temi] [gemo] [temigemo] [temiŋemo]
- All in Japanese orthography (hiragana), which does not distinguish [g] and [ŋ]
- Created by manipulating two factors: nasality of preceding seg & mora length (2-10)

e.g.,

preceding seg \ mora length	5 moras	8 moras
V (a, i, u, e, o)	dotsu'-guko'se	kasaka'so-gosoki'shi
N	no'N-ga'mehi	pehe'kiN-goro'doki

- Frequency ≡ 1, thus relative frequency was controlled
- All moras: CV, /N/ or /Q/ (/Q/ cannot end a word due to phonotactics)

A wug test (cont.)
procedure

- Consent form / audio check / instruction → 3 practice trials with real word → test

This is temi.
That was my temi.
The temi I saw yesterday is good.
[temi]

This is gemo.
That was my gemo.
The gemo I saw yesterday is good.
[gemo]

This is temig(ŋ)emo.
That was my temig(ŋ)emo.
The temig(ŋ)emo I saw yesterday is good.
[no audio played here]

Rate the relative naturalness between two potential forms of each nonce compound word in audios.

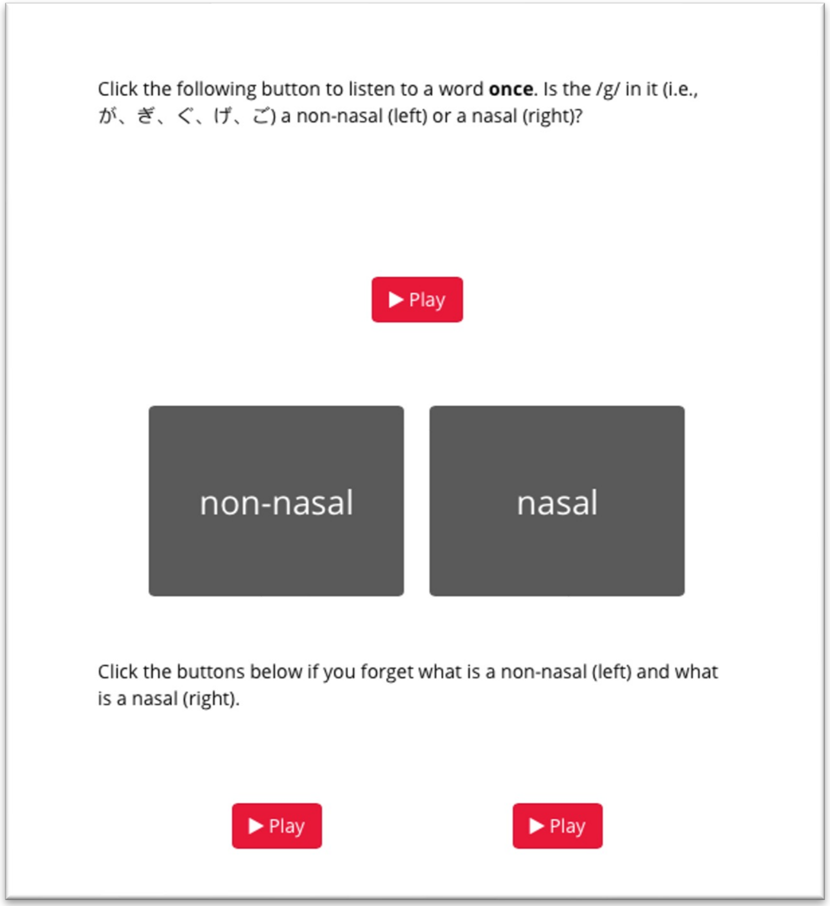
[temigemo] 1 2 3 4 5 6 7 [temiŋemo]

[Next]

- An attention check mixed in a random order with trials

A wug test (cont.)
An ABX test of distinguishing [g] & [ŋ]

- 10 trials
- Audio selected from compounds in the test trials
- Could only be played **once**
- 'Is the /g/ (i.e., が、ぎ、ぐ、げ、ご) in the audio a non-nasal or a nasal?'
- Accuracy rate < 8/10 : excluded



Results

Results

Turning the 7-scale bar to a binary variable

[temigemo] 1 ¦ 2 3 4 5 6 7 [temiŋemo]

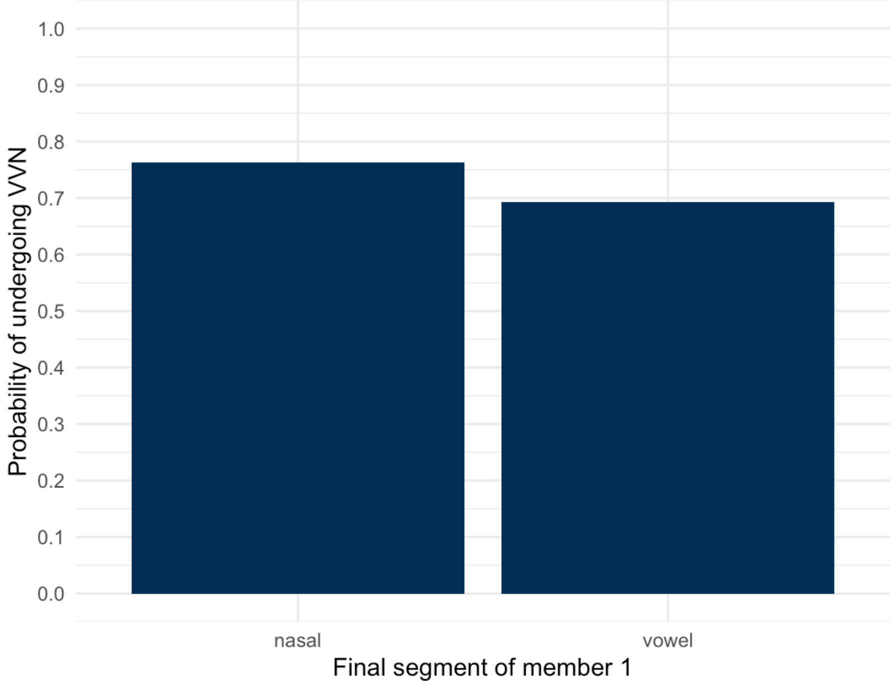
Not accept [ŋ] ¦ Accept [ŋ]

- Transforming a variable with 7 ordinal scales into a binary variable (consistent with the corpus data)
- Why 7 scales?

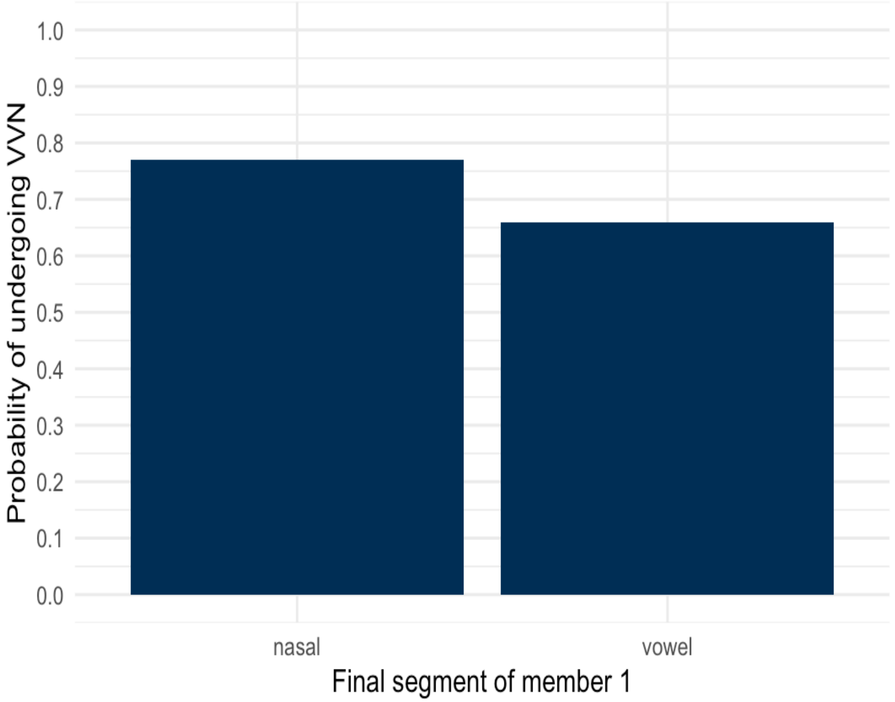
· To impartially present both forms, without implying that ‘[g] is the underlying form’ (vs. 2 options: [ŋ] is acceptable; [ŋ] is not acceptable)

· To prevent the risk of categorizing less probable [ŋ] as entirely impossible when the scale is too coarse. (vs. 3 options: only [g]; equal; only [ŋ])

Results
nasality of the preceding segment

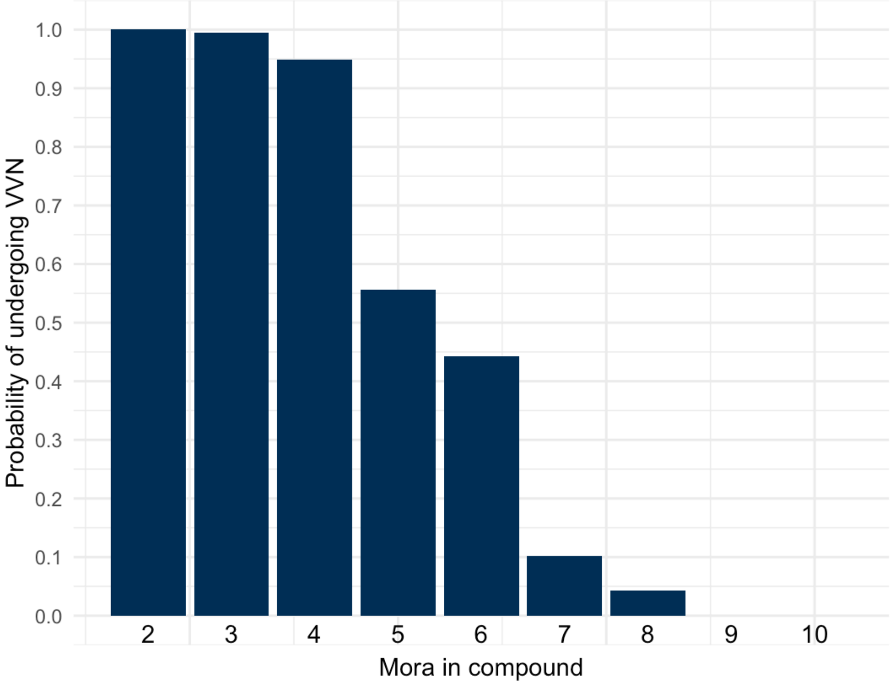


Trend in real lexicon

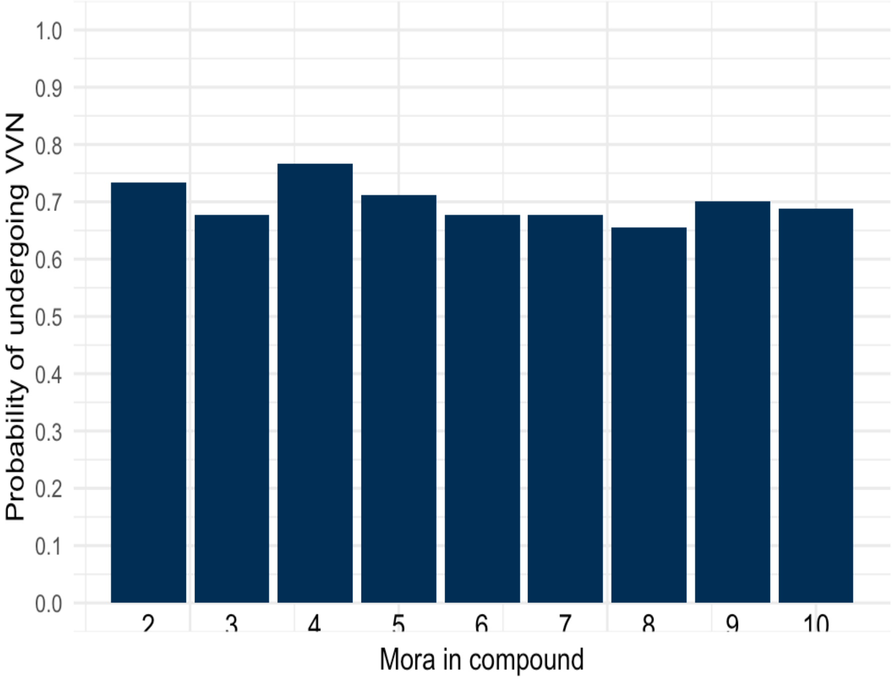


Trend in nonce words

Results
mora length



Trend in real lexicon



Trend in nonce words

- Mean value: 0.705 (real); 0.699 (nonce)
  Frequency matching (Hayes et al. 2009)

Results
Statistics

- Mixed-effect logistic model (with *max random effect*)
- glmer(Nasalized_Response ~ nas + length + *(1 + nas + length | subject)* + *(1 + nas + length | word)*, data = data, family = binomial)

- Results of fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	2.37357	1.10502	2.148	0.0317	*
nas	1.35168	0.53248	2.538	0.0111	*
length	-0.15144	0.09671	-1.566	0.1174	

Discussion

A learning bias against counting

- There exists a learning bias against counting-involved alternations.
- Second-order phonotactics is learned more slowly and with greater difficulty (although can be learned) (for a review, see Warker & Dell 2006)
- Unnatural patterns are disfavored and tend to be underlearned (e.g., Hayes et al. 2009)

What is the cause of the mora-counting pattern?

- A ‘surfeit of the stimulus’ effect (Becker et al. 2011, 2012)
  - Just an accidental generalization
  - No synchronic explanation required
  - Possibly due to a now inactive diachronic process

An alternative: token frequency

- The longer a word is, the less frequently it appears. (Zipf 2013)
- 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. (Endress & Hauser 2011)

a hypothesis: for long /X-gY/ where [gY] is a free word, [ŋ] as an exceptional variant requires high token frequency, otherwise hard to be learned.

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

A following study:
artificial language learning

- Extend the nonce word test from Japanese to artificial language

  (with prior training to subjects)
- Conclusion: no counting-involved pattern can be learned, no matter what length (number of units) is stated in the context of a rule.

Conclusion

- Japanese speakers internalize the natural factor of **nasality of the preceding segment** conditioning the tendency of /g/ undergoing nasalization
- But they fail to directly internalize the unnatural factor of **the number of moras**, although it *significantly* conditions the trend in the lexicon

References I

- 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., 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.
- Carr, P. (2006). Universal grammar and syntax/phonology parallelisms. *Lingua*, *116*(5), 634-656.
- 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.

References II

- Itô, J., & Mester, A. (1996). Correspondence and compositionality: The ga-gyō variation in Japanese phonology.
- Hibiya, J. (1995). The velar nasal in Tokyo Japanese: A case of diffusion from above. *Language Variation and Change, 7*(2), 139-152.
- Hayes, B., Siptár, P., Zuraw, K., & Londe, Z. (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language*, 822-863.
- McCarthy, J. J., & Prince, A. (1999). Prosodic morphology 1986. *Phonological theory: the essential readings*, 238-288.
- Paster, M. (2019). Phonology counts. *Radical: A Journal of Phonology*, *1*, 1-61. White, J. (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition*, *130*(1), 96-115.
- Peperkamp, S., Skoruppa, K., & Dupoux, E. (2006). The role of phonetic naturalness in phonological rule acquisition.
- 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.
- Zipf, G. K. (2013). *The psycho-biology of language: An introduction to dynamic philology*. Routledge.

Acknowledgements

- My sincere gratitude to Andrew Nevins, James White, Andrew Lamont, Canaan Breiss and audiences at Isa summer institute in UMass Amherst for their valuable comments on this project.
- Thanks to the students/faculty in linguistics 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).
- Thanks to 7 anonymous reviewers for their feedback.

Thank you!

Hailang Jiang
hailang.jiang.22@ucl.ac.uk
hailangjiang.github.io