Rendaku is not blocked by two nasal consonants: A reply to Kim (2022)

Abstract

Rendaku is a morphophonological process in Japanese, in which the first consonant of the second element of compounds becomes voiced (e.g. /ao-sora/ → [ao-zora] 'blue sky'). In a recent paper, Kim (2022) argues that the presence of two nasals—but not one nasal—in the second element reduces the probability of Rendaku application (e.g. /umi-kamome/ → [umi-kamome], *[umi-gamome] 'seagull'). This claim has an important consequence for grammatical theories, as it would constitute a case in which the influences of two segments add up to block a phonological process, a pattern that is referred to as "super-additive counting cumulativity." We have reexamined the empirical basis of this claim, and have found that there are only four relevant lexical items. Moreover, two of these items actually undergo Rendaku almost all the time, and there are thus only two items whose Rendaku is clearly blocked. The empirical basis of the theoretical claim based on Rendaku thus seems to be weak at best. With this, we move on to the main contribution of this paper, which is a new nonce-word experiment that aimed to resolve the conflicting results from the previous studies on this topic. The results of the current experiment show that the applicability of Rendaku is not reduced by the presence of two nasal consonants.

1 Introduction

Rendaku is a morphophonological process in Japanese, in which the initial obstruent of the second element of a compound becomes voiced, as illustrated by the examples in (1). For the rest of this paper, we will abbreviate "the second element of a compound" as "E2."

(1) Examples of Rendaku

a. /ko+**t**akaɾa/ → [ko+**d**akaɾa] 'babies as treasures'

b. /usagi+**k**umi/ → [usagi+**g**umi] 'Team Rabbit'

c. /ao+**s**oɾa/ → [ao+**z**oɾa] 'blue sky'

9 d.　　/tsuɾi+**h**aɕi/ → [tsuɾi+**b**aɕi] 'hanged bridge'

10 There are many factors that affect the applicability of Rendaku, and whether Rendaku applies
11 or not to a particular compound is ultimately idiosyncratic and unpredictable, although we can
12 identify some statistical tendencies (see Vance 2022 for the most up to date comprehensive re-
13 view). Still, one almost exception-less generalization is that Rendaku does not occur when E2
14 already contains a voiced obstruent, as illustrated by the examples in (2). This blockage of Ren-
15 daku is known as Lyman's Law (Lyman 1894), and has been extensively studied in the theoretical
16 phonology literature (e.g. Ito & Mester 1986, 2003; Kubozono 2005; see Kawahara & Zamma 2016
17 for a review).

18 (2)　　Examples of Rendaku blockage due to Lyman's Law

19 a.　　/ɕi+**t**oka**g**e/ → [ɕi+**t**oka**g**e], *[ɕi+**d**oka**g**e] 'fire lizard'
20 b.　　/kana+**k**u**g**i/ → [kana+**k**u**g**i], *[kana+**g**u**g**i] 'iron nail'
21 c.　　/ɕijaɕi+**s**o**b**a/ → [ɕijaɕi+**s**o**b**a], *[ɕijaɕi+**z**o**b**a] 'cold soba'
22 d.　　/joko+**h**a**b**a/ → [joko+**h**a**b**a], *[joko+**b**a**b**a] 'horizontal length'

23 In a recent paper, Kim (2022) has proposed a novel generalization about Rendaku: the presence
24 of two nasal consonants in E2—but not that of one nasal consonant—probabilistically lowers the
25 applicability of Rendaku. If this claim is true, it implys that two segments can "gang-up" to block
26 a phonological process, a case which Kim (2022) refers to as "super-additive counting cumulativ-
27 ity" (see also Breiss 2020, Breiss & Albright 2022, Jäger & Rosenbach 2006 and Kawahara & Breiss
28 2021 on cumulativity and additivity). In order to account for this pattern, Kim (2022) proposes a
29 modification to MaxEnt Harmonic Grammar (e.g. Flemming 2021; Hayes 2022; Hayes & Wilson
30 2008), in which the effects of constraint violations are scaled by the number of violation marks
31 using a power function. In short, Kim (2022) has identified a hitherto unnoticed generalization
32 about Rendaku, a phenomenon that has been studied extensively in the literature (Vance 2022),
33 which is thus in and of itself important; in addition, this new finding has important theoretical
34 consequences, as it would require a non-trivial modification to a widely-used theoretical frame-
35 work such as MaxEnt Harmonic Grammar.

2　Re-examining the empirical basis of the claim

37 However, there were some reasons to be careful about this new finding. Kim (2022) is based on
38 Kim (2020), and Kawahara & Kumagai (2021) point out that reexamination of the data used by
39 Kim (2020) shows that the blockage of Rendaku in many examples can and should be explained in
40 terms of independently motivated restrictions on Rendaku. For example, some examples in Kim

(2020) included morphologically complex nouns (e.g. [tate-mono] 'building'), whose Rendaku is independently known to be blocked by a restriction called Right Branch Condition (Ito & Mester 1986; Kubozono 2005; Otsu 1980). Other examples involved violations of Lyman's Law (e.g. [tabe-mono] 'food'), and hence required no additional explanations.

Kim (2022) uses a method that is different from that of Kim (2020), and bases her claim on the Rendaku Database (Irwin et al. 2020), which is the most comprehensive database on Rendaku. It is reported that verbal compounds as well as those compounds whose E2 are not native words were excluded (see Vance 2022: §7.4 and §7.3, respectively, for discussion on the influences of these factors on Rendaku). Those E2s that are hetero-morphemic were also excluded, as Rendaku would be blocked by Right Branch Condition (more technically speaking, those morphemes that are written by more than one *kanji* characters were excluded). Care was taken to exclude other factors that are known to inhibit Rendaku. For example, coordinated compounds, which are also known not to undergo rendaku (Vance 2022: §7.6), are reported to have been excluded.

The crucial generalizations are found in Table 3 of Kim (2022): it is reported that, setting aside the cases that violate Lyman's Law and words that would not undergo Rendaku in the first place (e.g. sonorant-initial words), forms containing no nasals undergo Rendaku 85% of the time; those forms that contain one nasal undergo Rendaku 80% of the time. Forms that contain two nasals undergo Rendaku 52% of the time. A logistic regression analysis reported by Kim (2022) shows that the Rendaku inhibition effect of two nasals is statistically significant.

It is reported that there are 29 mono-morphemic forms that contain two nasals, but the paper provides only two examples, [kamome] 'seagull' and [tsumami] 'knob', in the text. It is unlikely that there are 29 mono-morphemic distinct lexical items containing two nasals, and thus this number is presumably a token frequency rather than a type frequency. To explore how many relevant lexical items there are—and what these lexical items actually are—we tried to replicate the procedure described by Kim (2022) by consulting the Rendaku Database ourselves (Irwin et al. 2020).[1] We extracted compounds whose E2 is a native word that contains two nasal consonants. There were three E2s that were written with two *kanji* characters and are heteromorphemic ([kana-mono] 'hardware', [kane-motɕi] 'rich people' and [ki-mono] 'kimono'), which were excluded. There was one token which violated moraic identity avoidance across a morpheme boundary ([mata+tanomi] 'asking indirectly'), whose Rendaku is independently known to be inhibited (Kawahara & Sano 2014), which was thus excluded. One example, [netami-sonemi] 'jealous,' was excluded as a coordinate compound. As with Kim (2022), we excluded examples whose Rendaku status cannot be unambiguously determined from the database (e.g. those tokens whose Rendaku application is optional).

As a result of this procedure, we found that there were four distinct lexical items (29 tokens in

---

[1]Since our analysis is based on a password-protected database, we cannot make our analysis publicly available. However, pending Mark Irwin's approval, we are happy to share the Excel file which was used for this analysis.

total) that would satisfy the relevant conditions, i.e. native words which can potentially undergo Rendaku and contain two nasal consonants, which are listed in (3), together with their token frequencies in the Rendaku Database.

(3)   The crucial examples of E2s with two nasal consonants

    a.   [kamome] 'sea gull', 11 tokens

    b.   [konom-i] 'favorite' (from [konom-u] 'to like'), 8 tokens

    c.   [tanom-i] 'plea' (from [tanom-u] 'to ask'), 8 tokens

    d.   [tsumam-i] 'knob' (from [tsumam-u] 'to grab'), 2 tokens

A closer examination shows that of the forms listed in (3), [tanom-i] undergo Rendaku all the time (8 tokens) and [konom-i] undergo Rendaku 7 out of 8 times. The last item [tsumam-i] (in addition to [konom-i] and [tanom-i]) is a deverbal noun with suffixal [i], and hence it is arguably morphologically complex. However, since deverbal nouns do undergo Rendaku in some environments, we can probably treat them as mono-morphemic. However, it is important to bear in mind that there are complex conditions on Rendaku applications in deverbal nouns as well (see Fukasawa 2020, 2022; Kozman 1998; Yamaguchi 2011; Vance 2022). For example, when E1 is an argument of E2, Rendaku rarely occurs, and a recent experimental study by Fukasawa (2022) shows that specific thematic relationships between E1 and E2 influence the applicability of Rendaku in a complex fashion. There is at least one example in the Rendaku Database, [hana-tsumami] 'nose holding,' whose Rendaku is blocked for this reason. Thus, there is only one instance of [tsumami] ([çito-tsumami] 'one grab') which would instantiate the blockage of Rendaku by two nasal consonants.

    Therefore, [kamome], which never undergoes Rendaku in the database, is the only lexical item that unambiguously supports the generalization that two nasals block Rendaku. One may argue that one token of [tsumami] can also be used as an argument for this generalization.[2] This result raises the question of how generalizable this Rendaku blockage effect is. We should bear in mind that the application of Rendaku can be idiosyncratic and unpredictable (Vance 2022), and therefore, making a generalization based on a few lexical items can be dangerous.

    To conclude, we find this empirical basis to be weak at best, although perhaps not entirely unreliable. We thus felt that it is important to reexamine this claim with experimentation, especially given that the theoretical consequence of Kim (2022) is an important one. With this, we now turn to the major contribution of the current paper, a nonce word experiment which examined the claim that two nasals affect the applicability of Rendaku.

---

[2] To this, we can add [kaname] 'the pivot,' [hanamuke] 'present' and [kaminaɾi] 'lightening' as examples of mono-morphemic words with two nasals which do not seem to undergo Rendaku. [tanoɕim-i] 'fun' and [kanaɕim-i] 'sadness' are other possible examples, although they are deverbal nouns.

3 The current experiment

There are two nonce-word experiments that have previously tested the effects of two nasals on Rendaku. The first experiment is Kumagai (2017), which like Kim (2020, 2022), found that two nasal consonants reduce the Rendaku application rate. However, Kumagai (2017) had only three items for each condition, and moreover, all the items for the relevant condition were of the form [hVnama]; therefore, generalizability of this finding is suspicious. The next experiment is Kawahara & Kumagai (2021) (their Experiment 2), which had 6 items per each condition. Their baseline condition was forms in which second and third syllables had an obstruent onset. They found that forms that had a nasal consonant in the second and third syllables undergo Rendaku slightly less often than the baseline condition, but the magnitude of the difference was very small (about 3.5%), and this difference was not statistically credible in their Bayesian analysis.

In the current experiment, we tried to examine whether the Rendaku blockage effect by two nasal consonants can be identified in a more robust fashion. Since the difference between the two critical conditions in Kawahara & Kumagai (2021) was in the direction that was expected from Kumagai's (2017) results, there remains a possibility that that experiment was underpowered (although their data is based on responses from 143 speakers). There were a few other aspects that the current experiment attempted to expand on Kawahara & Kumagai (2021). First, the experiment by Kawahara & Kumagai (2021) did not directly compare the two nasal condition with forms that contain a voiced obstruent (i.e. forms that violate Lyman's Law), which would have been informative, as the latter should clearly lower Rendaku responses. Another limitation, which actually pertains through both of the previous experiments (Kawahara & Kumagai 2021; Kumagai 2017), is that all the target nonce words began with [h].

The current experiment was designed to overcome these limitations. First, the current experiment tested all the consonants that can potentially undergo Rendaku ([t], [k], [s], and [h], see (1)). Second, the experiment had 16 items per each condition to make sure that the results are not solely due to the effects of particular forms (e.g. [hVnama] forms used by Kumagai 2017). Third, we aimed to collect data from more speakers than Kawahara & Kumagai (2021) to examine if the non-credible difference found by Kawahara & Kumagai (2021) was due to an insufficient number of participants (for specifics, see below). Fourth, the experiment directly compared forms with two nasals and forms that contain a voiced obstruent (i.e. those that violate Lyman's Law).

3.1 Methods

Following the spirit of the open science initiative in linguistics (see e.g. Berez-Koreker et al. 2018 and Winter 2019), the raw experimental data file, the R markdown file as well as the Bayesian

posterior samples are available at an Open Science Framework (OSF) repository.[3]

3.1.1 Overall design

The experiment consisted of four conditions: (1) the baseline condition, in which the second and third syllables contained an obstruent onset, (2) the one nasal condition, in which the second syllable contained a nasal consonant, (3) the two nasal condition, in which the second and third syllable contained a nasal onset, and (4) the Lyman's Law condition, which contained a voiced obstruent in the second syllable.

If Kumagai (2017) and Kim (2022) are on the right track, the third condition, but not the second condition, should show lower Rendaku responses than the first condition. The second condition was included to make sure that it is additive effects of two nasals, not the presence of one nasal consonant, that would reduce the Rendaku applicability. The fourth condition was included to compare the effects of Lyman's Law and those of two nasal consonants. Since Lyman's Law is almost exception-less in contemporary Japanese (Vance 2022), while the blockage of Rendaku by two nasals is at best probabilistic (Kim 2022), the fourth condition may show lower Rendaku responses than the third condition.[4] However, if the blockage force against Rendaku by two nasal consonants is as productive as that of Lyman's Law, then the third and fourth conditions should show comparable Rendaku responses rates.

3.1.2 Stimuli

The list of the stimuli used in the current experiment is shown in Table 1. The experiment tested all four sounds that can potentially undergo Rendaku (=/t/, /k/, /s/ and /h/) with 4 nonce items each, resulting in 64 stimuli in total (4 conditions × 4 consonants × 4 items).

None of these words are existing words; neither do they become a real word when Rendaku is applied. All the stimuli consist of three light CV syllables. The vowel qualities across the four conditions were controlled as much as possible; however, sometime that would result in some stimuli sounding too similar to an existing word, which was avoided. Since it has been shown that Rendaku may be substantially inhibited when it results in identical CV mora sequences across a morpheme boundary (Kawahara & Sano 2014), no forms began with [se], since the E1 used in the experiment was [nise] 'fake' (see below).

---

[3]`https://osf.io/ceadx/?viewonly=2929cd72cbe24c04856a054715ed8b43`

[4]In her theoretical modeling of the data, Kim (2022) indeed uses two separete constraints for Lyman's Law and the blockage effect by two nasal consonants, so that the theory may predict that the third and fourth conditions show different Rendaku applicability rates. See also Nasukawa (2012) for possible connections between the features [voice] and [nasal], which may imply that blockage by nasal consonants and that by a voiced obstruent may be unified in some fashion.

Table 1: The list of nonce words used as E2s in the experiment.

	vls-vls	one nasal	two nasals	Lyman's Law
/t/	[tasake]	[tamake]	[taname]	[tazake]
	[takise]	[tamise]	[tamine]	[tagise]
	[tesaka]	[tenaka]	[tenano]	[tezaka]
	[tokasa]	[tonosa]	[tonomo]	[togosa]
/k/	[kasato]	[kamato]	[kanamo]	[kazato]
	[katesa]	[kamasa]	[kamena]	[kadesa]
	[ketase]	[kenase]	[kenane]	[kedase]
	[kotasa]	[tazanu]	[komana]	[kodasa]
/s/	[sakato]	[samato]	[sanamo]	[sagato]
	[sakike]	[samike]	[samine]	[sagike]
	[sotaka]	[sonaka]	[sonano]	[sodaka]
	[sotake]	[sonake]	[soname]	[sodake]
/h/	[hatasa]	[hamasa]	[hanomo]	[hadasa]
	[hakise]	[hamise]	[hamine]	[hagise]
	[hesaka]	[henaka]	[henano]	[hezaka]
	[hotosa]	[honosa]	[honoma]	[hodasa]

3.1.3 The participants

The experiment was conducted online using SurveyMonkey. A total of 236 native speakers of Japanese completed the online experiment. The responses were collected using a snow ball sampling method using Twitter. Only those who reported to be a native speaker of Japanese and those who have not heard about Rendaku were allowed to participate. The participants took part in the experiment completely voluntarily. There were no compensations, monetary or otherwise.

3.1.4 Procedure

Before starting the experiment, the participants read through the consent form, which was approved by the first author's institute. In the pre-experimental instructions, the participants were told that when we create compounds in Japanese, some combinations undergo voicing (i.e. Rendaku) while others do not, so that the participants become explicitly aware that Rendaku can apply in some items but not in others.

In the main session, the participants were asked to take each stimulus item and combine it with [nise] 'fake' as E1. They were then asked whether the resulting compound would sound more natural with or without Rendaku. A sample question, therefore, is, "given a nonce word [tasake], when it is combined with [nise], which form sounds more natural, [nise-tasake] or [nise-dasake]?" The stimuli were written in the *hiragana* orthography, which is used to represent native

words in Japanese. Before the main session, the participants went through two practice trials with existing compounds. The stimuli in the main trial session were presented to the participants as nonce words.[5] The order of the stimuli in the main trial sessions was randomized for each participant using a randomization function offered by SurveyMonkey.

3.1.5 Statistical analyses

The results were analyzed using a Bayesian mixed effects logistic regression model, implemented with the `brms` package (Bürkner 2017) and R (R Development Core Team 1993–). Bayesian statistics take prior information and the data obtained by an experiment to yield a posterior distribution for each parameter that we would like to estimate (for accessible introductions to Bayesian modeling, see e.g. Franke & Roettger 2019; Kruschke 2014; Kruschke & Liddell 2018; McElreath 2020; Vasishth et al. 2018). One advantage of Bayesian analyses is that we can interpret these posterior distributions as directly reflecting our certainty about the estimates. As a useful heuristic, we can examine the middle 95% of the posterior distribution, known as 95% Credible Interval (henceforth, 95% CrI). If a 95% CrI does not include 0, then we can interpret that effect to be meaningful. If it includes 0, then we can examine its posterior distribution more carefully to determine with how much certainty we can conclude the null effect. This ability to be able to test null effects is another advantage of Bayesian analyses (Gallistel 2009).

The details of the current logistic regression model are as follows. The dependent variable was whether each item was judged to undergo Rendaku or not (yes-Rendaku response = 1 vs. no-Rendaku response = 0). One fixed explanatory variable included in the model was the four conditions shown in Table 1, and we set the first condition (forms with only voiceless obstruents) to be the baseline. Another fixed factor included in the model was four types of segments. Since we had no a priori expectation about the differences between the four segments, [h] was arbitrarily chosen as the baseline. The interaction term between the two factors was also coded, since we wanted to see if the differences between the four conditions may depend on the segment type. A random intercept of items was included in the model, and so were the random intercepts and slopes of participants for the two fixed factors and their interaction term.

The prior specifications were as follows: for all slope coefficients, we used a Cauchy distribution prior with scale of 2.5, following Gelman et al. (2018), and for the intercept, we used Normal(0, 1) weakly informative priors (Lemoine 2019). We ran four chains with 4,000 iterations each, and discarded the first 1,000 iterations from each chain as warmups.[6] All the R̂-hat values

---

[5]Rendaku primarily applies in native words (Vance 2022: §7.3); however, a previous experiment has shown that presenting the stimuli either as nonce words or obsolete native words does not substantially affect the Rendaku responses, or the influence of Lyman's Law on Rendaku (Kawahara 2012).

[6]Running only 2,000 iterations with 1,000 warmups resulted in divergent transitions and R̂-hat values that were higher than 1.00. The materials prepared for "the Bayesian Analysis for the Speech Sciences (B4SS)" workshop

associated with the fixed effects were 1.00 and there were no divergent transitions, which suggest that the four chains mixed successfully. See the R Markdown file for complete details, which also includes additional analyses, such as illustration of conditional effects and a posterior predictive check.

3.2 Results

Figure 1 is a violin plot of the results, which shows the normalized probability distributions of Rendaku responses for each condition, separately for each segment shown in a different facet. Transparent circles represent averaged responses from each participant. Solid red circles represent grand averages across all the participants. Within each facet, violins are shown in the order of the baseline condition, the one nasal condition, the two nasal condition, and the Lyman's Law condition.
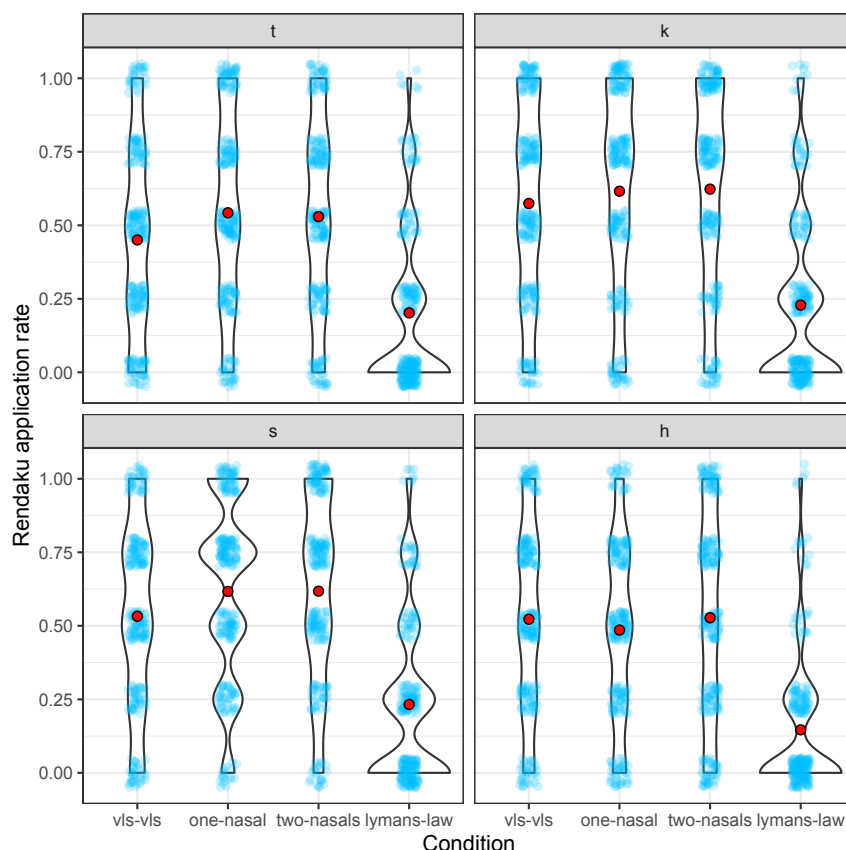

Figure 1: The distributions of Rendaku responses for each condition, separated by each segment type.

---

offer accessible introduction to these analytical concepts (`https://learnb4ss.github.io/`). See also the tutorial articles cited at the beginning of this section.

Across the four segments, we can observe that the first three conditions show comparable Rendaku application percentages. The Lyman's Law condition shows clearly low Rendaku application rates. The grand averages of the four conditions were 52.0%, 56.5%, 57.4% and 20.2%. The crucial comparison—the comparison between the baseline condition and the two nasal condition—does not seem substantial. If anything, the two nasal condition shows higher Rendaku responses than the baseline condition, especially for [t], [k] and [s].

Table 2 shows the model summary of the Bayesian regression model. For all the interaction terms, their 95% CrI include 0, which suggest that the differences among the four conditions are comparable across the four segments, which makes it easier for us to interpret the main effects. The 95% CrI for the main effect of segments also all include 0, which suggests that the four segments behaved similarly at the baseline level. In fact, the only credible coefficient in the model is the difference between the baseline condition and the Lyman's Law condition, the latter of which showed credibly lower Rendaku responses. This result replicates many previous nonce-word experiments on Rendaku which have established the productivity of Lyman's Law (e.g. Ihara et al. 2009; Kawahara 2012; Kawahara & Kumagai 2021; Vance 1980).

Table 2: Summary of the Bayesian mixed effects logistic regression model.

		β	error	95% CrI
(a) intercept		0.02	0.23	[-0.44, 0.45]
(b) condition	one nasal	-0.08	0.30	[-0.69, 0.51]
	two nasal	0.12	0.31	[-0.49, 0.73]
	Lyman's Law	-2.48	0.34	[-3.15, -1.81]
(c) sound	[t]	-0.28	0.32	[-0.92, 0.34]
	[k]	0.34	0.32	[-0.27, 0.97]
	[s]	0.15	0.32	[-0.46, 0.79]
(d) interactions	one nasal:[t]	0.53	0.44	[-0.36, 1.37]
	one nasal:[k]	0.32	0.44	[-0.53, 1.19]
	one nasal:[s]	0.51	0.44	[-0.36, 1.36]
	two nasal:[t]	0.27	0.44	[-0.59, 1.14]
	two nasal:[k]	0.18	0.45	[-0.71, 1.06]
	two nasal:[s]	0.34	0.44	[-0.53, 1.20]
	LL:[t]	0.63	0.47	[-0.29, 1.54]
	LL:[k]	0.25	0.46	[-0.65, 1.17]
	LL:[s]	0.58	0.46	[-0.33, 1.48]

The coefficient for the comparison between the baseline condition and the two nasal condition was positive (0.12), reflecting the observation above that the two nasal condition showed slightly higher Rendaku responses. However, its 95% CrI [-0.49, 0.73] include 0, suggesting that

this difference is not meaningful.

This result itself does not show that we can accept the conclusion that there is no difference between the two conditions. In Bayesian analyses, we can calculate how confident we can be about null effects using a so-called ROPE (Region Of Practical Equivalence) analysis (Kruschke & Liddell 2018; Vasishth & Gelman 2021). In this analysis, we define a range that is practically equivalent to 0. Following the suggestion by Makowski et al. (2019), we took the effect size of 0.1, a negligible effect size (Cohen 1988), to be practically equivalent of 0, which in logistic regression models corresponds to [-0.18, 0.18]. We used `bayestestR` package (Makowski et al. 2020) to calculate how many posterior samples of the posterior estimates are included in this ROPE and found that 42% of them are. In short, we can only be 42% certain about the null effect.

However, more importantly, it is at least the case that the two nasals do not lower Rendaku responses. As the final step of the analysis, we calculated how many posterior samples of this coefficient were negative, and found that only 35.4% of them were. Therefore, the productivity of the Rendaku blockage effect by two nasal consonants is not supported by the current experiment.

4 Conclusion

We have reexamined the recent claim by Kim (2022) that two nasal consonants reduce the applicability of Rendaku in Japanese. The analysis of the Rendaku Database (Irwin et al. 2020) shows that there are only four relevant lexical items, and two of them actually undergo Rendaku. Although we could come up with a few additional examples, the robustness of the empirical generalization is still questionable, if not entirely unreliable. With this result in mind, in order to resolve conflicting results from the previous studies, we also ran a new nonce-word experiment, which includes many more items and data from more participants, compared to the previous experimental studies. The emerging conclusion seems to be that there is no strong evidence that two nasals lower the applicability of Rendaku, which undermines the claim that Rendaku instantiates a case of super-additive counting cumulativity. We hasten to add, however, that Kim (2022) reports another case study from Korean compounding, and we have nothing to say about that pattern. With this said, one of the two empirical bases of Kim's (2022) theoretical proposal should be considered to be not very robust.

Data availability/Supplementary files

The experimental data and the analysis files are available at the OSF repository. See the method section.

Ethics and consent

Funding information

Acknowledgements

Competing interests

There are no competing interests to declare.

Authors' contributions

Conception of the study: XX and YY. Analyses of the Rendaku Database XX and YY. Designing the experiment: YY and XX. Statistical analyses: XX. Writing up the paper: XX and YY.

References

Berez-Koreker, Andrea, Lauren Gawne, Susan Smythe Kung, Barbara F. Kelly, Tyler Heston, Gary Holton, Peter Pulsifer, David I. Beaver, Shobhana Chelliah, Stanley Dubinsky, Richard P. Meier, Nick Thieberger, Keren Rice & Anthony C. Woodbury. 2018. Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics* 56(1). 1–18.

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

Breiss, Canaan & Adam Albright. 2022. Cumulative markedness effects and (non-)linearity in phonotactics. *Glossa* 7(1).

Bürkner, Paul-Christian. 2017. brms: An R Package for Bayesian Multilevel Models using Stan. R package.

Cohen, Jacob. 1988. *Statistical power analysis for the behavioral science.* Hillsdale: Lawrence Erlbaum Associates.

Flemming, Edward. 2021. Comparing MaxEnt and Noisy Harmonic Grammar. *Glossa* 6(1). 141.

Franke, Michael & Timo B. Roettger. 2019. Bayesian regression modeling (for factorial designs): A tutorial. Ms. `https://doi.org/10.31234/osf.io/cdxv3`.

Fukasawa, Michiko. 2020. Rendaku in syntax-phonology interface: A corpus study on deverbal noun compounds. In Michael Barrie (ed.), *Japanese korean linguistics 27*, CSLI Publications.

Fukasawa, Michiko. 2022. When syntax and semantics of compounds matter to voicing alternations: An experimental investigation of effects of argument structure on rendaku. Ms. University of Hawaii.

Gallistel, Randy C. 2009. The importance of proving the null. *Psychological Review* 116(2). 439–453.

Gelman, Andrew, A. Jakulin, M.G. Pittau & Y-S. Su. 2018. A weakly infmroative default prior distribution for logistic and other regression models. *Annual Applied Statistics* 1360–1383.

Hayes, Bruce. 2022. Deriving the wug-shaped curve: A criterion for assessing formal theories of linguistic variation. *Annual Review of Linguistics* 8. 473–494.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39. 379–440.

Ihara, Mutsuko, Katsuo Tamaoka & Tadao Murata. 2009. Lyman's Law effect in Japanese sequential voicing: Questionnaire-based nonword experiments. In The Linguistic Society of Korea (ed.), *Current issues in unity and diversity of languages: Collection of the papers selected from the 18th International Congress of Linguists*, 1007–1018. Seoul: Dongam Publishing Co., Republic of Korea.

Irwin, Mark, Mizuki Miyashita, Kerrim Russel & Yu Tanaka. 2020. The Rendaku Database v4.0.

Ito, Junko & Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17. 49–73.

Ito, Junko & Armin Mester. 2003. *Japanese morphophonemics*. Cambridge: MIT Press.

Jäger, Gerhard & Anette Rosenbach. 2006. The winner takes it all—almost: Cumulativity in grammatical variation. *Linguistics* 44(5). 937–971.

Kawahara, Shigeto. 2012. Lyman's Law is active in loanwords and nonce words: Evidence from naturalness judgment experiments. *Lingua* 122(11). 1193–1206.

Kawahara, Shigeto & Canaan Breiss. 2021. Exploring the nature of cumulativity in sound symbolism: Experimental studies of Pokémonastics with English speakers. *Laboratory Phonology* 12(1). 3, `https://doi.org/10.5334/labphon.280`.

Kawahara, Shigeto & Gakuji Kumagai. 2021. Does Lyman's Law count? Ms. Keio University and Kansai University.

Kawahara, Shigeto & Shin-ichiro Sano. 2014. Identity avoidance and rendaku. *Proceedings of Phonology 2013* .

Kawahara, Shigeto & Hideki Zamma. 2016. Generative treatments of rendaku. In Timothy Vance & Mark Irwin (eds.), *Sequential voicing in Japanese compounds: Papers from the NINJAL rendaku Project*, 13–34. Amsterdam: John Benjamins.

Kim, Seoyoung. 2020. Modeling super-gang effects in MaxEnt: Nasal in Rendaku. *Proceedings of NELS 49* 175–188.

Kim, Seoyoung. 2022. A maxent learner for super-additive counting cumulativity. *Glossa* 7(1).

Kozman, Tam. 1998. The psychological status of syntactic constraints on *rendaku*. In David Silva (ed.), *Japanese/Korean linguistics 8*, 107–120. Stanford: CSLI.

Kruschke, John K. 2014. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan.* Waltham: Academic Press.

Kruschke, John K. & Torrin M. Liddell. 2018. The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychological Bulletin and Review* 25. 178–206.

Kubozono, Haruo. 2005. Rendaku: Its domain and linguistic conditions. In Jeroen van de Weijer, Kensuke Nanjo & Tetsuo Nishihara (eds.), *Voicing in Japanese*, 5–24. Berlin & New York: Mouton de Gruyter.

Kumagai, Gakuji. 2017. Super-additivity of OCP-nasal effect on the applicability of rendaku. Talk presented at GLOW in Asia XI.

Lemoine, N.P. 2019. Moving beyond noninformative priors: Why and how to choose weakly informative priors in bayesian analyses. *Oikos* 128. 912–928.

Lyman, Benjamin S. 1894. Change from surd to sonant in Japanese compounds. *Oriental Studies of the Oriental Club of Philadelphia* 160–176.

Makowski, Dominique, Mattan S. Ben-Shachar, Annabel S.H. Chen & Daniel Lüdecke. 2019. Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology* https://doi.org/10.3389/fpsyg.2019.02767.

Makowski, Dominique, Daniel Lüdecke, Mattan S. Ben-Shachar, Michael D. Wilson, Paul-Christian Bürkner, Tristan Mahr, Henrik Singmann, Quentine F. Gronau & Sam Crawley. 2020. bayestestR. R package.

McElreath, Richard. 2020. *Statistical Rehinking: A Bayesian Course with Examples in R and Stan, 2nd edition.* London: Taylor & Francis Ltd.

Nasukawa, Kuniya. 2012. *A unified approach to nasality and voicing.* Berlin: Mouton De Gruyter.

Otsu, Yukio. 1980. Some aspects of rendaku in Japanese and related problems. In Ann Farmer & Yukio Otsu (eds.), *MIT working papers in linguistics*, vol. 2, 207–228. Cambridge, Mass.: Department of Linguistics and Philosophy, MIT.

R Development Core Team. 1993–. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing Vienna, Austria.

Vance, Timothy. 1980. The psychological status of a constraint on Japanese consonant alternation. *Linguistics* 18. 245–267.

Vance, Timothy. 2022. *Irregular phonological marking of Japanese compounds.* Berlin: Mouton de Gruyter.

Vasishth, Shravan & Andrew Gelman. 2021. How to embrace variation and accept uncertainty in linguistic and psycholinguistic data analysis. *Linguistics* 59(5). 1311–1342.

Vasishth, Shravan, Bruno Nicenboim, Mary Beckman, Fangfang Li & Eun Jong Kong. 2018. Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics* 71. 147–161.

Winter, Bodo. 2019. *Statistics for linguists.* New York: Taylor & Francis Ltd.

Yamaguchi, Kyoko. 2011. Accentedness and *rendaku* in Japanese deverbal compounds. *Gengo Kenkyu [Journal of the Linguistic Society of Japan]* 140. 117–134.