Partially activated morpheme boundaries in Japanese surnames[*]

Shigeto Kawahara & Yu Tanaka

Keio University, Doshisha University

Abstract

The theory of Gradient Symbolic Representation posits that phonological structures can be partially activated, going beyond the traditional binary presence vs. absence dichotomy. In this paper, we propose that morphological structures can also be partially activated. Empirical evidence comes from the patterns of Rendaku in Japanese surname compounds. Rendaku is a morphophonological process in which the initial voiceless obstruent of the second member of a compound becomes voiced. Lyman's Law blocks this process if there is already a voiced obstruent in the second member, but a voiced obstruent in the first member does not block Rendaku; i.e. Lyman's Law does not apply across a morpheme boundary. However, in surname compounds, the presence of a voiced obstruent in the first member substantially reduces the applicability of Rendaku, unlike in regular compounds. At the descriptive level, surname compounds behave ambiguously between monomorphemic and heteromorphemic words: they undergo the compound voicing process but are also probabilistically subject to the morpheme-internal restriction against multiple voiced obstruents. We demonstrate that this observation can be formally modeled by positing a partially activated morpheme boundary in surname compounds. A quantitative modeling of the data is developed using Gradient Symbolic Representation coupled with Maximum Entropy Harmonic Grammar.

1. Introduction

The theory of Gradient Symbolic Representation (GSR) posits that phonological structures can be gradiently activated, going beyond the traditional binary presence vs. absence dichotomy (Smolensky and Goldrick 2016). This proposal is a part of the larger theory of Gradient Symbolic Computation, a general theory of cognition integrating gradient neutral

[*]We would like to thank Canaan Breiss, Donna Erickson, Hironori Katsuda, Jason Shaw, Timothy Vance as well as the participants at NELS 51 for comments on this project. Eric Rosen and Brian Hsu provided very helpful comments on previous versions of this paper, for which we are very grateful. All remaining errors are ours.

networks and categorical symbolic representations. While the original proposal was illustrated with an analysis of French liaison, its generality was clearly anticipated by Smolensky and Goldrick (2016). In fact, since its initial proposal in 2016, the theory of the GSR has been extended to account for various phonological and morphophonological phenomena across a wide variety of languages (e.g., Guekguezian and Jesney 2020, Hsu 2019, Hsu and Jesney 2016, 2017, Rosen 2016, 2018, Smolensky et al. 2020, Walker 2020, Zimmermann 2018, 2019, 2020). Particularly relevant to the current study is the proposal made by Shih (2020), who argues that category membership can be gradiently activated.[1] She demonstrates, for example, that in the patterns of sound symbolism, unisex names behave ambiguously between male names and female names in English—in Shih's proposal, some morphemes can be "partially male," suggesting that morphosyntactic features can be partially activated.

Our current research is partially theory-driven. If we take the GSR seriously as a general theory of linguistic knowledge, there is nothing that prevents *morpheme boundaries* from being partially activated.[2] Put in a more theory-neutral term, we expect that some compounds should behave ambiguously between monomorphemic words and heteromophemic words (as discussed by Hay 2003 and Hay and Baayen 2005 from a perspective that is slightly different from ours). We argue that this prediction of the GSR is borne out; the empirical evidence comes from the pattern of Rendaku in Japanese surnames.

2. Rendaku and Lyman's Law

The pattern that we analyze in this paper is Rendaku, a well-known and well-studied morphophonological process in Japanese, by which the morpheme-initial obstruent of the second member (=N2) in a compound undergoes voicing, as in (1). Rendaku is blocked when N2 already contains a voiced obstruent, as in (2), a generalization that is known as Lyman's Law (for more on Rendaku and Lyman's Law, see the collection of papers in Vance and Irwin 2016 and references cited therein).

(1) *Examples of Rendaku*

a.	/oɕi+**h**ana/	→	[oɕi+**b**ana]	'dried flower'
b.	/nise+**t**anuki/	→	[nise+**d**anuki]	'fake raccoon'
c.	/juki+**k**umi/	→	[juki+**g**umi]	'Snow Team'
d.	/hoɕi+**s**ora/	→	[hoɕi+**z**ora]	'starry sky'

[1] The idea of "fuzzy" category membership is actually much older, going back at least to Ross (1972) and Lakoff (1973), who claimed that there can be different degrees of "noun-y-ness." The idea of fuzzy category membership is likely to be even older, going back to Edward Sapir, Ludwig Wittgenstein and possibly older scholars. See Aarts (2004) for a historical review of these ideas.

[2] Indeed, see §6.1 of Smolensky and Goldrick (2016), which foreshadows the current proposal that morpheme structures too can be gradiently activated.

(2) *Blocking of Rendaku by Lyman's Law*

a.	/çito+**h**a**d**a/	→	[çito+**h**a**d**a]	*[hito+**b**a**d**a]	'dried flower'
b.	/nise+**t**oka**g**e/	→	[nise+**t**oka**g**e]	*[nise+**d**oka**g**e]	'fake raccoon'
c.	/çito+**k**a**g**e/	→	[çito+**k**a**g**e]	*[hito+**g**a**g**e]	'people's shadow'
d.	/moɾi+**s**o**b**a/	→	[moɾi+**s**o**b**a]	*[moɾi+**z**o**b**a]	'cold soba'

Whether Rendaku applies is subject to various factors, both linguistic and idiosyncratic (Vance and Irwin 2016); however, the blocking of Rendaku due to Lyman's Law exemplified by the examples in (2) is rather robust, and there are only three well-established lexical exceptions: [nawa-**b**aɕi**g**o] 'string ladder', [X-**z**a**b**uɾoo] 'personal name', and [ɸun-ʑi**b**aɾu] 'to tie tightly' (Vance 2015).

Unlike a voiced obstruent in N2, a voiced obstruent in the first member (=N1) of a compound does not block Rendaku; i.e. Lyman's Law is operative within a morpheme but does not apply across a morpheme boundary, as shown by the examples in (3) (Irwin 2014, Vance 2015).[3]

(3) *Lyman's Law does not apply across a morpheme boundary*

a.	/ka**g**e+**h**oɕi/	→	[ka**g**e+**b**oɕi]	'dried in the shade'
b.	/ka**b**e+**k**ami/	→	[ka**b**e+**g**ami]	'wall paper
c.	/ha**d**a+**s**amui/	→	[ha**d**a+**z**amui]	'skin cold'
d.	/ta**b**i+**t**oɾi/	→	[ta**b**i+**d**oɾi]	'travel bird'

This much is very well-known in the literature. A less-well known puzzle about Rendaku, which we would like to address in this paper, is as follows. On the one hand, Lyman's Law is blocked by a morpheme boundary in normal compounds, as in (3). On the other hand, Rendaku appears to be less likely to apply when N1 contains a voiced obstruent in surname compounds (Kubozono 2005, Sugito 1965, Tanaka 2017, Zamma 2005, Zamma and Asai 2017). In some cases, Rendaku is blocked by a voiced obstruent in N1 as in (4a-b); in other cases, Rendaku seems to be rendered optional by a voiced obstruent in N1, as in (4c-d).

(4) *Effects of a voiced obstruent in N1 in surname compounds*

a.	/ka**d**o+**k**awa/	→	[ka**d**o+**k**awa]	*[ka**d**o+**g**awa]	'corner river' (surname)
b.	/na**g**a+**s**awa/	→	[na**g**a+**s**awa]	*[na**g**a+**z**awa]	'long swamp' (surname)
c.	/ɕi**b**a+**s**aki/	→	[ɕi**b**a+**s**aki] ~ [ɕi**b**a+**z**aki]		'brushwood cape' (surname)
d.	/mi**z**o+**ɸ**utɕi/	→	[mi**z**o+ɸutɕi] ~ [mi**z**o+**b**utɕi]		'trench pool' (surname)

Sugito (1965) is the first study which offered some quantitative data on this observation. She studied the frequency of Rendaku application in a set of surnames whose N2 is /ta/ 'rice field,' which is by far the most common morpheme appearing in surnames. The data are reproduced in Table 1. As can be seen, when a voiceless obstruent or a nasal appears in the last syllable of N1, the Rendaku-undergoing form is dominant (=158/174, 90.8%). On

[3]Old Japanese behaved differently from Modern Japanese with respect to how Lyman's Law applied, and we will come back to this issue in the Appendix.

the other hand, when N1 ends with a syllable containing a voiced obstruent, no Rendaku-undergoing form is observed. For this particular case, it appears that a voiced obstruent in N1 blocks Rendaku completely.

N1 last cons.	[**t**a]	[**d**a]
[s, t, k, m, n]	16	158
[z, b, d, g]	30	0

Table 1: *Patterns of Rendaku for the [ta]-[da] allomorphs, reported in Sugito (1965).*

Subsequent studies extended this observation by investigating surname compounds with N2s other than /ta/ (Kubozono 2005, Tanaka 2017, Zamma 2005, Zamma and Asai 2017). They all show that, although the effect may not be as strong as in the case of surnames with /ta/, a voiced obstruent in N1 does make Rendaku application less likely in other surnames as well.

3. The current data set: Tanaka (2017)

Building on these studies, Tanaka (2017) built a large scale quantitative data set on the application of Rendaku in surname compounds. Tanaka created a new corpus of Japanese surnames by consulting social networking services, which resulted in a total of 1,181 instances of surnames. Through a detailed analysis of this corpus, he has shown that, like Rendaku in normal compounds, various factors affect the applicability of Rendaku. In this paper, we zoom in on one of his core findings—the partial blockage of Rendaku by a voiced obstruent in N1.

The essence of Tanaka's (2017) findings that are relevant to the current study can be summarized as follows:

(5) *Tanaka's (2017) findings*

a. If neither N1 nor N2 contains a voiced obstruent, Rendaku applies about 48% of the time.
b. A voiced obstruent in N2 categorically blocks Rendaku.
c. A voiced obstruent in N1 reduces Rendaku application to 10.3%.

In the corpus of Japanese surnames, Lyman's Law is categorical within a morpheme. More interestingly, it is gradiently active across a surname compound boundary. One way to characterize this observation is that a markedness restriction which holds in a smaller domain—Lyman's Law within a morpheme—"leaks" to a larger domain (Martin 2011).

The guiding intuition, which motivated our formal and quantitative analysis of the observations in (5), was that Japanese surnames consisting of N1 and N2 are only partially heteromorphemic, as also anticipated by Tanaka (2017) (cf. Hay and Baayen 2005). Semantically speaking, surnames are not at all compositional; for example, the last name *Kawahara* consists of *kawa* 'river' and *hara* 'field,' but the whole name has nothing to do with either river or field. Likewise, the last name *Tanaka* consists of *ta* 'rice field' and *naka* 'middle', but the semantics of these parts do not contribute to the meaning of the whole surname. The fact that these names consist of multiple morphemes becomes clear when we take their orthographic representations into account.

In what follows, we will analyze Tanaka's (2017) data of Rendaku in surnames and formalize this guiding intuition using the GSR coupled with Maximum Entropy Harmonic Grammar (Goldwater and Johnson 2003, Hayes and Wilson 2008), a happy marriage that was anticipated by Smolensky and Goldrick (2016).

4. The GSR analysis

Our analysis significantly relies upon the previous OT analysis of Rendaku developed by Ito and Mester (2003). Like Ito and Mester, we make use of three constraints that any constraint-based theory of Rendaku would need, which are shown in (6). Minimally, there has to be a constraint that triggers Rendaku (here, simply RENDAKU),[4] which must be ranked above the faithfulness constraint that prohibits unmotivated voicing (=IDENT-IO(VOICE)). Since RENDAKU does not create a morpheme that contains two voiced obstruents, RENDAKU has to be dominated by a constraint that prohibits this configuration: in this paper, we simply name the constraint LYMAN'S LAW.[5]

(6) *The three key constraints*

a. LYMAN'S LAW: Assign a violation mark for each morpheme that contains two voiced obstruents.
b. RENDAKU: Assign a violation mark for each voiceless obstruent at the beginning of a second member of a compound.
c. IDENT-IO(VOICE): Assign a violation mark for each underlying voiceless obstruent that surfaces as a voiced obstruent.

Crucially, a voiced obstruent in N1, according to this definition, does not block Rendaku, because the domain of the constraint LYMAN'S LAW is a morpheme. This definition works well for regular compounds, as Rendaku is not blocked by a voiced obstruent in N1, as shown by the examples in (3).

However, we would like to account for the (probabilistic) blockage of Rendaku by a voiced obstruent in N1 in surname compounds. We would thus need a mechanism in which a voiced obstruent in N1 partially violates LYMAN'S LAW. To formalize this intuition, one key innovation that we are proposing to the theory of the GSR is that a morpheme boundary

[4]This constraint can be understood as a constraint which requires phonological realization of a compound-linking morpheme [voice] (Ito and Mester 2003), or a morphologized version of a constraint that requires intervocalic voicing (Ito and Mester 1996).

[5]This constraint should likely be interpreted as a specific instance of the general OCP constraint family, OCP(VOICE) (Ito and Mester 1986). See also Alderete (1997) and Ito and Mester (2003) for an alternative formulation of this constraint based on self local conjunction of markedness constraints (Smolensky 1995).

can be partially activated.[6] As a consequence, the markedness constraint LYMAN'S LAW can be partially violated by a voiced obstruent in N1 in surname compounds.

In order to develop an analysis which can achieve a numeric fit to the observed pattern, we also make use of Maximum Entropy Harmonic Grammar. In this theory, constraints bear numerical weights, as opposed to being strictly ranked as in OT (Prince and Smolensky 1993/2004), and the grammar assigns a probability distribution over the candidate sets. See, for example, Hayes (2020) for detailed explication of the mathematics that derives these probability distributions.[7]

As stated above, the constraint set we deploy does not differ from that of Ito and Mester (2003). We instead "stochasticize" Ito and Mester's OT analysis by using numerically-weighted constraints. The optimum weights of the three constraints were calculated using the Solver function in Excel (Fylstra et al. 1998), so as to maximize the log-likelihood of the data with respect to this constraint set. We find positing that a morpheme boundary in surname compounds is 0.8 active, allowing LYMAN'S LAW to be violated 0.2 times, achieves an almost perfect fit with the observed data, as illustrated in the MaxEnt tableau in (2). The Excel analysis file that we used to develop this analysis is available in our osf depository (`https://osf.io/gpqy8/`). Comparing the two rightmost columns in (2), the values that our MaxEnt grammar predict are almost identical to the observed values taken from Tanaka's (2017) study.

[6] There are two departures that we make from the original proposal by Smolensky and Goldrick (2016). First, in Smolensky and Goldrick (2016), only faithfulness constraints were scaled accordingly to gradiently activated structures. Second, relatedly, in Smolensky and Goldrick (2016), it was underlying segments that are gradiently activated. We are claiming that surface structures can also involve gradiently activated structures, and accordingly, violations of markedness constraints can be scaled, too. See Faust and Smolensky (2017), Shih (2020), Walker (2020), and Zimmermann (2018, 2019), who posit that surface representations can involve partially activated structures.

[7] To review the MaxEnt mathematics briefly, the harmony score of each candidate (H) is the weighted sum of constraint violations: $H = \sum w_i C_i(x)$, where w_i is the weight of the i-th constraint and $C_i(x)$ represents how many times the i-th constraint is violated. The harmony score is negatively exponentiated (e^{-H}). The e^{-H} values for all the candidates are summed (i.e. $Z = \sum(e^{-H})_j$), and the predicted probability of each candidate x_j, $p(x_j)$, is $\frac{e^{-H}(x_j)}{Z}$.

		LYMAN'S LAW w=10.54	RENDAKU w=0.62	IDENT-IO w=0.68	H	e^{-H}	Z	*Pred.*	*Obs.*
/jama-saki/	jama-saki		1		0.62	0.54	1.04	51.41	51.42
	jama-zaki			1	0.68	0.51	1.04	48.59	48.58
/siba-saki/	siba-saki		1		0.62	0.54	0.60	89.71	89.70
	siba-zaki	0.2		1	2.79	0.06	0.60	10.29	10.30
/oo-sugi/	oo-sugi		1		0.62	0.54	0.54	100.00	100.00
	oo-zugi	1		1	11.22	0.00	0.54	0.00	0.00

Table 2: *The MaxEnt tableaux for the Rendaku patterns in surnames. The last two columns show the predicted values by the MaxEnt grammar and the observed values in Tanaka's (2017) study, respectively.*

A potential objection that can be raised against the analysis we have developed above is that we are modeling three data points with three constraints, which is in and of itself not very interesting. However, our model is not just about Japanese surnames; rather it is about general Japanese phonological grammar instead. For instance, the high weight of LYMAN'S LAW, with respect to RENDAKU, models the blockage of Rendaku in normal compounds. The high weight of LYMAN'S LAW, relative to IDENT-IO(VOICE), accounts for the lack of morphemes that contain two voiced obstruents in native words (Ito and Mester 1986). LYMAN'S LAW is also known to cause devoicing of geminates in loanwords (Kawahara 2015b, Nishimura 2006). With an important caveat that native words and borrowed words may exhibit different strengths in terms of faithfulness constraints (Ito and Mester 1995, 2008), our analysis can and should be interpreted as (a part of) a model of general Japanese phonology.

One natural question that arises at this point is the model's prediction regarding data in normal compounds, and how it fits with the observed data. Quantitative data on how often Rendaku applies when N1 contains a voiced obstruent is not readily available (cf. Ohta 2015). However, results from a nonce word experiment are available (Kawahara and Sano 2014), so we make use of their data to address this question. Their experiment consisted of three conditions: (1) bimoraic N1s with no voiced obstruents (no violations of Lyman's Law: e.g., [aka]), (2) bimoraic N1s with a voiced obstruent in the second mora (local violation of Lyman's Law at the moraic level across a morpheme boundary: e.g., [i**b**o]), and (3) bimoraic N1s with a voiced obstruent in the first mora (non-local violation of Lyman's Law: e.g., [**g**omi]). The experiment had seven items for each condition, and 36 native speakers of Japanese participated in the experiment. The participants were provided with a combination of N1 and N2, where N1s were existing words and N2s were nonce words. The participants were asked if each combination should undergo Rendaku or not. Figure 1 reproduces the results of their experiment, in which the y-axis represents the probability of Rendaku application in each condition.

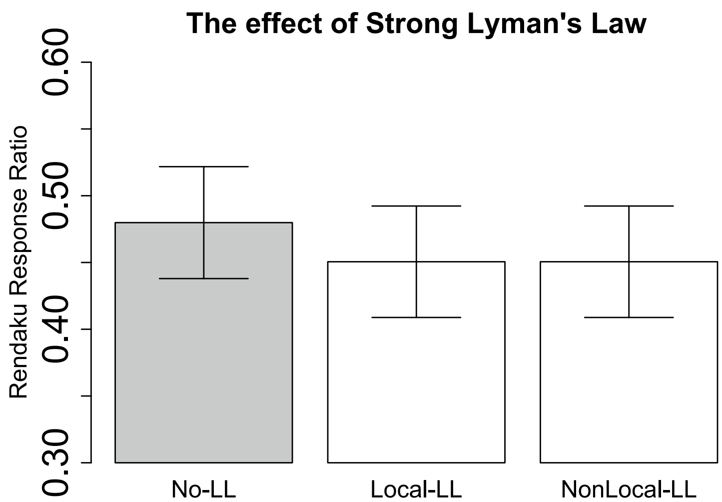

Figure 1: *Results of Kawahara & Sano (2014), their Figure 2. The y-axis represents the probability of Rendaku application, and the error bars represent the 95% confidence intervals. Reproduced here with permission.*

They found that the Rendaku percentage was about 48% in the first condition and about 45% in the second and third conditions. While the presence of a voiced obstruent in N1 did reduce the applicability of Rendaku, its magnitude was very small (about 3%). For the current purpose, we reanalyzed their data using a Bayesian mixed effects logistic regression model, which shows that the effect of a voiced obstruent in N1 is indeed not meaningful. The mean estimate of the slope coefficient is -0.07, and the Bayesian 95% Credible Interval is [-0.55, 0.43].[8] See the R markdown file available in the osf repository for complete details (`https://osf.io/gpqy8/`).

In summary, then, the Rendaku rates obtained in this experiment are comparable to Tanaka's (2017) surname data as well as the prediction of the GSR-MaxEnt analysis developed here: when Lyman's Law is completely irrelevant, either because there are no other voiced obstruents or because Lyman's Law is blocked by a fully activated morpheme boundary, Rendaku is expected about 45%~48% of the time.[9] Overall, we conclude that

[8]This result does not mean that the null effect of Lyman's Law is supported. In order for that to be the case, the 95% Credible Interval of the slope coefficient should be fully contained in the so-called ROPE (Region Of Practical Equivalence) of that point estimate, which is [-0.18, 0.18] in logistic regression models (Kruschke and Liddell 2018, Makowski et al. 2019). The number of participants is admittedly modest in that experiment, and the estimate may not be as accurate as desirable. With this caution in mind, we proceed by assuming that Lyman's Law did not meaningfully impact the Rendaku application rate across a morpheme boundary in that experiment.

[9]This analysis shows that, when Lyman's Law is irrelevant, the Rendaku application rates are comparable between surnames and in regular compounds. This result suggests that Rendaku application itself may not be affected by the semantic compositionality of compounds, although it could be that the compound-linking morpheme [voice] (Ito and Mester 2003), also, could have been activated to different degrees (cf. Rosen

the grammar we developed in this section is able to model general aspects of Japanese phonology, not just the Rendaku patterns in surnames.

5. Discussion

While our analysis successfully accounts for the Rendaku patterns in surnames—and the phonological properties of Japanese more generally—it also opens up several questions, which can be addressed only through more in-depth studies. We would like to discuss these ensuing questions briefly in this section.

First, one general question that arises from our results is how generally the patterns that Tanaka (2017) has identified hold in other domains of Japanese phonology. For example, it would be interesting to explore whether place names with a compound-like structure behave similarly with respect to the role of a voiced obstruent in N1.[10] Moreover, to the extent that gradient activation of morpheme boundaries has its roots in the semantic non-compositionality of surname compounds, as we speculated above, the current proposal predicts that regular compounds whose meanings are not very compositional should behave similarly. Quantifying degrees of semantic compositionality can be a challenge, but this is a topic that is worth some further attention in future research.

More generally speaking, we are also interested in the question of whether we would observe evidence for partially activated morpheme boundaries in patterns other than Rendaku in Japanese. Cases of exceptional prosodification patterns discussed by Hsu (2019) seem to be promising candidates. A recent study by Breiss et al. (to appear) also points to a potential case of partially activated morpheme boundaries. They studied patterns of velar nasalization in Japanese (Ito and Mester 1997), in which intervocalic [g] is nasalized into [ŋ]. Through a quantitative study of an NHK pronunciation and accent dictionary on the Yamanote dialect (NHK 2016), they show that the lexical frequencies of N1, quantified as differences in log-frequency between N1s and the whole compounds, negatively correlate with the application probability of velar nasalization. It looks as if the more frequent the N1, the less likely the whole compound behaves as "one chunk," resulting in a lower application rate of the intervocalic nasalization.[11]

As we stated at the outset of the paper, there is nothing in the theory of the GSR which prohibits morpheme boundaries from being partially activated—more careful cross-linguistic examination is warranted to test this prediction in natural languages.

One aspect of Rendaku which we abstracted away from our main analysis is the idiosyncratic lexical effects. Not all morphemes behave alike in terms of Rendaku: some

2016). It would be interesting to further explore which representations can and cannot be gradiently activated, as they relate to a single phonological phenomenon.

[10]Many Japanese surnames are actually derived from place names, so we expect that a similar pattern should hold, but a quantitative analysis is yet to be conducted.

[11]Brian Hsu (p.c.) pointed out an important difference between the velar nasalization pattern and the surname compound pattern analyzed in this paper. In the latter pattern, the compound boundaries appear to be "exceptionally weak" relative to normal compounds. In the veler nasalization pattern, it seems that N1s with higher lexical frequencies condition the boundary effects that are "stronger than expected." See Hsu (2019: 258-259) for relevant discussion on this difference. This is a good result, to the extent that the GSR is predicted to generate both types of patterns.

items are intrinsically more likely to undergo voicing than others. This observation led to the "lexicalist view of Rendaku" (Vance 2015), in which every piece of information about Rendaku application is stored in the lexicon, rather than Rendaku or Lyman's Law being governed by phonological grammar (though cf. Kawahara 2015a for a reply). Surnames also show idiosyncratic lexical effects as well (Tanaka 2017). See Rosen (2016) for a GSR-based approach to modelling such apparently idiosyncratic aspects of Rendaku in normal compounds, in which items can differ in how much they violate the RENDAKU constraint. We expect that a similar analysis can be extended to model idiosyncratic aspects of Rendaku in surname compounds. To the extent that this can be done, it would be interesting to investigate whether same morphemes behave similarly with respect to Rendaku in regular compounds and surname compounds. Our preliminary analysis of 22 morphemes that are commonly used as N2s in surnames shows that they indeed behave similarly in regular compounds and surname compounds (see the osf repository for this analysis), although further statistical exploration is needed to fully address this question.

Finally, there is an alternative analysis even within the theory of the GSR, namely positing categorical representations but using scaled constraints, as proposed by Hsu (2019) to account for exceptional prosodification effects observed in French liaison patterns. We can posit that compound boundaries in Japanese surnames are fully active, but they can scale a markedness violation of LYMAN'S LAW in such a way that they have a smaller scaling factor than regular compound boundaries. Recall that in our analysis we posited gradient representations and had a categorical constraint, whose violations are scaled according to the activity of gradient constraints. While both types of approach should be able to account for the current data set, these two theories make different predictions. The latter predicts that if there are other phenomena that are sensitive to morphological boundaries, they should be equally gradient. The first approach does not necessarily make that prediction (see Hsu and Jesney 2016). While we are unable to offer decisive evidence for one approach over the other here, the analysis presented here is more restrictive—a weakly activated morpheme should always be treated as weak by all phonological patterns relevant. Ultimately, these two alternatives should be examined from a cross-linguistic perspective against a wide range of patterns.

6. Conclusion

At the descriptive level, we have shown that Japanese surnames are ambiguous between being monomorphemic and heteromophemic (Tanaka 2017). Lyman's Law, a restriction that holds categorically within a morpheme, probabilistically holds across a morpheme boundary in Japanese surnames, i.e. Lyman's Law leaks across a surname compound boundary (Martin 2011). This is an intriguing observation given that Lymans's Law does not leak across a morpheme boundary in regular compounds (Vance 2015) or nonce words in a wug-experiment (Kawahara and Sano 2014). To account for this leakage effect specifically observed in surname compounds, we posited a partially activated morpheme boundary in surnames, which accords well with the semantic non-compositionality of such compounds. To the extent that our proposal is on the right track, it is possible for morpheme boundaries to be partially activated—and accordingly, markedness constraints can be partially violable—as expected from the general perspective of the GSR.

Appendix: Lyman's Law in Old Japanese

A potentially thorny challenge to the current proposal is how Lyman's Law behaved in Old Japanese—a voiced obstruent in N1 seems to have blocked Rendaku in Old Japanese (Ito and Mester 2003, Ramsey and Unger 1972, Unger 1975, Vance 2005). If we take our current proposal at its face value, it may mean, at first blush, that compound boundaries were not activated at all in Old Japanese. To us, this postulation sounds counter-intuitive. It is likely that compound junctures can get weaker as a result of diachronic changes, but it appears less likely that morpheme boundaries get stronger as a result of language changes.

But maybe this scenario is not entirely impossible—it could be the case that the development of the orthographic system, and its wide-spread use, made compound boundaries more salient. There is also evidence that speakers can actively assign morpheme boundaries to monomorphemic words, cases known as aggressive suffixation (Hammond 1999, Hayes and Jo 2020, Zuraw 2003). An alternative explanation is to posit that even a fully activated morpheme boundary did not—and perhaps does not—completely block the violation of LYMAN'S LAW, and its high weighting caused the blocking of Rendaku even across a fully activated morpheme boundary in Old Japanese.[12]

One easy way out of this potential challenge is to posit that there are two different versions of LYMAN'S LAW constraints, whose domain differ (Ito and Mester 2003): one constraint which is bound within a morpheme, and another constraint whose domain is a word. The one that is active in Old Japanese is the latter; in Modern Japanese, the former constraint is (almost) inviolable and the the latter constraint is probabilistically active only in surnames. We hesitate to take this route, because one strength of the current proposal is that it requires only one instance of LYMAN'S LAW constraint, and the domain restriction follows from different activation strengths of morpheme boundaries. More generally speaking, one advantage of deploying numerical weights, rather than a set of multiple constraints, is that it can simplify the constraint set (see Hsu 2019, Martin 2011, Smolensky and Goldrick 2016, Zuraw and Hayes 2017 for general discussion on this point). Allowing numeric weighting and expanding the constraint set at the same time does not seem to be desirable in terms of the restrictiveness of the theory.

References

Aarts, Bas. 2004. Conceptions of gradience in the history of linguistics. *Language Sciences 26*:343–389.

Alderete, John. 1997. Dissimilation as local conjunction. In *Proceedings of the North East Linguistics Society 27*, ed. by Kiyomi Kusumoto, 17–31. Amherst: GLSA.

[12]We should also perhaps bear in mind that voiced obstruents were pre-nasalized in Old Japanese, and that Lyman's Law was probably a restriction on two adjacent pre-nasalized consonants (Vance et al. 2021); therefore, direct comparison between Modern Japanese and Old Japanese is not as straightforward.

Breiss, Canaan, Hironori Katsuda, and Shigeto Kawahara. to appear. A quantitative study of voiced velar nasalization in Japanese. *Proceedings of Penn Linguistic Colloquium 45*.

Faust, Noam, and Paul Smolensky. 2017. Activity as an alternative to autosegmental association. Talk presented at Manchester Phonology Meeting 25.

Fylstra, Danial, Leon Lasdon, John Watson, and Allan Waren. 1998. Design and use of the Microsoft Excel Solver. *Interfaces* 28:29–55.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. *Proceedings of the Workshop on Variation within Optimality Theory* 111–120.

Guekguezian, Peter, and Karen Jesney. 2020. "Epenthetic" vowels are not all equal: Gradient representation in Yokuts roots and suffixes. Talk presented at WCCFL 38.

Hammond, Michael. 1999. English stress and cranberry morphs. Paper presented at the LSA.

Hay, Jennifer. 2003. *Causes and consequences of word structure*. New York & London: Routledge.

Hay, Jennifer, and Harald Baayen. 2005. Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences* 9:342–348.

Hayes, Bruce. 2020. Deriving the wug-shaped curve: A criterion for assessing formal theories of linguistic variation. Ms. UCLA.

Hayes, Bruce, and Jinyoung Jo. 2020. Balinese stem phonotactics and the subregularity hypothesis. Ms. UCLA.

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

Hsu, Brian. 2019. Exceptional prosodification effects revisited in Gradient Harmonic Grammar. *Phonology* 36:225–263.

Hsu, Brian, and Karen Jesney. 2016. Scalar positional markedness and faithfulness in Harmonic Grammar. *Proceedings of the Annual Meeting of the Chicago Linguistic Society* 51:241–255.

Hsu, Brian, and Karen Jesney. 2017. Loanword adaptation in Québec French: Evidence for weighted scalar constraints. *Proceedings of the 34th West Coast Conference on Formal Linguistics* 249–258.

Irwin, Mark. 2014. Rendaku across duplicate moras. *NINJAL Research Papers* 7:93–109.

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

Ito, Junko, and Armin Mester. 1995. The core-periphery structure of the lexicon and constraints on reranking. In *Papers in Optimality Theory*, ed. by Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, University of Massachusetts Occasional Papers 18, 181–210. Amherst: GLSA Publications.

Ito, Junko, and Armin Mester. 1996. Rendaku I: Constraint conjunction and the OCP. Ms. University of California, Santa Cruz.

Ito, Junko, and Armin Mester. 1997. Correspondence and compositionality: The ga-gyo variation in Japanese phonology. In *Derivations and constraints in phonology*, ed. by Iggy Roca, 419–462. Oxford: Oxford University Press.

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

Ito, Junko, and Armin Mester. 2008. Lexical classes in phonology. In *The Oxford handbook of Japanese linguistics*, ed. by Shigeru Miyagawa and Mamoru Saito, 84–106. Oxford: Oxford University Press.

Kawahara, Shigeto. 2015a. Can we use rendaku for phonological argumentation? *Linguistic Vanguard* 1:3–14.

Kawahara, Shigeto. 2015b. Geminate devoicing in Japanese loanwords: Theoretical and experimental investigations. *Language and Linguistic Compass* 9:168–182.

Kawahara, Shigeto, and Shin-ichiro Sano. 2014. Testing Rosen's Rule and Strong Lyman's Law. *NINJAL Research Papers* 7:111–120.

Kruschke, John K., and 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 *Voicing in Japanese*, ed. by Jeroen van de Weijer, Kensuke Nanjo, and Tetsuo Nishihara, 5–24. Berlin & New York: Mouton de Gruyter.

Lakoff, George. 1973. Fuzzy grammar and the performance/competence terminology game. *Papers from the Regional Meeting of the Chicago Linguistic Society* 9:271–291.

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

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

NHK. 2016. *NHK Japanese pronunciation accent dictionary*. Broadcasting Culture Research Institute.

Nishimura, Kohei. 2006. Lyman's Law in loanwords. *Phonological Studies* 9:83–90.

Ohta, Shinri. 2015. Effects of phonological and semantic factors on rendaku: A study using the rendaku database and logistic regression analyses [in Japanese]. Ms. University of Tokyo.

Prince, Alan, and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden and Oxford: Blackwell.

Ramsey, S. Robert, and J. Marshall Unger. 1972. Evidence for a consonant shift in 7th century Japanese. *Papers in Japanese Linguistics* 1:278–295.

Rosen, Eric. 2016. Predicting the unpredictable: Capturing the apparent semi-regularity of rendaku voicing in Japanese through harmonic grammar. *Proceedings of BLS* 235–249.

Rosen, Eric. 2018. Predicting semi-regular patterns in morphologically complex words. *Linguistic Vanguard* 4.

Ross, John R. 1972. The category squish: endstation hauptwort. *Papers from the Regional Meeting of the Chicago Linguistic Society* 8:316–328.

Shih, Stephanie. 2020. Gradient categories in lexically-conditioned phonology: An example from sound symbolism. *Proceedings of the 2019 Annual Meeting on Phonology*.

Smolensky, Paul. 1995. On the internal structure of the constraint component CON of UG. Talk presented at the University of California, Los Angeles.

Smolensky, Paul, and Matt Goldrick. 2016. Gradient symbolic representations in grammar: The case of French liaison. Ms. Johns Hopkins University and Northwestern University.

Smolensky, Paul, Eric Rosen, and Matthew Goldrick. 2020. Learning a gradient grammar of French liaison. *Proceedings of the 2019 Annual Meeting on Phonology*.

Sugito, Miyoko. 1965. Shibata-san to Imada-san: Tango-no chookakuteki benbetsu-ni tsuite-no ichi koosatsu [Mr. Shibata and Mr. Imada: An analysis of auditory distinction of words]. *Gengo Seikatsu* 165:64–72.

Tanaka, Yu. 2017. The sound pattern of Japanese surnames. Doctoral dissertation, University of California, Los Angeles.

Unger, Marshall. 1975. Studies in early Japanese morphophonemics. Doctoral dissertation, Yale University.

Vance, Timothy. 2005. Sequential voicing and Lyman's Law in Old Japanese. In *Polymorphous linguistics: Jim Mccawley's legacy.*, ed. by Salikoko S. Mufwene, Elaine J. Francis, and Rebecca S. Wheeler, 27–43. Cambridge: MIT Press.

Vance, Timothy. 2015. Rendaku. In *The handbook of Japanese language and linguistics: Phonetics and phonology*, ed. by Haruo Kubozono, 397–441. Berlin: Mouton de Gruyter.

Vance, Timothy, and Mark Irwin, ed. by. 2016. *Sequential voicing in Japanese compounds: Papers from the NINJAL Rendaku Project*. Amsterdam: John Benjamins.

Vance, Timothy, Shigeto Kawahara, and Mizuki Miyashita. 2021. The diachronic origins on Lyman's Law: Evidence from phonetics, dialectology, and philology. Manuscript.

Walker, Rachel. 2020. Gradient activity in Korean place assimilation. *Proceedings of NELS* 50:207–220.

Zamma, Hideki. 2005. Correlation between accentuation and rendaku in Japanese surnames: A morphological account. In *Voicing in Japanese.*, ed. by Jeroen van de Weijer, Kensuke Nanjoo, and Tetsuo Nishihara, 157–176. Berlin & New York: Mouton de Gruyter.

Zamma, Hideki, and Atsushi Asai. 2017. Sei-ni mirareru Sugitoo-no hoosoku-to kakuchoo-ban Lyman-no hoosoku-ni kansuru keitai-teki/onin-teki koosatsu [Morphological and phonological analyses of Sugito's law and the extended version of Lyman's Law in surnames]. In *Rendaku-no Kenkyuu [Research on Rendaku]*, ed. by Timothy J. Vance, Emiko Kaneko, and Seiji Watanabe, 147–179. Tokyo: Kaitakusha.

Zimmermann, Eva. 2018. Gradient symbolic representations in the output: A case study from Moses Columbian Salishan stress. *Proceedings of NELS* 275–284.

Zimmermann, Eva. 2019. Gradient symbolic representations and the typology of ghost segments. *Proceedings of AMP 2018*.

Zimmermann, Eva. 2020. Stronger and thus more beautiful: The phonological strength of templates. Talk presented at OCP 2020.

Zuraw, Kie. 2003. Aggressive reduplication. *Phonology* 19:395–439.

Zuraw, Kie, and Bruce Hayes. 2017. Intersecting constraint families: An argument for Harmonic Grammar. *Language* 93:497–548.

Shigeto Kawahara and Yu Tanaka
kawahara@icl.keio.ac.jp and yutanak@mail.doshisha.ac.jp