*Final draft; February 12, 2023
Tanaka, Yu (to appear) Learning biases in proper nouns. *Phonology*.

*Learning biases in proper nouns**

Yu Tanaka
Doshisha University

It has been proposed that there are cognitive biases in language learning that favour certain patterns over others. This study examines the effects of such bias factors on the learning of the phonology of proper nouns. I take up the phenomenon of compound voicing in Japanese surnames. The results of two judgment experiments show that, while Japanese speakers replicate various kinds of statistical regularities in existing names, they tend to extend only phonologically motivated patterns to novel names. This suggests that phonological naturalness plays a role even in the learning of a highly faithful category of words, namely proper nouns, and provides evidence for the relevance of learning biases in synchronic grammar.

1 Introduction

Some sound patterns are more commonly attested than others in languages of the world. One of the factors that has been proposed to account for such cross-linguistic tendencies is the notion of bias. Some researchers argue that language learners are biased for certain phonological properties. For example, patterns are easier to learn if they are structurally simple (e.g. Pycha *et al.* 2003; Moreton 2008, 2012; see Moreton & Pater 2012a for a review) or grounded in phonetic principles such as ease of articulation and perception (e.g. Wilson 2006; Myers & Padgett 2014; Martin & Peperkamp 2020; see Moreton & Pater 2012b for a review; Donegan & Stampe 1979; Hayes 1999 for general discussion). This kind of cognitive bias with respect to language learning can be referred to as 'analytic bias'. Others argue that bias lies rather

* E-mail: yutanak@mail.doshisha.ac.jp
I am particularly grateful to Bruce Hayes, Junko Ito, Megha Sundara, Kie Zuraw, and P-side members of the UCLA Linguistics Department for discussion at earlier stages of this research project. I also thank three anonymous reviewers and the editorial team at *Phonology* for their comments and suggestions, which have greatly helped improve this paper. The work has been supported by Kakenhi grants from the Japan Society for the Promotion of Science (20K13019; 22K13106).

in the way language sounds are transmitted. Speech signals can be systematically misperceived due to factors such as coarticulation on the part of the speaker and compensation for it on the part of the hearer. This can eventually lead to changes in phonological representations, often in the direction of phonetic naturalness (Ohala 1993; Blevins 2004; see Hansson 2008; Garrett & Johnson 2013 for reviews). This kind of bias that may cause transmission errors can be called 'channel bias'.

Note that most bias-based accounts, whether the biasing factors are assumed to be in speech transmission or in language learning, do not make an extreme argument that all sound patterns should be natural. An account based on channel bias, for example, may attribute the emergence of unnaturalness to historical quirks. An unnatural pattern may be created by a sequence of natural sound changes (Kenstowicz & Kisseberth 1977:64–65; Beguš 2018) or external factors such as language contact (Blevins 2017). An account based on analytic bias may also allow for the existence of unnaturalness. Natural patterns are favoured in language learning; yet those that do not fit the description are still learnable if there are sufficient input data (e.g. Hayes *et al.* 2009; White 2014; see Hayes 1999 for discussion).

Although the two approaches are not mutually exclusive (Moreton 2008; Beguš 2018), they do make different predictions as to how certain types of sound patterns are learned. Suppose that there is a phonological pattern that is structurally complex and has no phonetic motivation. Suppose also that there is another pattern that is structurally simpler and phonetically better motivated. An account based on analytic bias predicts that a learner will find the former more difficult to learn than the latter. An account based on channel bias, on the other hand, does not necessarily predict that there will be a difference in the learning of those patterns; the two patterns can be learned equally well as long as there is no difference in their phonetic precursors, which would lead to transmission errors between the speaker and the hearer (see Moreton 2008 for relevant discussion).

Along these lines, some previous studies have tried to tease apart the two approaches via artificial language learning experiments (see Moreton & Pater 2012a,b for reviews). Others have addressed the same question by conducting so-called 'surfeit-of-the-stimulus' experiments (Becker *et al.* 2011), which use data based on real language phonology. As stated above, a language may contain natural and unnatural patterns, and speakers of the language are exposed to both of them as statistical regularities. Researchers can investigate whether and if so how they generalise those patterns to novel items. If it turned out that speakers show preferences for natural patterns over unnatural ones in nonce word tasks, despite the same amount of experience with both kinds in real words, this would suggest that their grammar is biased for naturalness. A number of experimental studies have shown that patterns that 'phonologically make sense' are more readily reproduced or judged more acceptable than arbitrary patterns, even though their participants must have received data for both of them in their ambient languages (e.g. Hayes *et al.* 2009; Zhang & Lai 2010; Becker *et al.* 2011, 2012; Hayes & White 2013).

The current study aims to address the issues of learning biases through surfeit-of-the-stimulus experiments, making use of proper nouns as primary data. Names in general often show aberrant sound patterns that differ from those of other words (see Broad *et al.* 2015; Moreton *et al.* 2017; Jaber & Omari 2018 and references therein). They also tend to be highly lexicalised, and retain archaic characteristics that are no longer seen elsewhere in the language. Theoretically, these facts can be explained in terms of category-specific privilege effects (Smith 2011). Proper nouns constitute an independent category, and faithfulness constraints indexed to that category are inherently ranked high in grammar (Moreton *et al.* 2017; Jaber & Omari 2018). This does not mean, however, that proper nouns are simply exempt from all sorts of grammatical operations. Phonological processes may still apply productively in newly coined names (e.g. Moreton *et al.* 2017). One can then ask how the sound patterns of names are learned and extended to novel names. Studying the phonology of a highly faithful category of words and investigating what processes are learned, or underlearned, may provide a new window on the role of learning biases in phonology.

Compound voicing, known as *rendaku*, in Japanese surnames is a good case study subject in this respect. Rendaku in compound surnames exhibits peculiar patterns compared to its application in common noun compounds. Some of these patterns are difficult to define in terms of phonological features and lack clear phonetic motivations. Most of these peculiarities turn out to have roots in the sound patterns of Old Japanese, suggesting that surnames may simply retain archaic traits of the language. However, there is also evidence that rendaku in surnames is productive; the voicing alternation occurs in novel names. This suggests that Japanese speakers are exposed to surnames, somehow learn the rendaku application patterns in them, and extend those patterns to surnames they have not seen before. Questions to be addressed here are (i) what kinds of phonological patterns are found in existing names, and (ii) which among those are well learned and generalised to novel names. If phonologically natural patterns are preferred over unnatural ones even in the learning of proper nouns, which are generally tolerant of idiosyncrasy, it will serve as yet another kind of evidence for the effects of analytic biases.

The paper is organised as follows. §2 gives descriptions of the rendaku patterns in Japanese surnames. §3 proposes possible accounts of the phenomenon and discusses how they relate to the issues of learning biases. In §4 and §5, I report the results of rendaku judgment experiments using real and nonce name stimuli. §6 gives a general discussion and §7 concludes the paper.

2 Rendaku in surnames

2.1 Background: Rendaku

Rendaku is a process in Japanese whereby the initial obstruent of the second member of a morphologically complex word becomes voiced, as shown in (1) below.

(1) Compound formation involving rendaku[1]

a.	maki	+	susi	→	maki-**z**usi	'rolled-sushi'
b.	ken	+	tama	→	ken-**d**ama	'sword-ball'
c.	ori	+	kami	→	ori-**g**ami	'folding-paper'
d.	ike	+	hana	→	ike-**b**ana	'enlivening-flower'

As can be seen, if the second element of a compound (henceforth referred to as 'E2') starts with a voiceless obstruent such as /s/, /t/, /k/, or /h/, the consonant becomes voiced as a result of compounding. Note that /h/ alternates with /b/ for historical reasons (Ueda 1898; Miyake 2003:66–77).

Rendaku is intrinsically variable, and its applicability is affected by a number of phonological and non-phonological factors (see Vance 2015a; Kawahara & Zamma 2016 for overviews). One such factor is lexical idiosyncrasy. Rosen (2001) shows that some morphemes very often undergo voicing when they appear as E2 of a compound, hence he calls them 'rendaku lovers'; there are other morphemes dubbed 'rendaku haters' that typically do not voice, and 'rendaku immune morphemes' that in fact never voice (Rosen 2001; Irwin 2016a; also see Rosen 2016 for additional observations and analysis.)

Not only E2 but compound words themselves may show idiosyncratic behaviours. There are cases where the exact same E2, whether a lover or hater, undergoes rendaku in one compound but not another, as shown in (2).

(2) Irregularity of rendaku in regular compounds

a.	kage	+	kuti	→	kage-**g**uti	'behind-mouth (gossip)'
b.	haja	+	kuti	→	haja-**k**uti	'fast-mouth (fast speech)'
c.	waru	+	kuti	→	waru-**k**uti ∼ waru-**g**uti	'bad-mouth (slander)'

Rendaku is also seen in proper nouns. Many Japanese surnames are compounds; about 96% of the 10,019 most common surnames are composed of multiple morphemes (see Shirooka & Murayama 2011). A compound surname qualifies for rendaku application if its E2 starts with a voiceless obstruent. As in regular compounds, rendaku in surnames is not an iron-clad rule. Certain names typically show voicing while others do not, and there are also some that oscillate, as shown in (3).[2]

(3) Compound surnames and rendaku

a.	oka	+	ta	→	oka-**d**a	岡田	'hill-paddy'
b.	saka	+	ta	→	saka-**t**a	坂田	'slope-paddy'
c.	naka	+	ta	→	naka-**t**a ∼ naka-**d**a	中田	'central-paddy'

It is worth noting here that rendaku in surnames is usually not reflected in orthography. By convention, surnames are written in Chinese characters, or

[1] Throughout the paper, I use broad transcriptions largely based on the *kunrei* romanisation system in Japanese, except that I transcribe the palatal approximant with /j/ and vowel length with /ː/ as in IPA.

[2] 'Oscillation' here does not mean that one single person's surname can be pronounced either with or without rendaku at each utterance, but rather that we find people who have the rendaku reading and others who have the non-rendaku reading, even though their surnames contain the same morphemes and are written with the same Chinese characters.

kanji, which do not indicate voicing resulting from rendaku. Notice that in (3), the E2 morpheme /ta/ 'paddy' is always written with the same kanji 田, whether it is realised as [ta] or [da]. Thus, when speakers encounter a surname in written form, especially an unknown one, they need to make a judgment on whether it should be pronounced with or without voicing.

The literature shows that rendaku in surnames exhibits some peculiarities. In what follows, I describe the patterns of voicing in compound surnames, highlighting their differences from those in regular compounds.

2.2 Strong Lyman's Law

Previous studies have shown that rendaku application in surnames, though inherently variable, is predictable to some extent. Sugito (1965) points out that voicing is conditioned by the onset consonant of the final syllable of the first element (henceforth 'E1'). As shown in (4a), if the consonant in question (underlined) is /s/, /t/, /k/, /m/, or /n/, rendaku is commonly observed. Contrariwise, if it is a voiced obstruent, rendaku never applies, as in (4b).

(4) Voiced obstruents in E1 block rendaku in surnames

a.	asa-**da**	matu-**da**	take-**da**	jama-**da**	sana-**da**
b.	kazi-ta	siba-ta	kado-ta	naga-ta	

Although Sugito's (1965) focus was limited to surnames with /ta/ 'paddy' as E2, subsequent studies (Kubozono 2005; Zamma 2005; Tanaka 2017; Zamma & Asai 2017, among others) have further shown that the generalisation holds mostly true for many other names.

This blocking of rendaku by a voiced obstruent in E1 may, at first glance, seem to be the effect of oft-cited Lyman's Law (Lyman 1894; also see Motoori 1790-1822), namely a ban on multiple voiced obstruents. The law is understood as a general morpheme structure constraint against the co-occurrence of voiced obstruents (see Morita 1977; Ito & Mester 1986; Yamaguchi 1988). One may thus think that if there is already a voiced obstruent in E1, rendaku should be blocked, as its application would create two voiced obstruents within the name.

Although the description itself may be on the right track, it raises the issue of the domain of Lyman's Law. Under the most widely accepted definition, the law stipulates that there may be no multiple voiced obstruents within a single stem of Yamato (native) Japanese origin. In other words, the law is bounded by the 'stem' and not by the 'prosodic word' (Ito & Mester 2003). In regular compounds, rendaku is categorically blocked if there is already a voiced obstruent in 'E2', as shown in (5a). A voiced obstruent in 'E1', on the other hand, does not necessarily inhibit rendaku, as in (5b).

(5) Only voiced obstruents in E2 block rendaku in regular compounds

a.	kuro	+	kabi	→	kuro-kabi,	*kuro-**g**abi	'black-mold'
cf. b.	kabe	+	kami	→	kabe-**g**ami		'wall-paper'

Even though examples like (5b) are found, it has been suggested that voiced obstruents in E1 still dampen compound voicing (e.g. Kindaichi *et al.* 1988:264; Sato 1989; Labrune 2012:120–121). This alleged restriction on multiple voiced obstruents at the word level is often referred to as 'Strong Lyman's Law', as opposed to the stem-bounded version of the law.

The validity of Strong Lyman's Law has been the subject of controversy, however, and the evidence supporting it is mixed. Irwin (2014, 2016a) and Sano (2015) investigated different corpora of regular compounds, and both argue that its effect, if any, is negligible. Ohta's (2015) corpus study employed the same data source as Irwin (2014) but different analytical methods. He suggests that Strong Lyman's Law is partially active, reporting that a subset of voiced obstruents in E1 do lower rendaku applicability. Asai (2014) also found some weak effects of the word-bounded law in his original data of compounds compiled from magazines. It should be noted, however, that his corpus included both regular words and proper nouns (person and place names), and the results may not be suitable for our purposes. Kawahara & Sano (2014b) tested the psychological reality of Strong Lyman's Law experimentally, and found null results. In fact, when Lyman (1894) first described what would be later known as Lyman's Law, he also made it clear that a voiced obstruent in the final syllable of E1 does not affect rendaku, based on his own observation of words in a Japanese-English dictionary.

Part of the problem here is that Strong Lyman's Law has never been formally defined, and that there seem to be different views on what counts as an active constraint. Given what has been reported in the literature, a plausible interpretation would be that Lyman's Law, which is usually stem-bounded, also shows some gradient or probabilistic effects on the word level. It is known that phonotactic restrictions that have a categorical effect within a smaller domain (e.g. stem) may also have a weaker effect in, or 'leak into', a bigger domain (e.g. word or sentence) in languages (Martin 2011; Breiss & Hayes 2020). Strong Lyman's Law might then be another case of leakage of stem-internal phonotactics.

The question still arises as to why the law's effects differ in surnames and normal words. As stated above, Strong Lyman's Law is very much in force in compound surnames. In regular compounds, it is not strictly enforced, with only weaker effects at best.

2.3 Identity and Similarity Avoidance

Previous studies have revealed another factor affecting rendaku in surnames. Tanaka (2017) and Zamma & Asai (2017) show that avoidance of consonantal identity or similarity promotes voicing. That is, rendaku is more likely to occur when a compound surname underlyingly has a sequence of homorganic voiceless obstruents in E1 and E2. In such cases, application of compound voicing can yield the dissimilation of identical consonants (e.g. /s-s, t-t, k-k/), which I call Identity Avoidance (Yip 1998), or the dissimilation of similar consonants more broadly (e.g. /s-t, t-s/), which I call Similarity Avoidance (see Frisch 2004; Frisch *et al.* 2004). Examples are given in (6).

(6) Homorganic voiceless obstruents promote rendaku in surnames

a.	matu	+	ta	→	matu-**d**a	*matu-ta	'pine-paddy'
b.	nisi	+	sima	→	nisi-**z**ima	*nisi-sima	'West-island'
c.	naka	+	kawa	→	naka-**g**awa	*naka-kawa	'central-river'
d.	asa	+	ta	→	asa-**d**a	*asa-ta	'shallow-paddy'

Research to date has not fully confirmed that Identity Avoidance and Similarity Avoidance play any active roles as rendaku triggers in regular compounds. Most corpus studies of rendaku do not devote particular attention to the issue (Irwin 2014, 2016a; Ohta 2015; Sano 2015; cf. Asai 2014). Other studies only discuss identity in a narrower sense. Toda (1988) investigates compound words in Early Modern Japanese, and states that rendaku is more likely to occur if there are 'identical moras' across the E1-E2 boundary (e.g. /aka-kami/ → [aka-**g**ami] 'red-paper'). Kawahara & Sano (2014b) conduct a nonce-word judgment experiment, and show that voicing change is more acceptable if it helps dissimilate 'identical moras' (e.g. /ika-kaniro/ → [ika-**g**aniro] 'squid-nonce'). These studies, however, do not specifically test the question of whether Identity Avoidance and Similarity Avoidance at the 'consonantal level' promote rendaku in regular compounds.[3]

2.4 Other patterns: Sonorants in E1

Rendaku in surnames shows further complications. Kubozono (2005) observes that, in Sugito's (1965) data of surnames with /ta/ as E2, the presence of /r/ in E1's final syllable tends to block rendaku. Many surnames with E1-/r/ indeed resist rendaku, as shown in (7a), with just a few exceptions, as in (7b).

(7) /r/ in E1 inhibits rendaku in surnames with E2-/ta/

a.	nari-**t**a	kuri-**t**a	kura-**t**a	mura-**t**a	mori-**t**a	ari-**t**a	...
b.	hara-**d**a	kuro-**d**a	tera-**d**a				

Later studies have confirmed that this rendaku-inhibiting effect of E1-/r/ is also found in surnames with other E2 morphemes (Asai 2014; Zamma & Asai 2017).

There has been a claim that /r/ in E1's final syllable acts as a rendaku blocker in regular compounds as well (Hirano 2013 cited by Irwin 2016a:97; also see Asai 2014; Vance & Asai 2016; Asai & Vance 2017). Irwin (2016a) calls this into question, however, noting that the rendaku rates of words with E1-/r/ are only slightly lower than expected in his large corpus data. Toda (1988) also states impressionistically that E1-/r/ dampens rendaku somewhat, but not considerably, in Early Modern Japanese. Lastly, Asai (2014) reports that /r/ in E1 does inhibit rendaku in his magazine-based corpus, noting that the effect is weak but statistically significant. As stated above, however, his

[3] It has also been proposed that avoidance of consonantal identity may function as a 'rendaku blocker' when combined with Strong Lyman's Law (e.g. Sato 1989; Takayama 1992). See Irwin (2014, 2016a); Kawahara & Sano (2014a, 2016) for discussion and complications.

data include proper nouns, and thus the result should be taken with a grain of salt.

The nasals /n/ and /m/ also pose puzzles. It has been argued that onset nasals in E1-final syllables promote rendaku in surnames. As shown in (4) above, Sugito (1965) classifies them as consonants that trigger voicing along with voiceless obstruents. Zamma (2005) and Zamma & Asai (2017) report that the rendaku-promoting effects of nasals are also found in surnames with some E2 morphemes other than /ta/ ‘paddy’, even though they are not as robust. They further state that the two nasal segments show different behaviours; /n/ triggers rendaku more than /m/. Putting these observations together, they conclude that nasals in E1 promote rendaku only in surnames with certain E2 morphemes, and that /m/’s effect is weaker than /n/’s.

Do we find the same or similar patterns in regular compounds? The literature does not offer a definitive answer. Irwin’s (2016a) corpus study reports the average rendaku rates of all compounds (.769), as well as those of compounds with E1-/n/ (.765) and E1-/m/ (.726), but no statistical test is conducted to specifically test their differences. Ohta (2015) shows that some specific cases of /m/ in ‘E2’ make rendaku application less likely (see §3.2 for details), but does not include /n/ nor /m/ in ‘E1’ in his list of the factors that significantly affect rendaku, much less mention the difference between the two.

These facts about sonorants again highlight the peculiarities of rendaku in surnames. Several studies report that the type of the last sonorant in E1 affects rendaku applicability in compound surnames, but the literature has not reached a consensus as to whether the same effects are found in regular compounds.

2.5 Summary of rendaku in surnames

Segments	Effects
Voiced obstruent: /z, b, d, g/	inhibition (Strong Lyman’s Law)
Homorganic pair: /s-s, t-t, k-k, s-t, t-s/	promotion (Identity/Similarity Avoidance)
Liquid: /r/	inhibition
Nasal: /n, m/	promotion; but less by /m/

Table I
Effects on rendaku in surnames

Table I summarises the characteristics of rendaku in surnames based on the findings of previous studies. In the next section, I lay out two possible accounts of these patterns: one based on naturalness and the other based on diachrony. I then discuss how they can be relevant to the issues of learning biases.

3 Naturalness or diachrony?

3.1 A naturalness account

One possible way of accounting for rendaku in Japanese surnames is to attribute all of the patterns to principles based on phonological naturalness. Here, I use the term 'naturalness' in a broad sense, defining it from two perspectives: phonetic substance and structural complexity. A sound pattern is said to be natural in terms of phonetic substance, if it is motivated by the ease of articulatory or perceptual difficulties (Donegan & Stampe 1979; Hayes 1999; also see Hayes *et al.* 2004 and papers therein). A pattern is also considered natural from the structural point of view, if it can be easily defined without dependencies of distinctive features across multiple dimensions (see Moreton 2008, 2012; Moreton & Pater 2012a). I discuss below whether each of the rendaku patterns in Japanese surnames deserves such a naturalness-based explanation.

Strong Lyman's Law should be considered natural in terms of structural complexity. Formally, it is a segmental version of the Obligatory Contour Principle (OCP; McCarthy 1986, 1988); it simply bans multiple cases of [–sonorant, +voice] segments within a word (Ito & Mester 1986, 2003), without involving inter-dimensional dependencies (Moreton 2012). Meanwhile, the law is unnatural in its phonetic grounding (Kawahara 2008). According to Ohala (1981), dissimilation stems from perceptual errors of some kind; phonetic features that spread out across segments can cause perceptual confusions, which can in turn lead to a dissimilatory sound change due to hypercorrection by listeners. Ohala (1981, 1993) specifically claims that a voicing feature, whose main phonetic correlates do not typically stretch out, should not be subject to dissimilation, and that cases of synchronic voicing dissimilation must have originated from co-occurrence restrictions against features other than voicing (see Kawahara 2008; Vance *et al.* 2021; §3.2 for the origin of Lyman's Law). Taken together, Strong Lyman's Law is structurally simple but phonetically unmotivated.

The patterns of Identity Avoidance and Similarity Avoidance discussed in this study are natural on the grounds of both formal structure and phonetic substance. They are instances of dissimilation that are particularly concerned with place and manner, and can be formalised as versions of OCP targeting a sequence of total or partial identity (McCarthy 1986; Yip 1998; Frisch *et al.* 2004), without involving complex feature dependencies. A number of studies have also claimed that the principles have a functional motivation. Repetition of consonants that share place and manner features tends to induce errors in production and perception, which may stem from difficulties in language processing (see Frisch 2004; Alderete & Frisch 2007 and references therein). Avoidance of consonantal identity or similarity is thus well motivated in that it reduces those difficulties. The effects are seen as both static and dynamic patterns in many languages (see Suzuki 1998; Alderete & Frisch 2007 for overviews), including Japanese (Kawahara *et al.* 2006).

In contrast, the behaviour of /r/ does not seem very straightforward for a naturalness account. Recall that /r/ in E1 prevents rendaku in surnames.

To formulate a constraint specifically banning a sequence of /r/ and a voiced obstruent, which I tentatively dub *r...D for short, one would have to refer to not only the feature [voice] but also other features with disagreeing values. Such a constraint would not count as a normal OCP-type constraint, and it would involve relatively complex feature dependencies (see Moreton 2008, 2012). Some studies of rendaku in surnames have entertained the idea of treating /r/ as a voiced obstruent (see Kubozono 2005; Zamma 2005), but the proposal seems ad hoc unless there is supporting evidence outside this particular phenomenon. Furthermore, the constraint would look unnatural from the phonetic point of view; a sequence of /r/ and a voiced obstruent does not seem to pose any particular difficulty in production or perception.

That said, one could still possibly resort to dissimilation as the phonetic basis of *r...D. Japanese /r/ is most typically realised as an alveolar tap or flap, or some other close variant, depending on the context (Tsuzuki & Lee 1992; Vance 2008:89; Labrune 2014; Katz *et al.* 2018). These variants must be similar to the phonetic realisations of the coronal voiced obstruents /d/ and /z/. Indeed, in some dialects of Japanese, /d/, /z/, and /r/ are often misperceived as one another, leading to near phonemic mergers (see Sugito 1982 and work cited there). It could then be that a surname with E1-/r/ resists rendaku so as not to create a sequence of phonetically similar consonants, such as [r...d] and [r...z]. Note, however, that this explanation makes a particular prediction: /r/ in E1 should block rendaku when E2's initial consonant is /t/ or /s/, which would become [d] or [z] through voicing, but not when it is /k/ or /h/. Since previous studies have not considered this particular hypothesis, the possible phonetic grounding of /r/'s behaviour remains to be tested.

The effects of nasals are also hard to interpret in terms of naturalness. Again, an onset nasal in E1's final syllable is argued to cause voicing of E2's initial obstruent (e.g. /kane-ta/ → kane-**d**a 'gold-paddy'). The pattern is not to be confused with so-called post-nasal voicing, where a coda nasal voices the immediately following obstruent. Post-nasal voicing is cross-linguistically common and arguably has a phonetic motivation (e.g. Pater 1999; Hayes & Stivers 2000; also see Ito & Mester 1986 for its effect in Japanese). However, non-local voicing by an onset nasal seems rare and unmotivated.[4] Nasality itself is indeed a feature that may spread across segments, but it is unclear how it directly affects the voicing of a non-adjacent obstruent (see Ohala 1981, 1993). The fact that /m/ triggers rendaku less than /n/ also seems to have no phonetic basis. Lastly, formalisation of non-local post-nasal voicing would involve complex feature dependencies, as in the case of /r/, suggesting that the pattern is also unnatural in structural terms.

We have seen so far that rendaku in surnames poses challenges to a naturalness account. The discussion is summarised in Table II, which shows whether each pattern can be considered natural (✓) or unnatural (✗) from

[4] One possibly related phenomenon is found in Japanese. Kindaichi (1976/2005) reports that the adjectival and adverbial composite suffixes /-(a)si-i/ and /-(a)si-ku/ undergo voicing after a non-local nasal (e.g. [susam-a**z**i-i] 'tremendous'). However, there are many exceptions to this generalisation, and to the best of my knowledge, no other suffixes show such patterns.

the viewpoints of formal structure and phonetic substance. A question mark means that it is still unclear with the evidence at hand.[5]

Segments	Effect	Structure	Substance
Voiced obstruent: /z, b, d, g/	inhibition (Strong Lyman's Law)	✓	✗
Homorganic pair: /s-s, t-t, k-k, s-t, t-s/	promotion (Identity/Similarity Avoidance)	✓	✓
Liquid: /r/	inhibition	✗	✗ ?
Nasal: /n, m/	promotion; but less by /m/	✗	✗

Table II
Rendaku in surnames viewed from naturalness

3.2 A historical account

One can also seek a different kind of explanation and claim that the whole phenomenon of rendaku in surnames should be understood as a diachronic problem. Historical studies suggest that the sound patterns of surnames we see today actually have a number of similarities with patterns seen in earlier stages of the Japanese language.

Recall that Lyman's Law is active on the word level in current surnames. Historically, the domain of the law was in fact the prosodic word even in regular compounds (see Unger 1977). In Old Japanese (early 7c.–late 8c.), rendaku was blocked if either E1 or E2 contained /z/, /b/, /d/, or /g/, as it would create multiple voiced obstruents within the 'whole word' (e.g. /mîd̲u/ + /t̲öri/ → mîd̲u-**t**öri, *mîd̲u-**d**öri 'water-bird'; Vance 2005; the transcription is his). Vance (2005) and Vance & Irwin (2013) confirm the validity of this generalisation by scrutinising headwords in a comprehensive dictionary of Old Japanese (Jodaigo Jiten Henshu Iinkai 1967). Unger (1977) refers to this ban on multiple voiced obstruents within the whole compound as the 'strong version of Lyman's Law', and later studies simply call it 'Strong Lyman's Law', the term we have already seen. (For more details, also see Miyake 1932; Jodaigo Jiten Henshu Iinkai 1967:31; Ramsey & Unger 1972; Vance 2005; Vance *et al.* 2021.)

It is worth mentioning here that the phonetic realisations of voiced obstruents have also changed historically. Word-medial voiced obstruents were prenasalised in Old Japanese (see Hamada 1952; Miyake 2003:74; Frellesvig 2010:34–36). This may explain the origin of (Strong) Lyman's Law. Prenasalisation is a feature that often spreads out across segments, and multiple instances of prenasalised segments can cause perceptual confusion, which may

[5] These are merely simplified representations. It may be more accurate to state the naturalness of phonological patterns in relative rather than absolute terms. Also, a more quantitative evaluation of each effect will be shown by means of a statistical estimate in the analysis of experimental results in §4.2.

further lead to dissimilation through phonologisation (Ohala 1981, 1993). Kawahara (2008) suggests that Lyman's Law was originally a co-occurrence restriction against prenasalisation. Vance *et al.* (2021) further claim that it was a constraint banning prenasalised segments in consecutive syllables, without regard to morphological boundaries. Diachronic facts thus also explain the original phonetic motivation of the law, which has now been obscured by sound change and morphological bounding.

The properties of Lyman's Law are not the only similarities that current surnames share with Old Japanese compounds. Vance & Irwin (2013) report that /r/ in E1 affected rendaku application in Old Japanese. In their dictionary data, compounds with /r/ in E1's final syllable have a lower rendaku application rate (26%; e.g. /sira/ + /tama/ $\rightarrow$ sira-**t**ama, *sira-**d**ama 'white-jewel') than the overall average (41%). Although they do not conduct a statistical test, nor do they give an explanation as to why this pattern exists at all, they draw the conclusion that E1-/r/ acted as a rendaku-inhibiting segment in Old Japanese (see Vance & Asai 2016 for related discussion). Again, we see the same pattern in Japanese surnames today.

The descriptive statistics in Vance & Irwin (2013) further point to interesting facts about nasals in E1. Old Japanese compounds with E1-/m/ show a lower rendaku rate (around 30%) than those with E1-/n/ (around 50%). This apparent rendaku-inhibiting behaviour of /m/ may be attributed to its similarity to /b/. Historically, /m/ was often interchangeable with /b/ (Martin 1987:31–32; Unger 2004:331–332), which was prenasalised word-medially as [$^{\mathrm{m}}$b] (see above). Previous studies claim that this has some influence on the way rendaku applies in present-day Japanese (Nakagawa 1966; Irwin 2014, 2016a; Ohta 2015; Vance & Asai 2016): E2 morphemes containing /m/ that developed from /b/ (e.g. *ke$^{\mathrm{m}}$buri > kemuri 'smoke') tend to resist rendaku (e.g. /suna/ + /kemuri/ $\rightarrow$ suna-kemuri, *suna-**g**emuri 'sand-smoke'), as if Lyman's Law were in force. It is conceivable that /m/ in 'E1' also inhibited rendaku in Old Japanese, due to its similarity to /b/, as if 'Strong Lyman's Law' would apply. Then, the difference between E1-/m/ and E1-/n/ in current surnames could also be a historical relic.

In sum, many of the rendaku patterns in Old Japanese discussed here look parallel to those in current surnames summarised in Table II. Does this simply mean that surnames are old and have retained archaic sound patterns that once existed in Old Japanese but are no longer attested in regular words in present-day Japanese? The scenario seems compatible with the idea that proper nouns are inherently privileged in the faithfulness hierarchy (Smith 2014; Broad *et al.* 2015; Moreton *et al.* 2017). After some phonological processes lose their original motivations for markedness reduction and go extinct in regular words, they still survive as lexicalised patterns in proper nouns due to greater faithfulness requirements. If this is actually the case with Japanese surnames, it may even be that their quirky rendaku patterns are just historical vestiges and not actively produced by synchronic phonology.[6]

[6] Although not all of the current surnames are as old as Old Japanese (early 7c.–late 8c.), it is still conceivable that they have generally retained obsolete phonological traits. Not a few of them can actually be traced back to the names of clans in Classical Japan (6c.–late 12c.),

3.3 Motivation for experimentation

I have presented a naturalness account and a historical account of rendaku in Japanese surnames. The two accounts are in fact not mutually exclusive. It is possible that the patterns of voicing have their roots in diachrony but are still regulated by synchronic grammar, and naturalness also plays a role. It is thus of interest here to investigate whether and if so how present-day Japanese speakers generalise and reproduce rendaku patterns in surnames. To this end, I conduct two sets of judgment experiments: one experiment with real surnames and the other with nonce surnames as their stimuli.

First, we must examine the voicing patterns of real surnames in greater detail, since the descriptions reported in the literature are not adequate for our purposes. Rendaku is affected by variability and idiosyncrasy. Previous studies (Sugito 1965; Kubozono 2005; Zamma 2005; Tanaka 2017; Zamma & Asai 2017) have not fully taken these factors into account. Their data are based on judgments made by a limited number of speakers, sometimes including the authors themselves, or a web-based corpus with potential noise. Some just focus on certain frequent E2 morphemes. Also, the jury is still out on the naturalness status of one of the factors conditioning rendaku, namely E1-/r/. Running a systematic judgment experiment with a large number of speakers as participants and a large number of existing surnames as stimuli will allow us to better describe the general rendaku patterns of Japanese surnames.

Second, conducting a wug-test-style experiment using nonce words is an established way to test the productivity of a morphophonological process (Berko 1958). There is already some evidence that rendaku in Japanese surnames is productive. As stated above, rendaku is not usually shown in writing, and speakers occasionally need to judge whether there is voicing or not in names they have never seen. If it turns out that speakers systematically apply rendaku in novel surnames in experimental settings, it will corroborate the argument that rendaku in proper nouns is a productive process. Diachrony may give a good account of the peculiar behaviours of surnames. However, as long as they are productively replicated in present-day Japanese, they should not be simply dismissed as fossilised patterns, but should be treated as a synchronic issue.

Lastly, once the patterns of rendaku in real surnames have been established and the productivity of the process has been confirmed with nonce surnames, we can compare the results of the two experiments and see how the patterns are learned. The questions to be addressed here are: when Japanese speakers are presented with novel names, do they faithfully reproduce all the voicing patterns of real names, natural and unnatural patterns alike, or do they generalise certain patterns better than others? This is a kind of surfeit-of-the-

---

or the family names of aristocrats and warriors in Medieval Japan (late 12c.–16c.) (see Toyoda 1971/2012; Sakata 2006 for the history of Japanese surnames). There is a common belief that commoners had no surnames historically and created new ones in the Meiji era (1868–1912). However, records suggest that peasants and merchants did have unofficial surnames before then (Hora 1952; Sakata 2006:42–60). These surnames were probably created in or before medieval times and were modeled after already existing surnames and place names, which presumably had old sound patterns.

stimulus experiment (Becker *et al.* 2011) with a highly faithful category of words, namely proper nouns (Smith 2014; Moreton *et al.* 2017), serving as stimuli. If it turns out that natural patterns are learned and reproduced more readily than unnatural ones, even with respect to proper nouns that are inherently tolerant of idiosyncratic, phonologically unmotivated patterns, this will suggest that analytic biases for naturalness play a role in language learning.

4 Experiment 1: Rendaku in real surnames

4.1 Method

4.1.1 Stimuli

I created an original list of Japanese surnames by combining data from two existing databases. The main data come from the *Database of Japanese Surnames and their Rankings* (Shirooka & Murayama 2011), which lists the 25,000 most common surnames taken from telephone directories. Since the database is entirely based on kanji and no annotations for pronunciation are given, data from another online database (Suzaki 1999), which provides logically possible readings of surnames, were combined with it. I then extracted surnames that are potential rendaku-undergoers (i.e. E2 starts with an underlying voiceless obstruent), are written with two characters (i.e. bimorphemic), and are relatively frequent with more than 1,500 registered households.

Additional notes on the stimulus list are in order. A surname may have multiple possible readings for the kanji in E1. For instance, the kanji 小 'small' has readings such as /o/ and /ko/, and the surname 小川 'small-river' can be [o-kawa] or [o-gawa] with /o/ as E1, but could also be [ko-kawa] or [ko-gawa] with /ko/ as E1. In this study, I consider each of these pairs to be underlyingly different, assuming /o-kawa/ and /ko-kawa/ respectively, even though they are orthographically the same. Surnames may also be morphologically the same but orthographically different. The kanji for /sima/ 'island' has two variant forms 島 and 嶋. Surnames with these variants, such as 中島 /naka-sima/ 'central-island' and 中嶋 /naka-sima/ 'central-island' are treated as different surnames, following the original kanji-based databases.

Certain surnames have genitive /no/, /na/, or /ga/ between E1 and E2, which is not reflected in the orthography (e.g. /ki-no-sita/ 木下 'tree-Gen.-bottom'); they were excluded since genitive particles and rendaku usually do not cooccur (see Lyman 1894; Vance 2007).[7] There are also other factors for rendaku (see Vance 2015a) that are potentially at play in surnames but are not directly relevant to this study. I did not strictly control for all those factors, in order to keep the list of surnames as comprehensive as possible. These

[7] The fifth most common surname /wata-na-he/ 渡辺 'crossing-Gen.-edge' appears to be an exception to this statement as it usually shows voicing despite having genitive /na/, as in [wata-na-be]. Its E2 was in fact historically a different morpheme 部 /be/ 'vocational group' with an underlyingly voiced segment, and was later confused with the voiced form of 辺 /he/ [be] 'edge' (Toyoda 1971/2012:48). Most present-day speakers are not aware of the etymology, but the surname was excluded from the stimulus list for consistency, as it contained /na/.

additional factors are to be taken into consideration in statistical analysis.

The resulting list contained 1,176 surnames with 122 distinct E2 morphemes. Additionally, 12 common surnames composed of three elements (e.g. /o-jama-ta/ 小山田 'small-mountain-paddy') were included as stimuli for practice. A full list of the surnames is provided as supplementary material.

4.1.2 Procedure

The experiment was designed and run on the Internet-based experiment platform Experigen (Becker & Levine 2013). Participants were asked to go to the experiment website by clicking a link posted on a recruitment page. The first page showed a consent form. After agreeing to take the experiment, participants were directed to a general instruction page. They were told that they would be answering questions about the readings of Japanese surnames written in kanji. They were then asked to complete a practice session, which had three randomly selected three-element surnames. After the practice, they moved on to the main session, where they completed 120 trials with randomly selected two-element surnames.

At each trial, participants were presented with a surname written in kanji along with the honorific suffix *-san* 'Mr./Ms.' in a frame sentence, as in 'There is a person called 中田-san'. They were given two numbered options for the reading of the surname written in hiragana, one with rendaku voicing (e.g. 1. *nakada*) and the other without (e.g. 2. *nakata*). (The presentation order of the two types of readings was shuffled for each trial.) They were asked to read the surname out loud using both of the reading options, and to judge which one would sound more natural. To make their response, they clicked on a button marked '1' or '2' according to the number of their selection. Once the response was made, a proceed button would appear. On clicking on the button, they were taken to the next trial.

At the end of the experiment, participants were asked to fill out a questionnaire about personal information, such as their age and home prefecture. They were also asked whether they knew what the term 'Lyman's Law' means, which would indicate their knowledge of linguistics.

4.1.3 Participants

In all, 500 native speakers of Japanese were recruited on the crowdsourcing platform *CrowdWorks* and participated in the experiment. They received 200 Japanese Yen as a reward. The data of 26 participants were excluded as their CrowdWorks ID indicated that they had also participated in Experiment 2 reported in the next section.[8] The data of eight participants were also excluded as they reported that they knew the meaning of Lyman's Law. Some of the responses from three participants were not recorded properly on the data server's database, possibly due to connection issues, and their entire data

[8] Experiment 2 had been conducted earlier than Experiment 1 in the actual timeline. Participants were asked not to participate in both experiments at the time of recruitment, but some still did. I excluded their data from the results of Experiment 1 to be conservative.

were discarded. The data of 463 participants, aged 18 to 71 (mean: 39.03; SD: 10.67), were thus entered in the final analysis.

No recruitment criteria were set with respect to dialects, and speakers of any Japanese dialect were allowed to take part.[9] Although there may be regional differences in the rendaku patterns of surnames (e.g. Sugito 1965; Morioka 2011:23; Iwasaki 2013:42; but see Takemura *et al.* 2019 who report no clear regional effects on rendaku in place names), speakers of all dialects were included in the participant pool because the surname databases on which this study is largely based (Suzaki 1999; Shirooka & Murayama 2011) gathered data from all over Japan, without taking regional differences into consideration. The results here are thus meant to be a sample of the rendaku judgments of Japanese speakers as a whole. Since the task was orthography-based and no audio stimuli were used, the experimental design did not present biases for any particular dialect or any particular accent patterns that could affect participants' judgments (see Sugito 1965; Zamma 2005; Zamma & Asai 2017).

4.2 Results

To first present the descriptive statistics of the results, I plot in Figure 1 the average rendaku rates of surnames with obstruents (top) and sonorants (bottom) in E1-final syllables, each broken down by E2-initial consonant. Error bars represent 95% confidence intervals. Following the descriptions of previous studies (see Table I), I represent the conditions that are expected to promote rendaku with white (□) and light grey (□) bars, and those expected to inhibit rendaku with dark grey (■) bars. All the others are shown in regular grey color (■).

It can be seen that surnames containing voiced obstruents E1-finally, or the conditions labeled 'D', generally show lower rates of rendaku application, indicating that Strong Lyman's Law is in effect. Meanwhile, Identity Avoidance and Similarity Avoidance promote rendaku, as the rates of conditions such as 'E1-s & E2-s' are relatively high. Turning to sonorants, E1-/r/ seems to act as a rendaku inhibitor. Notice that the trend is observed not only for E2-/s/ or E2-/t/, where rendaku would yield a similar sequence such as [r...z] or [r...d], but also in the case of E2-/k/. This suggests that the blocking of rendaku by E1-/r/ is a general effect, rather than being motivated by avoidance of particular sequences. As for the nasals, E1-/n/ shows higher rendaku rates than E1-/m/ for the most part.

These results should not be taken at face value, however. The experiment has used existing surnames as stimuli without much experimental control. For example, each condition contains different numbers of E2 morphemes, including rendaku lovers and haters (Rosen 2001), which might possibly skew the results. Individual participants may also have different trends for rendaku responses. In order to assess the effects of the phonological factors in ques-

[9] The distributions of the 463 participants' home regions are as follows: Hokkaido: 22; Tohoku: 33; Tokyo/Kanto: 141; Tokai-Tosan: 57; Hokuriku: 12; Kinki/Kansai: 95; Chugoku: 30; Shikoku: 12; Kyushu: 50; Okinawa: 4; Other/No Answer: 7.

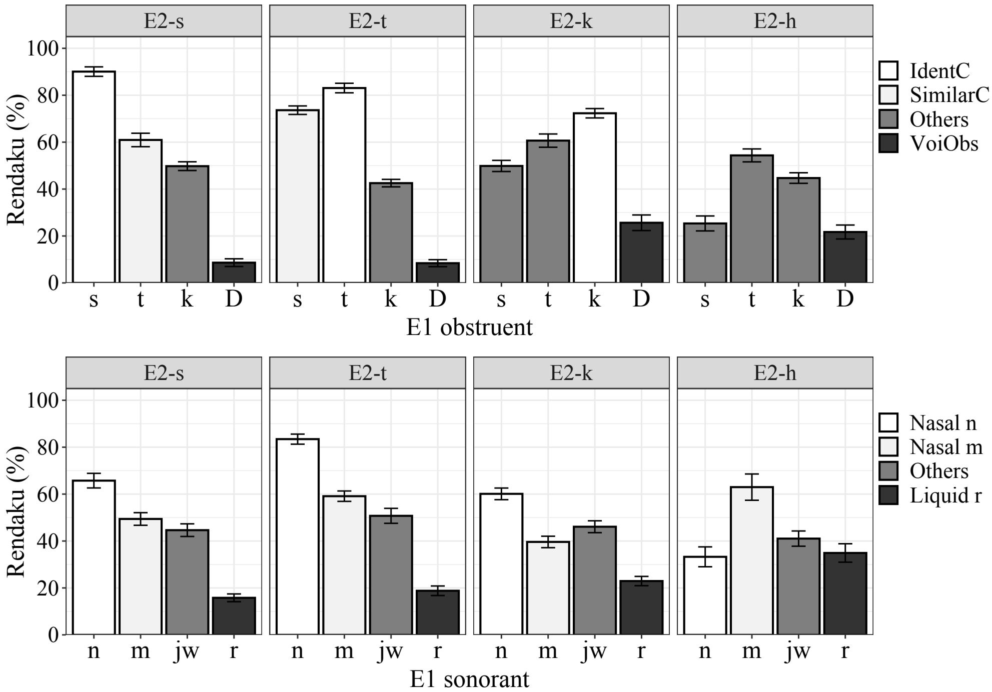

Figure 1
Average rendaku rates: Existing surnames

tion while considering other factors such as the idiosyncratic properties of lexical items and participants, I ran a mixed-effects logistic regression analysis. The model was constructed using the *glmer* function of the lmerTest package (Kuznetsova *et al.* 2017) built on lme4 (Bates *et al.* 2015) in R (R Core Team 2021).[10]

The model's response variable was participants' responses for rendaku application (rendaku or not). Predictors related to obstruents included the presence of a voiced obstruent in the final syllable of E1 ('E1-D'), as well as an underlying identical consonant sequence ('IdentC') and a similar consonant sequence ('SimilarC') across the E1-E2 boundary. Those related to sonorants included the presence of /r/ ('E1-r'), /n/ ('E1-n'), and /m/ ('E1-m') in the final syllable of E1. The obstruency of the consonant in the E1-final syllable ('E1-Obs') was also included in the model in order to capture the potential baseline difference between obstruents and sonorants.

Though not directly relevant to the discussion here, other factors that could affect rendaku application according to the literature were also included in the model: the presence of a so-called special mora in the E1-final position, such as a coda nasal ('E1-CodaN') and a non-nasal consonant ('E1-CodaQ'), the presence of a voiced obstruent in E2 ('E2-D'), and the length of E1 or E2 being equal to or greater than three moras ('LongE1/E2'). (See below for more details on these factors.) For the model's random structure, I set

[10] The model was run with bound optimisation by quadratic approximation with the *bobyqa* optimiser (Powell 2009), with the number of iterations increased to 200,000, in order to avoid convergence issues.

	β	SE	*z*	*p*	
Intercept	–0.278	0.329	–0.845	0.3981	
E1-Obs	0.202	0.279	0.723	0.4695	
IdentC	2.104	0.071	29.638	<0.0001	***
SimilarC	0.298	0.082	3.628	0.0003	***
E1-D	–2.627	0.401	–6.553	<0.0001	***
E1-n	1.292	0.424	3.050	0.0023	**
E1-m	0.351	0.424	0.826	0.4088	
E1-r	–2.048	0.399	–5.135	<0.0001	***
E1-CodaN	3.239	0.545	5.940	<0.0001	***
E1-CodaQ	–3.453	0.892	–3.873	0.0001	***
E2-D	–4.158	0.774	–5.371	<0.0001	***
LongE1/E2	3.262	0.486	6.720	<0.0001	***

Table III
Logistic regression model: Existing surnames

random intercepts for experimental participants, as well as intercepts for both E1 and E2 items, following the observation that each morpheme may show an idiosyncratic behaviour with respect to rendaku (Rosen 2016).[11]

A table of the coefficients is given in Table III. Note that the model presented here is a hypothesis-driven model that includes all the predictors introduced above, and is not necessarily the one that best fits the data with the fewest possible predictors. This is intended to make it easy to examine the effects of the phonological factors of interest in a single model, and also to compare the results with those of the nonce-name experiment presented later. The baseline intercept here refers to the condition where the last consonant of E1 is a sonorant and none of the factors in question is present (roughly corresponding to the 'E1-jw' bars in Figure 1).

The model predicts that both IdentC and SimilarC positively affect rendaku (β = 2.104; β = 0.298). That is, as expected, Identity Avoidance and Similarity Avoidance make compound voicing more likely to occur. The effect of E1-D, namely Strong Lyman's Law, is also significant in that it rather makes rendaku less likely (β = –2.627). E1-n and E1-m raise the probability of rendaku application numerically (β = 1.292; β = 0.351), but the effect of the latter is weaker and also not statistically significant ($p = 0.4088$). This seems more or less compatible with the literature: onset nasals in E1's final syllables tend to facilitate rendaku voicing, but the tendency is more clearly seen with /n/ than /m/. E1-r has a significantly negative effect (β = –2.048), meaning that /r/ in E1's final syllable generally inhibits rendaku. E1-Obs is not significant (β = 0.202, $p = 0.4695$), suggesting that there is no clear

[11] Further including random slopes for participants in the model causes a singular fit (overfitting) problem. I thus report here the model with random intercepts, but the predictions are essentially similar with or without random slopes.

baseline difference between obstruents and sonorants (but see the results of Experiment 2 in §5.2 for comparison).

Additionally, the following predictors have turned out to be significant. A coda nasal in E1 (E1-CodaN) raises rendaku applicability (e.g. /hon-ta/ → [hon-**d**a] 'original-paddy' 本田), indicating that post-nasal voicing is operative (Zamma & Asai 2017; cf. Irwin 2016a; Vance & Asai 2016 about regular compounds). Rendaku is less likely to occur if E1 ends in a non-nasal in the coda (E1-CodaQ); in such a case, the consonant must turn into the first half of a geminate (conventionally represented as 'Q' in Japanese linguistics) through regressive assimilation, and rendaku would further create a voiced geminate (e.g. /hor-ta/ → [hot-ta], $*^?$[ho**d**-**d**a] 'digging-paddy' 堀田), which is disfavored in the language (Ito & Mester 1986; Nishimura 2003; Kawahara 2006). A voiced obstruent in E2 (E2-D) also inhibits rendaku (e.g. /ko-sugi/ → [ko-sugi], $*^?$[ko-**z**ugi] 'short-cedar' 小杉), conforming to the stem-bounded version of Lyman's Law (Lyman 1894). Three-mora or longer E1 and E2 (LongE1/E2) promote rendaku, as is also observed in regular compounds (Rosen 2001, 2003; Vance 2015b; Irwin 2016a,b; cf. Tamaoka *et al.* 2009; Kawahara & Sano 2014c; Tanaka 2020 for complications).

To summarise the results overall, most of the generalisations proposed by previous research have proven to be valid. That is, rendaku in surnames contains statistical regularities that are both natural and unnatural. Of importance here is that these patterns are reflected in Japanese speakers' judgments on existing surnames. In order to further examine whether such generalisations are actually internalised in speakers' phonological grammars, I conduct another judgment experiment with nonce surnames.

5 Experiment 2: Rendaku in nonce names

5.1 Method

5.1.1 Stimuli

Non-existing surnames composed of nonce E1 and real E2 were used as stimuli. For E1, I created items looking like monomorphemic native stems. They were all two moras in length and of the (C)VCV configuration. See Table IV for some of the examples of the E1 items sorted by the last consonant type. (A full list of the words is given as supplementary material.)

Last C type	Last C	Examples				
Voiceless obstruent	s, t, k	hesa	hotu	moki	keso	...
Voiced obstruent	z, b, d, g	juza	sobe	hado	tegu	...
Nasal	m, n	kema	homi	wane	suno	...
Liquid	r	jora	niri	mere	saro	...
Approximant	j, w	kuja	ajo	seju	kewa	...

Table IV
Nonce E1 items

I also made word definitions and example phrases which would be presented along with nonce items in the trials. For example, for some participants, the word /hesa/ would be presented as a type of plant, with the example phrase ‘Leaves of *hesa* are colouring’.

For E2, I used 35 native morphemes with an initial voiceless obstruent that appeared as E2 more than five times in the 1,176 surnames used in Experiment 1. Table V shows the E2 items with their kanji, meanings, and raw frequencies in the real surname data.[12]

E2	Kanji	Freq.	Gloss	E2	Kanji	Freq.	Gloss
sawa	沢	71	‘mountain	kawa	川	89	‘river’
			stream’	ki	木	42	‘tree’
sima	島	67	‘island’	kuti	口	34	‘entrance’
saki	崎	52	‘cape’	kami	上	23	‘top’
saka	坂	21	‘slope’	kura	倉	21	‘warehouse’
se	瀬	21	‘riffle’	kaki	垣	9	‘hedge’
sita	下	9	‘bottom’	kosi	越	8	‘crossing’
seki	関	6	‘gate’	kisi	岸	6	‘shore’
sato	里	6	‘village’	ko	子	6	‘child’
siro	城	5	‘castle’	ki	城	5	‘castle’
ta	田	236	‘paddy’	hara	原	65	‘natural field’
tani	谷	26	‘valley’	hasi	橋	20	‘bridge’
tuka	塚	15	‘mound’	hajasi	林	16	‘wood’
tu	津	14	‘harbor’	hata	畑	10	‘cultivated field’
to	戸	8	‘door’	hama	浜	7	‘beach’
tuki	月	5	‘moon’	huti	渕	7	‘edge’
taka	高	5	‘high, height’	hori	堀	6	‘moat’
taira	平	5	‘plain’	ha	葉	5	‘leaf’

Table V
Real E2 items

E1 items and E2 items were then combined to create non-existing surnames. For example, with E1-/hesa/ and E2-/sima/, a surname /hesa-sima/ ‘hesa-island’ was created. The created surnames can be classified into conditions based on the last consonant of E1 and its combination with the initial consonant of E2. For example, a surname with a voiceless obstruent in the E1-final syllable may have an identical consonant sequence (e.g. /hesa-sima/) or a non-identical sequence (e.g. /eku-sima/) underlyingly. One with an E1-final voiced obstruent can be characterised as a potential Strong Lyman’s Law violator (e.g. /hoze-ta/ → [hoze-**d**a]).

The main session of the experiment had 128 judgment trials. It was designed so that the presented stimuli would be balanced based on the initial consonant of E2. That is, a participant would see 32 surnames with each

[12] Here, E2 morphemes with different kanji variants (e.g. /sawa/ 沢 ∼ 澤 ‘mountain stream’; /sima/ 島 ∼ 嶋 ‘island’; /kawa/ 川 ∼ 河 ‘river’) are counted as the same morpheme. In the experiment, the more frequent variant was used for presentation.

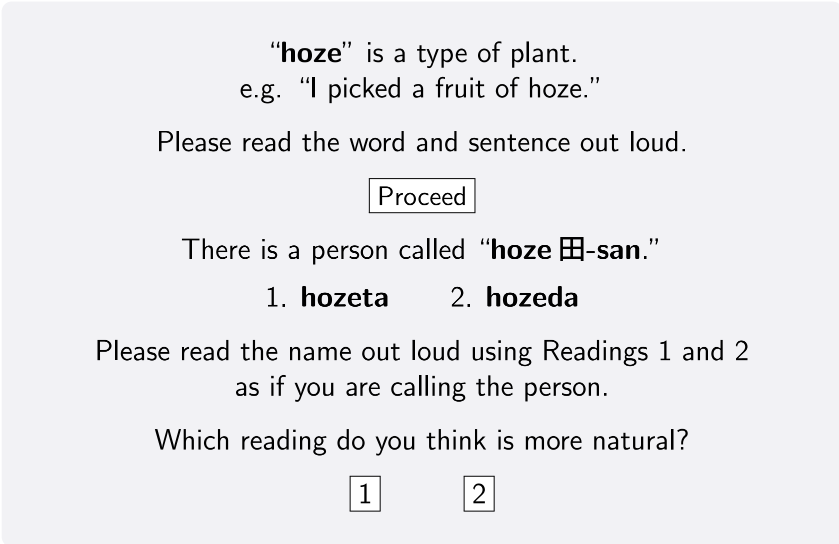

Figure 2
An image of the experimental task

of /s/, /t/, /k/, and /h/ in E2-initial position. Additionally, some E2 morphemes were set to appear more often than others, so that their frequency would roughly match the actual frequency. Those found more than 10 times in the real surname data, or 'frequent E2s', would appear twice as often as other 'infrequent E2s'. For example, for /s/-initial E2, frequent /sawa/, /sima/, /saki/, /saka/, and /se/ would be presented 24 times, while infrequent /sita/, /seki/, /sato/ and /siro/ would be presented 12 times in total. The stimulus set was also balanced based on the last consonant of E1. A participant would receive 16 surnames with each of the three voiceless obstruents (/s/, /t/, and /k/), 16 with a voiced obstruent (/z/, /b/, /d/, or /g/), 16 with each of the nasals (/m/ and /n/), 16 with (/r/), and another 16 with an approximant (/j/ or /w/).

5.1.2 Procedure

The experiment was implemented on Experigen (Becker & Levine 2013). The basic procedure was the same as that of Experiment 1, except that participants were also presented with nonce E1 items before making rendaku judgments. In the instruction session, they were told that they would be presented with some obsolete words or words from some regional dialects of Japanese that might sound unfamiliar to them. They were also told that they would be answering questions about the readings of uncommon Japanese surnames. They completed a practice session with two judgment trials. After the practice, they moved on to the main session, where they completed 128 trials.

An image of the task is shown in Figure 2. (The text is translated from Japanese into English.) At each trial, participants were first presented with an E1 morpheme written in the phonographic *hiragana* script along with its definition and an example sentence, and were asked to read them out loud. On clicking on the Proceed button, new text would appear. A surname composed of the previously presented morpheme as E1 and an existing morpheme as E2 was presented with the honorific suffix /-san/ attached. E1 was written in hiragana (e.g. ほぜ /hoze/) and E2 was written in kanji (e.g. 田 /ta/ 'paddy'). Participants were given the rendaku form (e.g. 1. ほぜだ *hozeda*) and the non-rendaku form (e.g. 2. ほぜた *hozeta*) in hiragana, and were asked to judge which would sound more natural as the reading of the surname. They made their response by clicking on one of the buttons according to the number of their selection. The order of the stimuli was randomised for each participant, and the order of the response options (rendaku or no rendaku) was shuffled for each trial.

5.1.3 Participants

A total of 150 native Japanese speakers were recruited through *CrowdWorks*, and received 360 Japanese Yen for participation. The data of two participants were excluded as they reported that they knew what Lyman's Law means, or did not answer the question. The data of three participants were also excluded as some of their responses were not recorded properly for some unknown reason. As a result, the data of 145 participants, aged 19 to 65 (mean: 39.84; SD: 10.24), were included in analysis. As in Experiment 1, the participants were from various regions of Japan.[13]

5.2 Results

Figure 3 graphs the average rendaku response rates of nonce surnames with E1-obstruents (top) and E1-sonorants (bottom) by E2's initial consonant. Error bars represent 95% confidence intervals calculated based on participant means. Would-be rendaku promoters are shown in white (□) and light grey (□), rendaku inhibitors in dark grey (■), and all the others in regular grey (■).

Surnames with E1-final voiced obstruents (D) appear to have lower rates than those with voiceless obstruents. This suggests that, by and large, Strong Lyman's Law is operative in rendaku in non-existing surnames as well. Identity Avoidance and Similarity Avoidance also seem to be at work, since surnames with 's-s', 't-t', 'k-k', and 't-s' show relatively high rates. In contrast, the rates of surnames with E1-sonorants appear to be neither very high nor low, hovering around 50%. Note that, compared to the results of real surnames, E1-/r/'s rendaku-inhibiting effects and the differences, if any, between E1-/n/ and E1-/m/ appear small and inconsistent.

A mixed-effects logistic regression model was fitted to the data in the same manner as in Experiment 1 with the same fixed predictors and random structure, except that the predictors did not include controlled factors such as the presence of a E1-final special mora and the length of E1 and E2.[14] The results are shown in Table VI. Again, the model's baseline intercept corresponds to

[13] Their home regions are as follows: Hokkaido: 11; Tohoku: 7; Tokyo/Kanto: 44; Tokai-Tosan: 20; Hokuriku: 2; Kinki/Kansai: 26; Chugoku: 11; Shikoku: 7; Kyushu: 13; Okinawa: 1; No Answer: 3.

[14] The nonce E1 items all have two moras. There are two E2 items that have three moras (Table V) but their potential idiosyncratic behaviours are already captured as random effects.

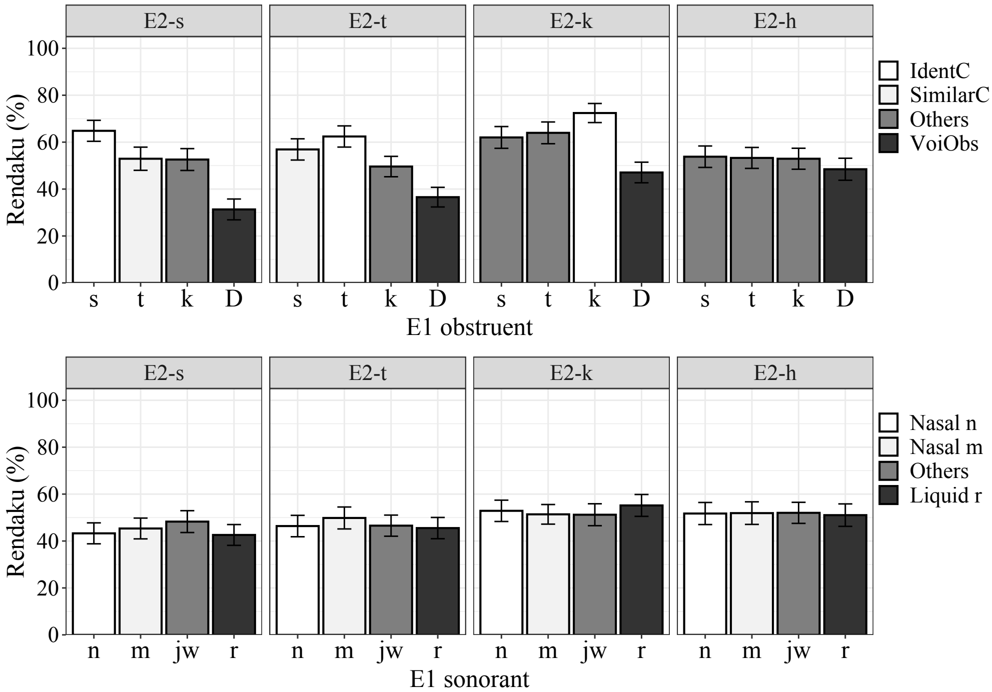

Figure 3
Average rendaku rates: Nonce surnames

the condition where the last consonant of E1 is a sonorant and none of the other factors is present (i.e. roughly the 'E1-jw' bars in Figure 3).

As it shows, IdentC and SimilarC raise rendaku applicability ($\beta = 0.705$; $\beta = 0.208$), while E1-D lowers it ($\beta = -0.812$). These results suggest that the patterns of voicing driven by Identity Avoidance, Similarity Avoidance, and Strong Lyman's Law are all productively reproduced in nonce surnames. In addition, E1-Obs has a positive effect ($\beta = 0.219$), indicating that the rendaku rates of E1-obstruents are generally higher than those of E1-sonorants. Although this trend is not found in real surnames, it may also be interpreted as being phonologically motivated. When there is a sequence of voiceless obstruents in general (whether homorganic or not), rendaku occurs to achieve dissimilation in voicing (e.g. /t...k/ → t...**g**).

On the other hand, the effects of E1-sonorants are less clear. According to the model, E1-n does not particularly increase rendaku applicability ($\beta = -0.026$), nor does E1-m ($\beta = -0.043$). E1-r is predicted to make rendaku less likely, but the effect is still relatively weak ($\beta = -0.144$) and is also not statistically significant at the conventional alpha level of 0.05 ($p = 0.0825$). These results suggest that the patterns of voicing involving E1-sonorants are not robustly extended to novel surnames.

The effects of the predictors were also assessed by means of model comparison. Simpler models were constructed by removing each predictor and were compared to the full model in terms of goodness of fit. Based on the results of likelihood ratio tests, all the predictors related to obstruents were confirmed to contribute to the full model's better fit. In contrast, the fit does not improve in a statistically significant manner by including E1-n ($\chi^2(1) =$

	β	SE	z	p	
Intercept	0.061	0.241	0.253	0.8006	
E1-Obs	0.219	0.073	3.000	0.0027	**
IdentC	0.705	0.073	9.599	<0.0001	***
SimilarC	0.208	0.084	2.485	0.0129	*
E1-D	–0.812	0.068	–11.895	<0.0001	***
E1-n	–0.026	0.083	–0.315	0.7524	
E1-m	–0.043	0.083	–0.521	0.6027	
E1-r	–0.144	0.083	–1.736	0.0825	.

Table VI
Logistic regression model: Nonce surnames

0.099, $p = 0.7529$) or E1-m ($\chi^2(1) = 0.270$, $p = 0.6036$). The same is true for E1-r ($\chi^2(1) = 2.949$, $p = 0.0859$) with the alpha level set at 0.05.

The non-significant results for E1-sonorants, especially the effect of E1-r, are not easily interpreted by themselves, but they should be contrasted with the results for E1-obstruents. In the very same experiment, factors such as IdentC and E1-D were found to be more sound and significant. The whole results can also be compared to those of Experiment 1. Recall that in real surnames, most of the effects on rendaku under discussion, including E1-n and E1-r, were in fact valid. The discrepancies suggest that the rendaku patterns with E1-sonorants are not strongly internalised in Japanese speakers' minds.

6 Discussion

6.1 Alternatives: Robustness and scope

The two experiments have shown that the rendaku patterns found in real surnames do not have the same degree of productivity in novel surnames. Given the naturalness statuses of these patterns (Table II), this can be taken as evidence that speakers are biased toward learning natural patterns. Before claiming that it is actually so, I consider alternative explanations.

An anonymous reviewer points out that the results could instead be explained in terms of the robustness of the patterns in question. According to the estimated coefficients of the regression model for real surnames (Table VII), the magnitude of rendaku promotion effect is larger for IdentC ($\beta = 2.104$, $z = 29.638$) than for E1-n ($\beta = 1.292$, $z = 3.050$), and that of inhibition effect is larger for E1-D ($\beta = –2.627$, $z = –6.553$) than for E1-r ($\beta = –2.048$, $z = –5.135$). It is then possible that the results of Experiment 2 merely reflect these differences in robustness. That is, the patterns that I have described as natural are just stronger in the data of real surnames, and it is easier for learners to generalise them.

Although it is difficult to completely rule out this possibility, especially regarding the effects discussed above, it seems that robustness is not a panacea either. Notice that the promotion effect of SimilarC in real surnames is relatively weak (β = 0.298, z = 3.628), but still shows up in nonce surnames (β = 0.208, z = 2.485) on top of the E1-Obs effect. This can be compared to the inhibition effect of E1-r, which is at least stronger in real surnames (β = –2.048, z = –5.135) but is greatly reduced in nonce surnames (β = –0.144, z = –1.736). This may not be a fair comparison, given that one is promotion and the other is inhibition, but it still suggests that not every robust pattern shows up as is, and that the difference remains to be explained.

The reviewer also suggests another possible confounding factor: the scope of the patterns in question. There remains uncertainty in the literature as to whether these voicing patterns are truly unique to surnames. This means that Japanese speakers might actually be exposed to the reported statistical regularities in other contexts of rendaku, and also possibly to just a subset of them in those contexts. Crucially, if it is the case that the patterns deemed to be natural are more prominent in a broader range of data, it should be easier for learners to generalise them.

In order to address this issue, I reexamined the patterns of rendaku in regular compounds. I used the latest version of the *Rendaku Database* made by Irwin *et al.* (2020), which contains compounds taken mainly from two large dictionaries of Japanese (see Irwin 2016a for details). I converted the '+' and '–' signs, which indicate 'rendaku' and 'no rendaku' respectively, into 1's and 0's.[15] I restricted my data to compounds with purely nominal E2 morphemes, excluding those with (de)verbal and (de)adjectival ones, so that the same analysis as for surnames could be applied.[16] I also excluded words if they were annotated as being used only as proper nouns. After the procedure, 20,476 words remained. I conducted a mixed-effects logistic regression analysis on them in the same manner as in Experiment 1, except that some of the main predictors (E1-Q and E2-D) and random intercepts for participants were not included, as there were no relevant data.

The results are shown in Table VII. Note that they are quite similar to those of Experiment 1 (Table III). Most of the main predictors have significant effects in the expected directions. This is perhaps a little surprising, given that I have described these patterns as peculiar characteristics of surnames. A closer look suggests that the tendencies are generally weaker than in existing surnames, but are still there in the large-scale data. It is possible that, given that the database includes not a few obsolete words (see Irwin 2016a), old traits such as Strong Lyman's Law and E1-r are made more visible. It might be that these effects would be weaker if other predictors were also considered (Ohta 2015), or even non-significant if they were tested with spontaneous speech data (Sano 2015) or in experimental settings (Kawahara & Sano 2016). I leave these issues for future research.

[15] Words that show variation are marked with '+/–' signs. If the two source dictionaries both give a '+/–', I discarded the word. If one gives a '+/–' but the other gives a '+' or '–', I adopted the description of the latter and assigned the word 1 or 0 accordingly.

[16] Verbs and adjectives tend to have specific phonological shapes, which would skew the results. Rendaku in (de)verbal compounds is also affected by morphological and semantic factors (see Vance 2015a and references therein).

	β	SE	*z*	*p*	
Intercept	0.532	0.147	3.612	0.0003	***
E1-Obs	0.092	0.101	0.911	0.3623	
IdentC	0.897	0.131	6.830	<0.0001	***
SimilarC	0.022	0.177	0.123	0.9021	
E1-D	–0.453	0.116	–3.908	<0.0001	***
E1-n	0.576	0.167	3.440	0.0006	***
E1-m	–0.043	0.128	–0.333	0.7394	
E1-r	–0.242	0.111	–2.186	0.0288	*
E1-CodaN	0.054	0.159	0.342	0.7325	
LongE1/E2	1.841	0.083	22.111	<0.0001	***

Table VII
Logistic regression model: Regular compounds

To return to the main point, the scope-based explanation discussed above seems no longer tenable. Both natural and unnatural patterns are present also in regular compounds, albeit somewhat weakly in relative terms. If these data affect the learning of rendaku in surnames in any way, the learning of both kinds of patterns should be boosted. One may still argue that robustness plays a role here, too. In regular words, IdentC and E1-D appear to have greater effects than E1-n and E1-r. Although that is true, the 'rendaku-promotion' effect of E1-n in nonce surnames looks much weakened (β = –0.026, z = –0.315, p = 0.7524), considering what would be expected simply from its robustness in regular words (β = 0.576, z = 3.440, p = 0.0006). Furthermore, the rendaku-promotion effect of SimilarC is not even significant in regular words (β = 0.022, z = 0.123, p = 0.9021), but still shows up in nonce surnames (β = 0.208, z = 2.485, p = 0.0129). In sum, the results of Experiment 2 cannot be fully explained by scope or robustness.

6.2 Naturalness biases in names

I now argue that the results should be explained in terms of learning biases for phonological naturalness. Again, speakers of present-day Japanese are exposed to both natural and unnatural patterns with respect to rendaku in real surnames, and also in regular words. Nevertheless, they replicate the former more productively in novel surnames. This is very much compatible with the idea that speakers are biased toward only generalising natural patterns.

It should be highlighted here that this study is concerned with rendaku in the domain of proper nouns in particular. Names tend to exhibit peculiar patterns, and this has been explained by a strong faithfulness requirement specific to this word category (Smith 2014; Broad *et al.* 2015; Moreton *et al.* 2017). Put differently, learners are generally willing to accept idiosyncratic patterns, or perhaps any patterns, in proper nouns, attributing them to input specifications. Rendaku is also known for its irregularity, posing challenges to linguists and possibly to learners. This has led some to argue that the phenomenon should just be treated as lexicalised (see Vance 2014; Kawahara 2015 for discussion). The patterns of rendaku in surnames as a whole could then well be learned as lexicalised patterns. Even so, speakers have still extracted phonological regularities that are well motivated and extended them to novel surnames. Meanwhile, they have underlearned unmotivated ones. I take this to indicate that they have failed to see them as real phonological regularities, and treat them as the lexicalised properties of individual surnames with name-specific faithfulness. Overall, these facts corroborate the general claim that a set of analytic biases for phonological naturalness exists; their effects appear even in the most faithful word category.

6.3 Structure or substance

As has been discussed in §3.1, analytic biases for phonological naturalness may be divided into two kinds: one based on formal structure and the other on phonetic substance. The literature suggests that the former exert stronger effects on language learning than the latter (e.g. Pycha *et al.* 2003; see Moreton & Pater 2012b). Does the case of Japanese surnames offer any insight into this issue? Table VIII presents a summary of the patterns of rendaku in surnames with their naturalness and productivity statuses indicated.

Segments	Effect	Structure	Substance	Productivity
Voiced obstruent: /z, b, d, g/	inhibition (Strong Lyman's Law)	✓	✗	Yes
Homorganic pair: /s-s, t-t, k-k, s-t, t-s/	promotion (Identity/Similarity Avoidance)	✓	✓	Yes
Liquid: /r/	inhibition	✗	✗?	Weakly?
Nasal: /n, m/	promotion; but less by /m/	✗	✗	No

Table VIII
Rendaku in surnames viewed from naturalness and productivity

Once again, Identity and Similarity Avoidance, which are considered natural both in terms of structure and substance, are productive. Strong Lyman's Law, which is formally simple but not phonetically motivated, has also shown to be effective. In contrast, the patterns related to sonorants, which are deemed unnatural from both perspectives, have not proven to be fully productive. Since the E1-/r/ effect has shown no phonetic motivation, there is no condition under which the phonetic substance could be specifically tested, with structure-based factors controlled. Thus, the results here do not provide a clear-cut answer to the question of which kind of bias is stronger. Given the productivity of Strong Lyman's Law, it can be said that the effect of formal structure is at least strong enough to manifest itself. The effect of phonetic substance alone as it relates to the phonology of proper nouns remains to be tested.

6.4 Other remaining issues

There are several remaining issues. First, the effect of E1-/r/ in nonce surnames was not statistically significant, but its weak trend should probably not be neglected. It could be, of course, that the design of the experiment was not sensitive enough to detect such a feeble effect, which would suggest that unnatural patterns are not completely unlearnable (e.g. Hayes *et al.* 2009; White 2014), even though they are at least harder to learn.

One reviewer asks what would happen if the nonce items used in Experiment 2 were presented as common nouns. Now that the natural and unnatural rendaku patterns have also been found as weaker trends in regular words, we could actually address the same question about learning biases with them. There is one experimental study (Kawahara & Sano 2014b) that has found no evidence for Strong Lyman's Law in regular compounds. It would indeed be interesting to investigate the effects of other factors and compare the results between proper nouns and common nouns.

Lastly, I have abstracted away the prosodic patterns of surnames, as I recruited experimental participants speaking various dialects of Japanese having different accent patterns. It is reported that, in Tokyo Japanese, rendaku often cooccurs with unnaccentedness (e.g. Sugito 1965; Zamma 2005; Zamma & Asai 2017).[17] Examining whether this holds in nonce surnames would be another research topic worth pursuing.

7 Conclusion

The current study has offered evidence for learning biases from a slightly new perspective. By focusing on the phonology of proper nouns, I have shown that speakers are biased toward generalising natural patterns even in a highly faithful word category. In generative linguistics, segmental alternations in names have received relatively little attention. As I have shown, however, alternations in proper nouns can be productive, and studying their patterns closely can also help clarify the effects of learning biases. Further investigations of the phonology of proper nouns with the aim of addressing learnability and other theoretical issues are awaited.

REFERENCES

Alderete, John D. & Stefan A. Frisch (2007). Dissimilation in grammar and the lexicon. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*. Cambridge:

[17] It has also been pointed out during the review process that the prosodic patterns of given names in Japanese show very regular patterns; most of them have antepenultimate accent or are unnaccented (Tanaka & Kubozono 1999; Tanaka & Sugawara 2018). This is true of surnames as well. Answering the question of why regularities and irregularities coexist in proper nouns is beyond the scope of this study. I leave it for future work.

Cambridge University Press. 379–398.

Asai, Atsushi (2014). Rendaku seiki no keikō to teichaku-ka [The tendency of rendaku application and its stabilization]. *NINJAL Research Papers* **7**. 27–44.

Asai, Atsushi & Timothy J. Vance (2017). Kobetsu onso to rendaku [Individual phonemes and rendaku]. In Vance *et al.* (2017). 47–68.

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker (2015). lme4: linear mixed-effects models using 'Eigen' and S4. *Journal of Statistical Software* **67**. 1–48. https://doi.org/10.18637/jss.v067.i01.

Becker, Michael, Nihan Ketrez & Andrew Nevins (2011). The surfeit of the stimulus: analytic biases filter lexical statistics in Turkish laryngeal alternations. *Lg* **87**. 84–125.

Becker, Michael & Jonathan Levine (2013). Experigen – an online experiment platform. http://becker.phonologist.org/experigen.

Becker, Michael, Andrew Nevins & Jonathan Levine (2012). Asymmetries in generalizing alternations to and from initial syllables. *Lg* **88**. 232–268.

Beguš, Gašper (2018). *Unnatural phonology: a synchrony-diachrony interface approach*. PhD dissertation, Harvard University.

Berko, Jean (1958). The child's learning of English phonology. *Word* **14**. 150–177.

Blevins, Juliette (2004). *Evolutionary phonology: the emergence of sound patterns*. Cambridge: Cambridge University Press.

Blevins, Juliette (2017). Between natural and unnatural phonology: the case of cluster splitting epenthesis. In Claire Bowern, Laurence Horn & Raffaella Zanuttini (eds.) *On looking into words (and beyond)*. Berlin: Language Science Press. 3–16.

Breiss, Canaan & Bruce Hayes (2020). Phonological markedness effects in sentence formation. *Lg* **96**. 338–370.

Broad, Rachel, Brandon Prickett, Elliott Moreton, Katya Pertsova & Jennifer L. Smith (2015). Emergent faithfulness to proper nouns in novel English blends. *WCCFL* **33**. 77–78.

Donegan, Patricia J. & David Stampe (1979). The study of Natural Phonology. In Daniel A. Dinnsen (ed.) *Current approaches to phonological theory*. Indiana: Bloomington. 126–173.

Frellesvig, Bjarke (2010). *A history of the Japanese language*. Cambridge: Cambridge University Press.

Frisch, Stefan A. (2004). Language processing and segmental OCP. In Hayes *et al.* (2004). 346–371.

Frisch, Stefan A., Janet B. Pierrehumbert & Michael B. Broe (2004). Similarity avoidance and the OCP. *NLLT* **22**. 179–228.

Garrett, Andrew & Keith Johnson (2013). Phonetic bias in sound change. In Alan C. L. Yu (ed.) *Origins of sound change: approaches to phonologization*. Oxford: Oxford University Press. 51–97.

Hamada, Atsushi (1952). Hatsuon to dakuon to no sōkan-sei no mondai [The issues in the relation between moraic nasals and voiced obstruents]. *Kokugo Kokubun* **21**. 18–32.

Hansson, Gunnar (2008). Diachronic explanations of sound patterns. *Language and Linguistics Compass* **2**. 859–893.

Hayes, Bruce (1999). Phonetically driven phonology: the role of Optimality Theory and inductive grounding. In Michael Darnell, Edith Moravcsik, Michael Noonan, Frederick Newmeyer & Kathleen Wheatley (eds.) *Functionalism and formalism in linguistics*. Vol. 1. Amsterdam: John Benjamins Publishing Company. 243–285.

Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.) (2004). *Phonetically based* *phonology*. Cambridge: Cambridge University Press.

Hayes, Bruce & Tanya Stivers (2000). The phonetics of post-nasal voicing. Ms, University of California, Los Angeles. Available (March 2022) at https://linguistics.ucla.edu/people/hayes/Phonet/NCPhonet.pdf.

Hayes, Bruce & James White (2013). Phonological naturalness and phonotactic learning. *LI* **44**. 45–75.

Hayes, Bruce, Kie Zuraw, Péter Siptár & Zsuzsa Londe (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Lg* **85**. 822–863.

Hirano, Takanori (2013). A rule application approach to rendaku. Paper presented at International Conference on Phonetics and Phonology 2013, Tokyo.

Hora, Tomio (1952). Edo jidai no ippan shomin wa hatashite myōji o motanakatta ka [Is it really the case that commoners in the Edo period did not have surnames?]. *Nihon Rekishi* **50**. 2–7.

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

Irwin, Mark (2016a). The rendaku database. In Vance & Irwin (2016). 79–106.

Irwin, Mark (2016b). Rosen's Rule. In Vance & Irwin (2016). 107–117.

Irwin, Mark, Mizuki Miyashita, Kerri Russell & Yu Tanaka (2020). The rendaku database. Version 4.0. Available (March 2022) at https://zaonotes.com/rendaku-database/.

Ito, Junko & Armin Mester (1986). The phonology of voicing in Japanese: theoretical consequences for morphological accessibility. *LI* **17**. 49–73.

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

Iwasaki, Shoichi (2013). *Japanese: revised edition*. London Oriental and African Language Library 17. Amsterdam and Philadelphia: John Benjamins.

Jaber, Aziz & Osama Omari (2018). Proper name subcategory: a prominent position. *Language Sciences* **69**. 113–124.

Jodaigo Jiten Henshu Iinkai (ed.) (1967). *Jidaibetsu kokugo daijiten Jōdaihen [The unabridged dictionary of the national language by age: Old Japanese]*. Tokyo: Sanseido.

Katz, William F., Sonya Mehta & Matthew Wood (2018). Effects of syllable position and vowel context on Japanese /r/: kinematic and perceptual data. *Acoustical Science and Technology* **39**. 130–137.

Kawahara, Shigeto (2006). A faithfulness ranking projected from a perceptibility scale: the case of voicing in Japanese. *Lg* **82**. 536–574.

Kawahara, Shigeto (2008). Phonetic naturalness and unnaturalness in Japanese loanword phonology. *Journal of East Asian Linguistics* **18**. 317–330.

Kawahara, Shigeto (2015). Can we use rendaku for phonological argumentation? *Linguistics Vanguard* **1**. 3–14. https://doi.org/10.1515/lingvan-2015-0001.

Kawahara, Shigeto, Hajime Ono & Kiyoshi Sudo (2006). Consonant co-occurrence restrictions in Yamato Japanese. In Timothy J. Vance & Kimberley Jones (eds.) *Japanese/Korean linguistics*. Vol. 14. Stanford: CSLI Publications. 27–38.

Kawahara, Shigeto & Shin-ichiro Sano (2014a). Identity Avoidance and Lyman's Law. *Lingua* **150**. 71–77.

Kawahara, Shigeto & Shin-ichiro Sano (2014b). Identity Avoidance and rendaku. In John Kingston, Claire Moore-Cantwell, Joe Pater & Robert Staubs (eds.) *Proceedings of the 2013 Annual Meeting on Phonology*. Washington, DC: Linguistic Society of America. https://doi.org/10.3765/amp.v1i1.23.

Kawahara, Shigeto & Shin-ichiro Sano (2014c). Testing Rosen's Rule and Strong

Lyman's Law. *NINJAL Research Papers* **7**. 111–120.
Kawahara, Shigeto & Shin-ichiro Sano (2016). Rendaku and Identity Avoidance: consonantal identity and moraic identity. In Vance & Irwin (2016). 47–55.
Kawahara, Shigeto & Hideki Zamma (2016). Generative treatments of rendaku and related issues. In Vance & Irwin (2016). 13–34.
Kenstowicz, Michael J. & Charles W. Kisseberth (1977). *Topics in phonological theory*. New York: Academic Press.
Kindaichi, Haruhiko (1976/2005). Rendaku no kai [An account of rendaku]. In *Kindaichi Haruhiko chosakushū [Haruhiko Kindaichi collection]*. Vol. 6. Tokyo: Tamagawa Daigaku Shuppanbu. 583–614. [Originally published in 1976 in *Sophia Linguistica* 2:1–22].
Kindaichi, Haruhiko, Oki Hayashi & Takeshi Shibata (eds.) (1988). *Nihongo hyakka daijiten [An encyclopaedia of the Japanese language]*. Tokyo: Taishūkan.
Kubozono, Haruo (2005). Rendaku: its domain and linguistic conditions. In van de Weijer *et al.* (2005). 5–24.
Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software* **82**. 1–26. https://doi.org/10.18637/jss.v082.i13.
Labrune, Laurence (2012). *The phonology of Japanese*. The phonology of the world's languages. Oxford: Oxford University Press.
Labrune, Laurence (2014). The phonology of Japanese /r/: a panchronic account. *Journal of East Asian Linguistics* **23**. 1–25.
Lyman, Benjamin (1894). The change from surd to sonant in Japanese compounds. In The Oriental Club in Philadelphia (ed.) *Oriental studies: a selection of the papers read before the Oriental Club in Philadelphia 1888-1894*. Boston: Ginn and Company. 160–176.
Martin, Andrew (2011). Grammars leak: modeling how phonotactic generalizations interact within the grammar. *Lg* **87**. 751–770.
Martin, Alexander & Sharon Peperkamp (2020). Phonetically natural rules benefit from a learning bias: a re-examination of vowel harmony and disharmony. *Phonology* **37**. 65–90.
Martin, Samuel E. (1987). *The Japanese language through time*. Yale University Press.
McCarthy, John (1986). OCP effects: gemination and antigemination. *LI* **17**. 207–263.
McCarthy, John (1988). Feature geometry and dependency: a review. *Phonetica* **43**. 84–108.
Miyake, Hideo Marc (2003). *Old Japanese: a phonetic reconstruction*. London: Routledge.
Miyake, Takeo (1932). Dakuonkō [An examination of voiced obstruents]. *Onsei no Kenkyū* **5**. 135–190.
Moreton, Elliott. (2008). Analytic bias and phonological typology. *Phonology* **25**. 83–127.
Moreton, Elliot (2012). Inter- and intra-dimensional dependencies in implicit phonotactic learning. *Journal of Memory and Language* **67**. 165–183.
Moreton, Elliott & Joe Pater (2012a). Structure and substance in artificial-phonology learning. Part I: Structure. *Language and Linguistics Compass* **6**. 686–701.
Moreton, Elliott & Joe Pater (2012b). Structure and substance in artificial-phonology learning. Part II: Substance. *Language and Linguistics Compass* **6**. 702–718.
Moreton, Elliott, Jennifer L. Smith, Katya Pertsova, Rachel Broad & Brandon Prick-

ett (2017). Emergent positional privilege in novel English blends. *Lg* **93**. 347–380.

Morioka, Hiroshi (2011). *Myōji no nazo [The mysteries of surnames]*. Tokyo: Chikuma Shobo.

Morita, Takeshi (1977). Nippo-jisho ni mieru goon-ketsugō-jō no ichi-keikō [A combinatorial tendency of sounds in Vocabulario da Lingoa de Iapam]. *Kokugogaku* **108**. 20–29.

Motoori, Norinaga (1790-1822). *Kojiki-den [Commentaries on the Kojiki, Records of Ancient Matters]*. Nagoya: Eirakuya.

Myers, Scott & Jaye Padgett (2014). Domain generalisation in artificial language learning. *Phonology* **31**. 399–433.

Nakagawa, Yoshio (1966). Rendaku, rensei (kashō) no keifu [A genealogy of sequential voicing and sequential unvoicing (working label)]. *Kokugo Kokubun* **35**. 302–314.

Nishimura, Kohei (2003). *Lyman's Law in loanwords*. Master's thesis, Nagoya University.

Ohala, J. John (1981). The listener as a source of sound change. In Carrie S. Masek, Roberta A. Hendrick & Mary Frances Miller (eds.) *Papers from the parasession on language and behavior: Chicago Linguistic Society, May 1-2, 1981*. Chicago: Chicago Linguistic Society. 178–203.

Ohala, John J. (1993). The phonetics of sound change. In Charles Jones (ed.) *Historical linguistics: problems and perspectives*. London: Longman Academic. 237–278.

Ohta, Shinri (2015). Onin-teki/imi-teki yōin ga rendaku ni ataeru eikyō: rendaku dētabēsu to rojisutikku kaiki bunseki o riyō shita kenkyū [The effects of phonological and semantic factors on rendaku: a study using a rendaku database and logistic regression analyses]. *On'in Kenkyū* **18**. 85–92.

Pater, Joe (1999). Austronesian nasal substitution and other NC̥ effects. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.) *The prosody-morphology interface*. Cambridge: Cambridge University Press. 310–343.

Powell, Michael J. D. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. Technical report. Department of Applied Mathematics and Theoretical Physics, Cambridge University. Available (March 2022) at http://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf.

Pycha, Anne, Pawel Nowak, Eurie Shin & Ryan Shosted (2003). Phonological rule-learning and its implications for a theory of vowel harmony. *WCCFL* **22**. 101–114.

R Core Team (2021). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Version 4.1.2. https://www.R-project.org/.

Ramsey, Robert & Marshall Unger (1972). Evidence for a consonant shift in 7th century Japanese. *Papers in Japanese Linguistics* **1**. 279–295.

Rosen, Eric (2001). *Phonological processes interacting with the lexicon: variable and non-regular effects in Japanese phonology*. PhD dissertation, University of British Columbia.

Rosen, Eric (2003). Systematic irregularity in Japanese rendaku: how the grammar mediates patterned lexical exceptions. *Canadian Journal of Linguistics* **48**. 1–37.

Rosen, Eric (2016). Predicting the unpredictable: capturing the apparent semi-regularity of rendaku voicing in Japanese through Gradient Symbolic Computation. *BLS* **42**. 235–249.

Sakata, Satoshi (2006). *Myōji to namae no rekishi [History of surnames and given names]*. Rekishi Bunka Library. Tokyo: Yoshikawa Kobunkan.

Sano, Shin-ichiro (2015). Universal markedness reflected in the patterns of voicing process. *NELS* **45**. 49–58.

Sato, Hirokazu (1989). Fukugō-go ni okeru akusento kisoku to rendaku kisoku [Accent rules and rendaku rules in compounds]. In *Kōza nihongo to nihongo kyōiku 2: Nihongo no onsei, on'in (jō) [Japanese and Japanese teaching 2: Japanese phonetics, phonology 1]*. Tokyo: Meiji Shoin. 233–265.

Shirooka, Keiji & Tadashige Murayama (2011). A database of Japanese surnames and their rankings. Available (March 2022) at http://hdl.handle.net/10297/00025667.

Smith, Jennifer (2011). Category-specific effects. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.) *The Blackwell companion to phonology*. Vol. 4: Phonological interfaces. Malden: Wiley-Blackwell. 2439–2463.

Smith, Jennifer L. (2014). Prototypical predicates have unmarked phonology. In John Kingston, Claire Moore-Cantwell, Joe Pater & Robert Staubs (eds.) *Proceedings of the 2013 Annual Meeting on Phonology*. Washington DC: Linguistic Society of America. https://doi.org/10.3765/amp.v1i1.37.

Sugito, Miyoko (1965). Shibata-san to Imada-san: tango no chōkakuteki benbetsu ni tsuite no ichi kōsatsu [Shibata-san and Imada-san: an examination of auditory distinction of words]. *Gengo Seikatsu* **165**. 64–72.

Sugito, Miyoko (1982). Kinki hōgen-ni okeru za-gyōon, da-gyōon, ra-gyōon no kondō ni tsuite [Perceptual confusions of /z/, /d/, and /r/ in Kinki dialects]. In Kiichi Iitoyo, Sugezumi Hino & Ryoichi Sato (eds.) *Kōza hōgengaku [Courses in dialectology]*. Vol. 7. Tokyo: Tosho Kankokai. 299–325.

Suzaki, Haruo (1999). A private on-line database of 120,000 Japanese family names. Available (March 2022) at https://suzaki.skr.jp/index40.html.

Suzuki, Keiichiro (1998). *A typological investigation of dissimilation*. PhD dissertation, University of Arizona.

Takayama, Michiaki (1992). Rendaku to renjōdaku [On sequential voicing and sequential post-nasal voicing]. *Kuntengo to Kunten Shiryō* **88**. 115–124.

Takemura, Akiko, Thomas Pellard, Hyun Kyung Hwang & Timothy J. Vance (2019). Rendaku in place names across Japanese dialects. *Reports of the Keio Institute of Cultural and Linguistic Studies* **50**. 79–89.

Tamaoka, Katsuo, Mutsuko Ihara, Tadao Murata & Hyunjung Lim (2009). Effects of first-element phonological-length and etymological-type features on sequential voicing (rendaku) of second elements. *Journal of Japanese Linguistics* **25**. 17–38.

Tanaka, Shin'ichi & Haruo Kubozono (1999). *Nihon-go no hatsuon kyōshitsu: riron to renshū [A course in Japaense pronunciation: theory and practice]*. Tokyo: Kurosio Publishers.

Tanaka, Yu (2017). Phonotactically-driven rendaku in surnames: a linguistic study using social media. *WCCFL* **34**. 519–528.

Tanaka, Yu (2020). Testing Rosen's Rule yet again: an experimental study. In Michael Barrie (ed.) *Japanese/Korean linguistics*. Vol. 27. Stanford: CSLI Publications. 355–364.

Tanaka, Yu & Ayaka Sugawara (2018). Revisiting accent in Japanese given names: stem-like accent with foot faithfulness. In Céleste Guillemot, Tomoyuki Yoshida & Seunghun J. Lee (eds.) *Proceedings of the 13th Workshop on Altaic Formal Linguistics*. 217–228.

Toda, Ayako (1988). Wago no hi-rendaku kisoku to rendaku keikō: Nippo Jisho to Waei Gorin Shūsei kara [The rendaku-inhibiting rules and tendencies for rendaku application in native Japanese words: an examination of the Vocabulario da Lingoa de Iapam and Hepburn's Japanese-English and English-Japanese Dictionary]. *Dōshisha Kokubungaku* **30**. 80–98.

Toyoda, Takeshi (1971/2012). *Myōji no rekishi [The history of surnames].* [Originally published in 1971 by Chuo Koronsha]. Tokyo: Yoshikawa Kobunkan.

Tsuzuki, Masaki & Hyun-Bok Lee (1992). A phonetic study of the Korean and Japanese lateral, flap and nasal. In The Linguistic Society of Korea (ed.) *Proceedings of the 1992 Seoul International Conference on Linguistics*. Seoul: Koryo University. 761–780.

Ueda, Kazutoshi (1898). Gogaku sōken: p-on-kō [A new linguistic perspective: on the sound /p/]. *Teikoku Bungaku* **4**. 41–46.

Unger, J. Marshall (2004). Alternations of *m* and *b* in Early Middle Japanese: the deeper signifiance of the sound-symbolic stratum. *Japanese Language and Literature* **38**. 323–337.

Unger, Marshall (1977). *Studies in Early Japanese morphophonemics*. Bloomington: Indiana University Linguistics Club.

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

Vance, Timothy J. (2007). Have we learned anything about rendaku that Lyman didn't already know? In Bjarke Frellesvig, Masayoshi Shibatani & John Charles Smith (eds.) *Current issues in the history and structure of Japanese*. Kurosio Publishers. 153–170.

Vance, Timothy J. (2008). *The sounds of Japanese*. Cambridge: Cambridge University Press.

Vance, Timothy J. (2014). If rendaku isn't a rule, what in the world is it? In Kaori Kabata & Tsuyoshi Ono (eds.) *Usage-based approaches to Japanese grammar: towards the understanding of human language*. Amsterdam & Philadelphia: John Benjamins. 137–152.

Vance, Timothy J. (2015a). Rendaku. In Haruo Kubozono (ed.) *The handbook of Japanese phonetics and phonology*. Berlin: De Gruyter Mouton. Handbooks of Japanese language and linguistics 2. 397–441.

Vance, Timothy J. (2015b). Rendaku no fukisokusei to Rōzen no hōsoku [Rendaku's irregularities and Rosen's Rule]. *NINJAL Research Papers* **9**. 207–214.

Vance, Timothy J. & Atsushi Asai (2016). Rendaku and individual segments. In Vance & Irwin (2016). 119–137.

Vance, Timothy J. & Mark Irwin (2013). A rendaku database for Old Japanese. Paper presented at the 21st International Conference on Historical Linguistics, Oslo.

Vance, Timothy J. & Mark Irwin (eds.) (2016). *Sequential voicing in Japanese: papers from the NINJAL Rendaku Project*. Amsterdam & Philadelphia: John Benjamins.

Vance, Timothy J., Emiko Kaneko & Seiji Watanabe (eds.) (2017). *Rendaku no kenkyū [Research on rendaku].* Tokyo: Kaitakusha.

Vance, Timothy J., Shigeto Kawahara & Mizuki Miyashita (2021). The diachronic origins of Lyman's Law: evidence from phonetics, dialectology and philology. *Phonology* **38**. 479–511.

Weijer, Jeroen van de, Kensuke Nanjo & Tetsuo Nishihara (eds.) (2005). *Voicing in Japanese*. Berlin & New York: Mouton de Gruyter.

White, James (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition* **130**. 96–115.

Wilson, Colin (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* **30**. 945–982.

Yamaguchi, Yoshinori (1988). Kodai-go no fukugō-go ni kansuru ichi-kōsatsu: rendaku o megutte [A study of compounds in Old Japanese: on rendaku]. *Nihongogaku* **7**. 4–12.

Yip, Moira (1998). Identity avoidance in phonology and morphology. In Steven G. LaPointe, Diane K. Brentari & Patrick M. Farrell (eds.) *Morphology and its relation to phonology and syntax*. Stanford: CSLI Publications. 216–246.

Zamma, Hideki (2005). The correlation between accentuation and rendaku in Japanese surnames: a morphological account. In van de Weijer *et al.* (2005). 157–176.

Zamma, Hideki & Atsushi Asai (2017). Sei ni mirareru Sugitō no hōsoku to kakuchōban Raiman no hōsoku ni kansuru keitai-teki/onin-teki kōsatsu [Morphological and phonological analyses of Sugito's Law and Strong Lyman's Law seen in surnames]. In Vance *et al.* (2017). 147–179.

Zhang, Jie & Yuwen Lai (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology* **27**. 153–201.