Stochastic phonological knowledge and word formation in Japanese

Gakuji Kumagai and Shigeto Kawahara
NINJAL and Keio University

January 2018
Gengo Kenkyu [Journal of the Linguistic Society of Japan], in press

Abstract: The question of whether linguistic knowledge is binary (i.e., grammatical vs. ungrammatical) or stochastic is one of the most important questions in general linguistic inquiry. Much recent work in the last few decades has argued that phonological knowledge is stochastic (e.g., Hayes & Londe 2006). Building on this body of research, we show that in Japanese, gradient phonological knowledge affects several word formation patterns in stochastic ways. Concretely, we show that identity avoidance effects hold at both the segmental and the CV-moraic levels and stochastically affect two types of word formation patterns in Japanese: group name formation and rendaku. We show that Maximum Entropy Grammar (Goldwater & Johnson 2003), together with multiple OCP constraints (Coetzee & Pater 2008), successfully models both of the observed morphological word formation patterns without any further stipulation. In addition to this theoretical contribution, one of the patterns discussed in this paper—group name formation—has not been analyzed from the perspective of formal phonological theories before, and hence this paper has descriptive novelty as well.*

Key words: word formation; Maximum Entropy Grammar; OCP effects; sonority; rendaku

1. Introduction

Whether linguistic knowledge is dichotomous/binary (grammatical vs. ungrammatical) or can be gradient is one of the most important questions in current linguistic inquiry. At the outset of the generative enterprise, sentences were divided into two distinct sets: those that could be generated by the posited grammar ("grammatical sentences"), and those that could not ("ungrammatical sentences") (Chomsky 1957). In reality, however, acceptability judgment patterns in syntax often show gradient patterns, as indicated by the common use in the syntactic literature of a variety of prefixal diacritic symbols (?, ??, ???, ?*,*?,*) showing different degrees of (un)grammaticality in sentential judgments (see e.g., Chomsky 1965; Lasnik 2004; Lasnik & Saito 1984; Pullum 2013a, b; Schütze 1996, 2016; Sprouse 2015, among many others).

However, it is still debated whether syntactic knowledge itself is dichotomous, making a fundamental grammatical vs. ungrammatical distinction; some researchers argue that grammar/competence makes only a binary distinction (yes grammatical vs. no grammatical), and it is other cognitive processes classified as performance that yield graded judgments (e.g., Neeleman 2013; Schütze 1996, 2016; Sprouse 2007a, b). Other researchers, including Adli (2010), Bresnan and Hay (2008), Keller (2006), Lasnik (2000), Lasnik and Saito (1984), Pullum (2013a, b), and Sorace and Keller (2005), accept the thesis that syntactic knowledge itself can be gradient and maintain that linguistic models should be able to capture this gradiency. Specific proposals have been put forward to capture the gradient nature of syntactic knowledge, such as Linear Optimality Theory (Keller 2006) and Model Theoretic Syntax (Pullum 2013a, b).

As with generative syntax, generative phonology began with the assumption that phonological knowledge is binary; a famous example is that whereas *brick* and *blick* are well-formed in English, *bnick* is not (Halle 1978). One of the fundamental tenets of early generative phonology is that phonological grammar should be able to capture this binary, grammatical vs. ungrammatical distinction between possible and impossible words (rather than existing and non-existing words). However, it has become increasingly clear that phonological knowledge is, if not entirely, at least partly stochastic, that is, probabilistic rather than a simple matter of possible vs. impossible. (see also Cohn 2006 and Pierrehumbert 1997 for historical reviews). First, phonotactic judgment patterns have now long been known to be stochastic; i.e., the intuition about whether a particular string can be a word is usually not a matter of a yes/no dichotomy. This gradient nature of phonotactic judgments was shown, for example, by the word-likeliness judgment experiment reported in Greenberg and Jenkins (1964). For instance, native speakers of English tend to judge [klæb] to be more natural—or more "English-sounding"—than [kleb], although both forms should be "grammatical" in English. It is also known that consonant clusters with a sonority plateau (e.g., [bdif]) are judged by English speakers to be better than clusters with falling sonority (e.g., [lbif]), despite the fact that both types of clusters should be "ungrammatical" in English (Berent et al. 2007 *et seq*.). See Shademan (2007) and Daland et al. (2011) for recent extensive results showing gradient phonotactic judgment patterns in English and a relevant discussion of the gradient nature of phonotactic knowledge.

Another well-known type of gradient phonotactics is the pattern of similarity avoidance, found in many Semitic languages, in which pairs of similar adjacent consonants are underrepresented in their lexicon. In the similarity avoidance pattern, the more similar two paired consonants are, the less likely it is that that pair exists in the lexicon (Frisch et al. 2004). These sorts of gradient phonotactic identity avoidance effects have been observed in many languages besides Semitic languages, including English (Berkley 1994), Muna (Coetzee & Pater 2008), Russian (Padgett 1992), and native words in Japanese (Kawahara et al. 2006), among others (see also Alderete & Frisch 2007; Yip 1998; Zuraw & Lu 2009 for other cases of identity avoidance). In short, phonotactic distribution patterns, as well as native speakers' judgments on word-likeliness, can undoubtedly be gradient, and thus cannot be reduced to a yes/no dichotomy. This observation led to the recent development of theories with numerically weighted constraints, such as Harmonic Grammar (Coetzee & Pater 2008) and MaxEnt Grammar (Goldwater & Johnson 2003; Hayes & Wilson 2008). Hayes and Wilson (2008: 382) explicitly declare that they "consider the ability to model gradient intuitions to be an important criterion for evaluating phonotactic models" Gradiency in phonotactics is now generally considered an essential aspect of grammar that any grammatical theory is required to capture, at least in phonology.

What has been less clear is whether phonological *alternations* can show systematic stochastic variations. However, recent work again demonstrates that some phonological alternations show patterned, stochastic variations (e.g., Boersma & Hayes 2001; Hayes 2017; Hayes & Londe 2006; McPherson & Hayes 2016; Moore-Cantwell & Pater 2016; Zuraw 2000, 2010). For example, Hayes and Londe (2006), in a paper titled "Stochastic Phonological Knowledge," have demonstrated that the probabilities of suffixes undergoing vowel harmony in Hungarian are different for different suffixes, and their likelihood of undergoing vowel harmony is affected by various phonological considerations. Zuraw (2000, 2010) shows that in Tagalog, different segments undergo nasal substitution with different probabilities in the lexicon, and that native speakers are sensitive to these gradient—yet regular—patterns, when they are tested with nonce words. These phonological patterns are not only *optional* but *systematic* in the sense that their patterns make phonological sense (see Hayes 2017 for recent discussion). Although the issue of whether or not phonological alternations can be systematically stochastic may be less well-established than the issue of the gradient nature of phonotactics, in the last few decades we have witnessed a growing body of evidence that suggests the stochastic nature of phonological alternation patterns. One impetus for our research is to add more case studies to address the question of the gradience of phonological alternations.

Before delving into our own case studies, we would like to address one question raised by two anonymous reviewers. So far, we have described the situation as a strict dichotomy: "grammar is binary" vs. "grammar is stochastic." However, there is a third position: grammar can be both binary and stochastic. One instantiation of the view is rather simple and non-problematic. There are many non-stochastic phonological restrictions; for example, neither English nor Japanese stochastically allows clicks in their speech sounds. Similarly, neither language uses front rounded vowels. In this sense, we fully admit that phonological knowledge can sometimes be categorical. As we will see below, the analytical tool that we employ in the paper, MaxEnt model, can yield both stochastic and non-stochastic patterns.

There is a different version of the third position. For the sake of discussion, let us take Sprouse's (2007a, b) view that grammar (or competence) must be binary, but performance gives the stochastic flavor to our linguistic behavior (see also Neeleman 2013; Schütze 1996, 2016). We will return to this issue once we present our experimental results, but our objection to this idea in short is that each phonological constraint affecting the word formation patterns in question is undoubtedly a matter of competence in nature and should not be relegated to a matter of performance. We thus disagree with the view that gradience simply arises in performance. In short, our claim is that at least part of phonological knowledge (or competence, for that matter) can be stochastic.

With these theoretical issues in mind, this paper offers two new pieces of evidence for stochastic phonological knowledge from Japanese, both of which affect word formation patterns. To the best of our knowledge, the issue of stochastic phonological knowledge has not been seriously tested using Japanese (except perhaps in a few works such as Kawahara 2013, Kilbourn-Ceron & Sonderegger 2017, and Tanaka 2017). Moreover, the current paper shows that such patterns can be successfully analyzed using Maximum Entropy (MaxEnt) Grammar (e.g., Colavin et al. 2014; Goldwater & Johnson 2003; Hayes 2017, Hayes & Wilson 2008; Hayes, Zuraw, Siptar & Londe 2009; Hayes et al. 2012; Jäger & Rosenbach 2006; Kumagai 2017; Martin 2011; McPherson & Hayes 2016; Shih 2016; Shih & Inkelas 2016; Tanaka 2017; White 2017; Wilson 2006; Zhang et al. 2011; Zuraw & Hayes 2017) by positing multiple OCP constraints (Coetzee & Pater 2008). Again, this paper is one of the first attempts to fit a MaxEnt grammar to Japanese data (though see also Tanaka 2017).[1]

The first case study, developed in Section 2, deals with the formation of names for a group consisting of two members created by combining the name of each member. As far as we know, this paper is the first attempt to describe and analyze this word formation pattern in the formal linguistic literature. Japanese speakers sometimes make up a group name for a pair of people. For example, the group consisting of two identical twin sister actresses, *mana* and *kana*, is called *mana-kana*. The current project started with the simple question of why the group name is *mana-kana* instead of *kana-mana*. Our hypothesis is that phonological considerations affect the formation of such group names. For example, *ka**na**-**mana*** is disfavored because of the three consecutive CV-moras[2] with nasal onset. This is reminiscent of the blockage of -*ly* adverb formation in English, in which -*ly* cannot be attached to roots that already end in -*ly* (e.g., **friendly-ly* and **silly-ly*: Katamba 1993). Shih (2014) likewise showed through a corpus study that in English names, name pairs are subject to a similar phonotactic restriction such that, for example, *Josh Smith* [ʃ-s] is less likely to occur than *Jack Smith* [k-s] as a full name (see also Yip 1998 for other similar cases). Shih and Zuraw (to appear) show that avoidance of a sequence of nasals can affect or even determine the inherently variable adjective-noun word ordering in Tagalog (e.g., *magandá* 'beautiful' + *babáe* 'woman' + -*ng* (LINK) → *magandá-ng babáe*/ *babáe-ng magandá* 'beautiful woman'). Their corpus study shows that when the nasal-initial linker -*ng* or *na* is inserted between an adjective and a noun, the word that follows it is more likely to begin with a non-nasal; for example, the order *manggá-ng diláw* 'mango-LINK yellow' is more frequent than the opposite, *diláw* ***na*** ***m****anggá* 'yellow-LINK mango'.

The experiment reported below in Section 2 is designed to test the hypothesis that identity avoidance constraints help determine the order of two elements. The results show that identity avoidance restrictions do indeed affect group name formation patterns, although it is not the case that names violating the identity avoidance constraint are categorically prohibited. To model the results, we develop a MaxEnt analysis and demonstrate that positing multiple OCP constraints following Coetzee and Pater (2008) successfully models the results without further stipulation.

The second type of word formation that this paper explores in depth is rendaku in Section 3, which is a well-studied morphophonological process. Rendaku is the phenomenon in which initial voiceless obstruents of the second member of a compound appear as voiced (e.g., /nise+**t**anuki/ → /nise+**d**anuki/ 'fake raccoon') (McCawley 1968; Tanaka 2017; Vance 1980, 1987, 2015; Vance & Irwin 2016, among many others; see Irwin 2016 for an extended bibliography). We build upon the results of Kawahara and Sano (2016), who show that identity avoidance restrictions apply stochastically to the application of rendaku in nonce words. Kawahara and Sano (2016) demonstrated with a nonce-word experiment that the more similar the pairs of segments that rendaku creates, the less likely it is to apply. In one condition of their experiment, two consonants across the word boundary were identical after rendaku applied (e.g., schematically, /i**g**a+**g**omoke/ from /i**g**a/+/**k**omoke/); in the other condition, the two consonants across the word boundary are not identical, even after rendaku applies (e.g., schematically, /i**ga**+**da**niro/ from /i**ga**/+/**ta**niro/). The results show that rendaku was less likely to occur when it resulted in consecutive identical consonants than under the control condition in which no identity violations were involved (that is, forms like /i**g**a+**g**omoke/ are avoided); furthermore, the applicability of rendaku was even more reduced when rendaku resulted in adjacent identical CV-moras (that is, forms like /i**ga**+**ga**niro/ are even more strongly avoided). Importantly, it is not the case that either of the identity avoidance constraints blocks rendaku entirely; they reduce the probability of rendaku applying. As is the case with group name formation, these results can be modeled by multiple OCP constraints and a MaxEnt grammar. This analysis supports the generality of the analysis that we develop in Section 2.

To summarize, in this paper we show empirically that phonological knowledge can stochastically and systematically affect Japanese word formation patterns beyond a dichotomous grammatical vs. ungrammatical distinction, and that theoretically, a MaxEnt grammar is a useful tool with which to model that stochastic knowledge.[3] We also emphasize the descriptive value of what we report in Section 2, which has hitherto not been analyzed in the theoretical literature.

2. Group name formation in Japanese

2.1. Background

This section explores the compound formation pattern of group names in which two names are combined. As mentioned in Section 1, the pair of Japanese identical twin sister actresses, *mana* and *kana*, is called *mana-kana*. Another example is a pair of two Japanese ping-pong players, *mima* and *miu*, which is *miu-mima, not *mima-miu*. In both of these examples, the possible-yet-unattested forms—*ka**n**a#**m**a**n**a* and ***m**i**m**a#**m**iu*—contain three onset nasal consonants across the word boundary, whereas the attested examples—***m**a**n**a#ka**n**a* and ***m**i_u#**m**i**m**a*—contain no sequence of onset nasal consonants across the word boundary.[4] A further example came to our attention during our revision phase of the paper: a new pair of ping-pong players, *mima* and *hina*, whose group name is ***m**i**m**a-hi**n**a* instead of *hi**n**a-**m**i**m**a*. In the rest of this section, such a sequence of nasals is referred to as presenting nasal clash (cf. “stress clash”: Prince 1983).

We experimentally examine whether nasal clash generally affects compound formation patterns in Japanese. We also examine whether degrees of similarity (e.g., /m/-/m/ vs. /m/-/n/) matter. Previous studies (e.g., Coetzee & Pater 2008; Frisch et al. 2004; Kawahara & Sano 2016) have shown that the more similar sequences are, the more strongly they are disfavored; hence it is predicted that the degree of similarity should impact the formation of Japanese group names as well. On the other hand, in some languages, total identity has been found to provide "an escape hatch" for similarity avoidance restrictions (e.g., Berent & Shimron 1997; Frisch et al. 2004; Kawahara et al. 2006), and hence it may be the case that an /m/-/m/ pair may be favored over an /m/-/n/ pair. This is an empirical question that remains unsettled in the phonology of Japanese (though see Kawahara et al. 2006 and Kawahara & Sano 2016 for discussion).

Going beyond the segmental level, we also test identity effects in the CV-mora. Recall that in Kawahara and Sano's (2016) experiment, rendaku was more likely to be blocked when it resulted in CV moraic identity (e.g., *[… **ga-ga**…]) than when it resulted in mere consonantal identity (e.g., *[**g**a…**g**o]). Therefore, Japanese speakers may disfavor a sequence of two identical CV moras in general, which may affect group name formation as well.

Although an inquiry into the nasal clash effect—more generally, the effect of similarity avoidance—is the main focus of this paper, another phonological factor taken into consideration in this experiment is sonority (e.g., Clements 1990; Kenstowicz 1994; Parker 2002, 2011): In the general sonority hierarchy, although some details are debated, segments are ordered as follows: stop < fricative < nasal < liquid < glide. In English, when two words are combined with *and*, the word with the more sonorous onset tends to come first.[5] Some existing examples include, for example, ***l**ovey-**d**ovey*, ***w**alkie-**t**alkie*, and ***w**illy-**n**illy* (Parker 2002: 246). Parker (2002) experimentally examined this tendency by presenting participants with several pairs of compounds such as *weeby-leeby* and *leeby-weeby.* The results showed that *weeby-leeby* was indeed preferred to *leeby-weeby*, which suggests that English speakers prefer to place the word with the more sonorous consonant at the beginning of the derived word.[6] Given this observation, we needed to make sure that the preference for *mana-kana* over *kana-mana* does not (solely) come from a sonority-based preference rather than an avoidance of consecutive nasal onset consonants; it could be the case that Japanese speakers, just like English speakers, may order names in such a way that more sonorous consonants are placed word-initially, which would result in a preference for *mana-kana* over *kana-mana*, although this sonority-based theory cannot explain the *miu-mima* example.

To summarize, in this experiment we examine whether various similarity-related factors affect word formation patterns in Japanese; in particular, (i) whether nasal clash is avoided, and if so, (ii) whether the number of nasal clashes matters, (iii) whether consonantal identity and moraic identity show different degrees of influence, and in addition, (iv) whether, as with English, sonority matters when speakers combine two words to make a larger word. In what follows, we express general nasal clash as the effects of OCP(nasal), nasal clash with identical nasal consonants as OCP(C), and nasal clash in identical CV moras as OCP(CV), respectively (where OCP = the Obligatory Contour Principle: Goldsmith 1976; Leben 1973; McCarthy 1986).

2.2. Stimuli

The current experiment used disyllabic Japanese girls' names as stimuli. All of the names used were existing (or at least possible) names.[7] Sets 1 and 3 consisted of pairs that could result in two nasals in sequence, either non-identical (e.g., *ha**n**a-**m**oka*), or identical (e.g., *ha**n**a-**n**iko*). Sets 2 and 4 consisted of pairs that could result in three nasals in sequence (e.g., *ha**n**a-**m**o**n**a* and *ku**m**i-**m**i**n**a*).

Table 1. The overall stimulus structure

	Number of nasals	Non-nasal segment
Set 1	2	obs
Set 2	3	obs
Set 3	2	son
Set 4	3	son

The number of nasal consonants involved in nasal clash was included as a condition in the experiment, because, as in the case of the *mana-kana* and *miu-mima* examples, it may be the sequence of three consecutive nasal onset consonants that makes the unattested *ka**n**a-**m**a**n**a* and ***m**i**m**a-**m**iu* unviable options; we were interested in whether two consecutive nasal onset consonants were sufficient to affect group name formation patterns.

Sets 1 and 2 consisted of pairs in which one word begins with an obstruent and the other with a nasal (e.g., *ha**n**a* and ***m**oka*),[8] and Sets 3 and 4 consisted of pairs in which one word begins with a liquid and the other with a nasal (e.g., *ri**n**a* and ***m**oka*). Recall that we wanted to tease apart the effects of identity avoidance and sonority.

Within each set, there were three conditions that were characterized in terms of different OCP violation profiles (i.e., OCP(nasal); OCP(nasal)+OCP(C); OCP(nasal)+

OCP(C)+OCP(CV)). In Set 1, shown in Table 2, the first syllable of one word had a nasal onset, and the second syllable of the other word had a nasal onset (e.g., ***m****oka* and *ha****n****a*). The word that did not begin with a nasal began with an obstruent (e.g., ***h****ana*). The condition in Table 2a was used to test whether the violation of OCP(nasal) is avoided. If *moka-hana* is preferred over *ha****n****a-****m****oka*, this would indicate that nasal clash (i.e., …***na****-****mo***…) is avoided. The condition in Table 2b was used to test the effects of identical consonants in addition to the occurrence of two nasals, i.e., the effects of OCP(C). Given *niko* and *hana*, *ha****n****a-****n****iko* has a sequence of identical nasals (i.e., …***na****-****ni***…), thus violating OCP(C) in addition to OCP(nasal). The condition in Table 2c was used to test the OCP(CV) in addition to the OCP(C) and OCP(nasal). If *natu-hana* is favored over *ha****na-na****tu*, this might indicate an avoidance of an identical mora across the word boundary (i.e., ...***na****-****na***...). There are four possible combinations for each condition, and thus Set 1 consists of 12 combinations in total, as shown in Table 2.

Table 2. Set 1: Two nasals (M = /m/; N = /n/; O = an obstruent; R = a sonorant). Sequences with nasal clash are underlined.

	α	+	*β*	→	*α-β or β-α*
a.	*moka* (MO)	+	*hana* (ON)	→	***m****oka-ha****n****a* (MOON) or *ha****n****a-****m****oka* (ONMO)
	moka (NO)	+	*kana* (ON)	→	***m****oka-ka****n****a* (MOON) or *ka****n****a-****m****oka* (ONMO)
	natu (NO)	+	*kumi* (OM)	→	***n****atu-ku****m****i* (NOOM) or *ku****m****i-****n****atu* (OMNO)
	natu (NO)	+	*fumi* (OM)	→	***n****atu-fu****m****i* (NOOM) or *fu****m****i-****n****atu* (OMNO)
b.	*niko* (NO)	+	*hana* (ON)	→	***n****iko-ha****n****a* (NOON) or *ha****n****a-****n****iko* (ONNO)
	niko (NO)	+	*kana* (ON)	→	***n****iko-ka****n****a* (NOON) or *ka****n****a-****n****iko* (ONNO)
	moka (MO)	+	*kumi* (OM)	→	***m****oka-ku****m****i* (MOOM) or *ku****m****i-****m****oka* (OMMO)
	moka (MO)	+	*fumi* (OM)	→	***m****oka-fu****m****i* (MOOM) or *fu****m****i-****m****oka* (OMMO)
c.	*natu* (NaO)	+	*hana* (ONa)	→	***n****atu-ha****n****a* (NaOONa) or *ha****n****a-****n****atu* (ONaNaO)
	natu (NaO)	+	*kana* (ONa)	→	***n****atu-ka****n****a* (NaOONa) or *ka****n****a-****n****atu* (ONaNaO)
	mika (MiO)	+	*kumi* (OMi)	→	***m****ika-ku****m****i* (MiOOMi) or *ku****m****i-****m****ika* (OMiMiO)
	mika (MiO)	+	*fumi* (OMi)	→	***m****ika-fu****m****i* (MiOOMi) or *fu****m****i-****m****ika* (OMiMiO)

Set 2, shown in Table 3, was prepared to examine whether three consecutive nasals would be avoided more strongly than two consecutive nasals. Sequences with different OCP violation profiles were also examined, as in Set 1. The nasal clash in Table 3a violates only OCP(nasal), the nasal clash in Table 3b violates OCP(nasal) and OCP(C), and the nasal clash in Table 3c violates all three, OCP(nasal), OCP(C), and OCP(CV).

Table 3. Set 2: Three nasals (M = /m/; N = /n/; O = an obstruent; R = a sonorant). Sequences with nasal clash are underlined.

	α	+	*β*	→	*α-β or β-α*
a.	*mona* (MN)	+	*hana* (ON)	→	*mona-hana* (MNON) or *hana-mona* (ONMN)
	mona (MN)	+	*kana* (ON)	→	*mona-kana* (MNON) or *kana-mona* (ONMN)
	nami (NM)	+	*kumi* (OM)	→	*nami-kumi* (NMOM) or *kumi-nami* (OMNM)
	nami (NM)	+	*fumi* (OM)	→	*nami-fumi* (NMOM) or *fumi-nami* (OMNM)
b.	*nina* (NN)	+	*hana* (ON)	→	*nina-hana* (NNON) or *hana-nina* (ONNN)
	nina (NN)	+	*kana* (ON)	→	*nina-kana* (NNON) or *kana-nina* (ONNN)
	mona (MN)	+	*kumi* (OM)	→	*mona-kumi* (MNOM) or *kumi-mona* (OMMN)
	mona (MN)	+	*fumi* (OM)	→	*mona-fumi* (MNOM) or *fumi-mona* (OMMN)
c.	*nami* (NaM)	+	*hana* (ONa)	→	*nami-hana* (NaMONa) or *hana-nami* (ONaNaM)
	nami (NaM)	+	*kana* (ONa)	→	*nami-kana* (NaMONa) or *kana-nami* (ONaNaM)
	mina (MiN)	+	*kumi* (OMi)	→	*mina-kumi* (MiNOMi) or *kumi-mina* (OMiMiN)
	mina (MiN)	+	*fumi* (OMi)	→	*mina-fumi* (MiNOMi) or *fumi-mina* (OMiMiN)

In Sets 3 and 4, shown in Tables 4 and 5, the word listed in β begins with a sonorant rather than an obstruent. If there is a sonority-driven word-ordering preference in Japanese, we would expect to observe different results between Sets 1 and 2 on the one hand and Sets 3 and 4 on the other.

Table 4. Set 3: Two nasals (M = /m/; N = /n/; O = an obstruent; R = a sonorant). Sequences with nasal clash are underlined.

	α	+	*β*	→	*α-β or β-α*
a.	*moka* (MO)	+	*rina* (RN)	→	***m**oka-ri**n**a* (MORN) or *ri**n**a-**m**oka* (RNMO)
	moka (MO)	+	*rena* (RN)	→	***m**oka-re**n**a* (MORN) or *re**n**a-**m**oka* (RNMO)
	natu (NO)	+	*rumi* (RM)	→	***n**atu-ru**m**i* (NORM) or *ru**m**i-**n**atu* (RMNO)
	natu (NO)	+	*remi* (RM)	→	***n**atu-re**m**i* (NORM) or *re**m**i-**n**atu* (RMNO)
b.	*niko* (NO)	+	*rina* (RN)	→	***n**iko-ri**n**a* (NORN) or *ri**n**a-**n**iko* (RNNO)
	niko (NO)	+	*rena* (RN)	→	***n**iko-re**n**a* (NORN) or *re**n**a-**n**iko* (RNNO)
	moka (MO)	+	*rumi* (RM)	→	***m**oka-ru**m**i* (MORM) or *ru**m**i-**m**oka* (RMMO)
	moka (MO)	+	*remi* (RM)	→	***m**oka-re**m**i* (MORM) or *re**m**i-**m**oka* (RMMO)
c.	*natu* (NaO)	+	*rina* (RNa)	→	***n**atu-ri**n**a* (NaORNa) or *ri**n**a-**n**atu* (RNaNaO)
	natu (NaO)	+	*rena* (RNa)	→	***n**atu-re**n**a* (NaORNa) or *re**n**a-**n**atu* (RNaNaO)
	mika (MiO)	+	*rumi* (RMi)	→	***m**ika-ru**m**i* (MiORMi) or *ru**m**i-**m**ika* (RMiMiO)
	mika (MiO)	+	*remi* (RMi)	→	***m**ika-re**m**i* (MiORMi) or *re**m**i-**m**ika* (RMiMiO)

Table 5. Set 4: Three nasals (M = /m/; N = /n/; O = an obstruent; R = a sonorant). Sequences with nasal clash are underlined.

	α	+	*β*	→	*α-β or β-α*
a.	*mona* (MN)	+	*rina* (RN)	→	***m**o**n**a-ri**n**a* (MNRN) or *ri**n**a-**m**o**n**a* (RNMN)
	mona (MN)	+	*rena* (RN)	→	***m**o**n**a-re**n**a* (MNRN) or *re**n**a-**m**o**n**a* (RNMN)
	nami (NM)	+	*rumi* (RM)	→	***n**a**m**i-ru**m**i* (NMRM) or *ru**m**i-**n**a**m**i* (RMNM)
	nami (NM)	+	*remi* (RM)	→	***n**a**m**i-re**m**i* (NMRM) or *re**m**i-**n**a**m**i* (RMNM)
b.	*nina* (NN)	+	*rina* (RN)	→	***n**i**n**a-re**n**a* (NNRN) or *re**n**a-**n**i**n**a* (RNNN)
	nina (NN)	+	*rena* (RN)	→	***n**i**n**a-re**n**a* (NNRN) or *re**n**a-**n**i**n**a* (RNNN)
	mona (MN)	+	*rumi* (RM)	→	***m**o**n**a-ru**m**i* (MNRM) or *ru**m**i-**m**o**n**a* (RMMN)
	mona (MN)	+	*remi* (RM)	→	***m**o**n**a-re**m**i* (MNRM) or *re**m**i-**m**o**n**a* (RMMN)
c.	*nami* (NaM)	+	*rina* (RNa)	→	***n**a**m**i-ri**n**a* (NaMRNa) or *ri**n**a-**n**a**m**i* (RNaNaM)
	mina (MiN)	+	*remi* (RMi)	→	***m**i**n**a-re**m**i* (MiNRM) or *re**m**i-**m**i**n**a* (RMiMiN)
	mina (MiN)	+	*rumi* (RMi)	→	***m**i**n**a-ru**m**i* (MiNRM) or *ru**m**i-**m**i**n**a* (RMiMiN)
	mina (MiN)	+	*remi* (RMi)	→	***m**i**n**a-re**m**i* (MiNRM) or *re**m**i-**m**i**n**a* (RMiMiN)

2.3. Participants and procedure

A total of 83 naive native speakers of Japanese participated in the experiment. All of the participants were undergraduate students at a Japanese university. There was no overlap of participants between the current experiment and that reported in Section 3. In the instruction session, they were told that they were to make up a group name for a pair of girls. In the test session, they were given two names and asked to choose one of the two combined forms (e.g., "Given two personal names, *mana* and *kana*, which order would you use to make up a group name, *mana-kana* or *kana-mana*?"). All the names were written in the Japanese *katakana* orthography, which is commonly used to write personal names. There were a total of 48 questions (4 sets*12 combinations). The order of the questions was randomized.

2.4. Results

For statistical analysis, a generalized mixed-effects logistic regression was fit to the response using the `glmer` function in *R* (e.g., Baayen 2008). Subjects and items were coded as random effects. The first model included all the fixed factors (obs vs. son; two nasals vs. three nasals; OCP(C); OCP(CV)); specific follow-up comparisons were made based on contrast analyses using more specific logistic regression models. The resulting figures below show the ratios of the responses that contain nasal clash on the y-axis. The results for Sets 1 and 2 are shown in Figure 1 and those for Sets 3 and 4 are shown in Figure 2. Error bars represent 95% confidence intervals.

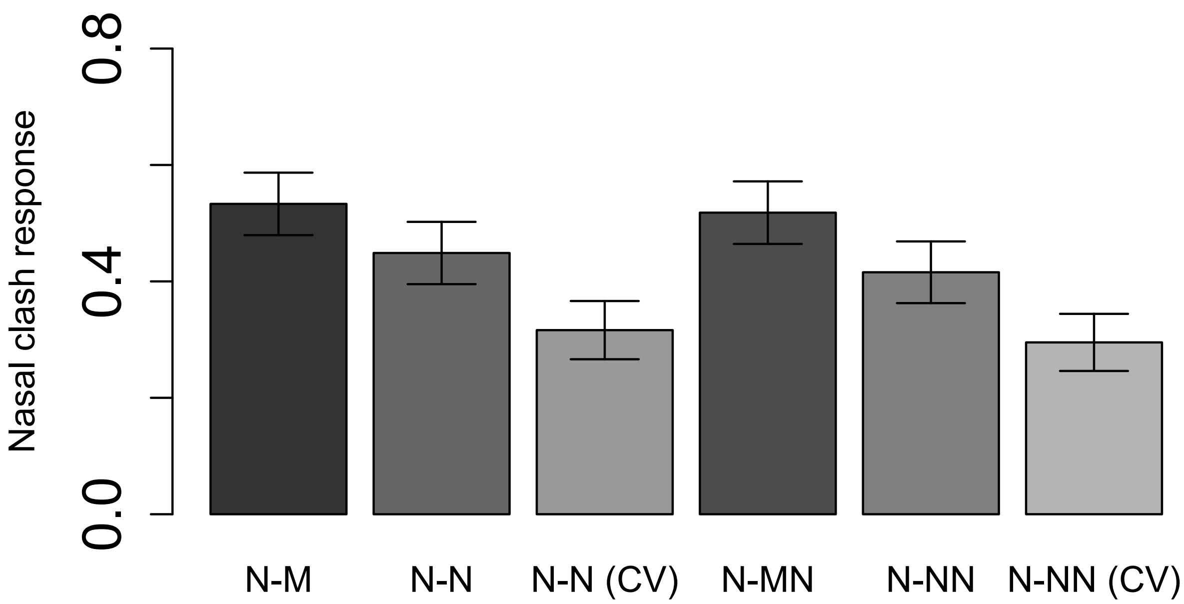

Figure 1: Nasal clash response ratio with 95% confidence intervals. Words that do not begin

with a nasal begin with an obstruent.

In Figure 1, the first three bars show cases in which *two* nasals are placed in adjacent syllables (e.g., *ha**n**a-**m**oka*), whereas the last three bars show cases in which *three* nasals are placed in proximity (*ha**n**a-**m**o**n**a*). Within each set, the three bars are ordered by the degree of similarity (non-identical nasals (N-M), identical nasals (N-N), identical CV moras with a nasal onset (N-N (CV)). The actual observed average values are: 0.53 vs. 0.45 vs. 0.32 for the first three bars and 0.52 vs. 0.42 vs. 0.30 for the last three. For the two-nasal condition (the leftmost three bars), there were significant differences between each condition: N-M vs. N-N, $z = -2.366$, $p < .05$; N-M vs. N-N (CV), $z = -6.035$, $p < .001$; N-N vs. N-N (CV), $z = -3.874$, $p < .001$. The same holds true of the three-nasal condition (the rightmost three bars) (N-MN vs. N-NN, $z = -2.885$, $p < .01$; N-MN vs. N-NN (CV), $z = -6.245$, $p < .001$; N-NN vs. N-NN (CV), $z = -3.618$, $p < .001$). We thus observe a clear tendency for avoidance of similar sequences. It is important to note here that the effects are gradient; we see a three-way distinction according to different violation profiles of OCP constraints. We maintain that this instantiates the effect of gradient phonological knowledge that affects the group name formation pattern.

There were no effects of the number of nasal consonants involved; i.e., there were no differences between corresponding bars in the first three and the last three bars ($z = 1.12$, *n.s.*). Finally, looking at the two N-M(N) conditions, the nasal clash response ratios are greater than 0.5 (i.e., 0.53 and 0.52), which is slightly higher than would be expected by chance. This may indicate that the avoidance of non-identical nasal consonants—OCP(nasal)—is not so strong as to show tangible effects in this experiment. The weak effect of OCP(nasal) will be made clearer in the MaxEnt analysis presented below, in which the weight of OCP(nasal) is low. As we will observe below, there may be a preference for less sonorous consonants to occur word-initially (Smith 2002), which would coerce nasal clash in this condition; i.e., ***h**ana-moka* is better than ***m**oka-hana* in that the former has a word-initial obstruent. This sonority-based effect may have “cancelled out” the effects of OCP(nasal).

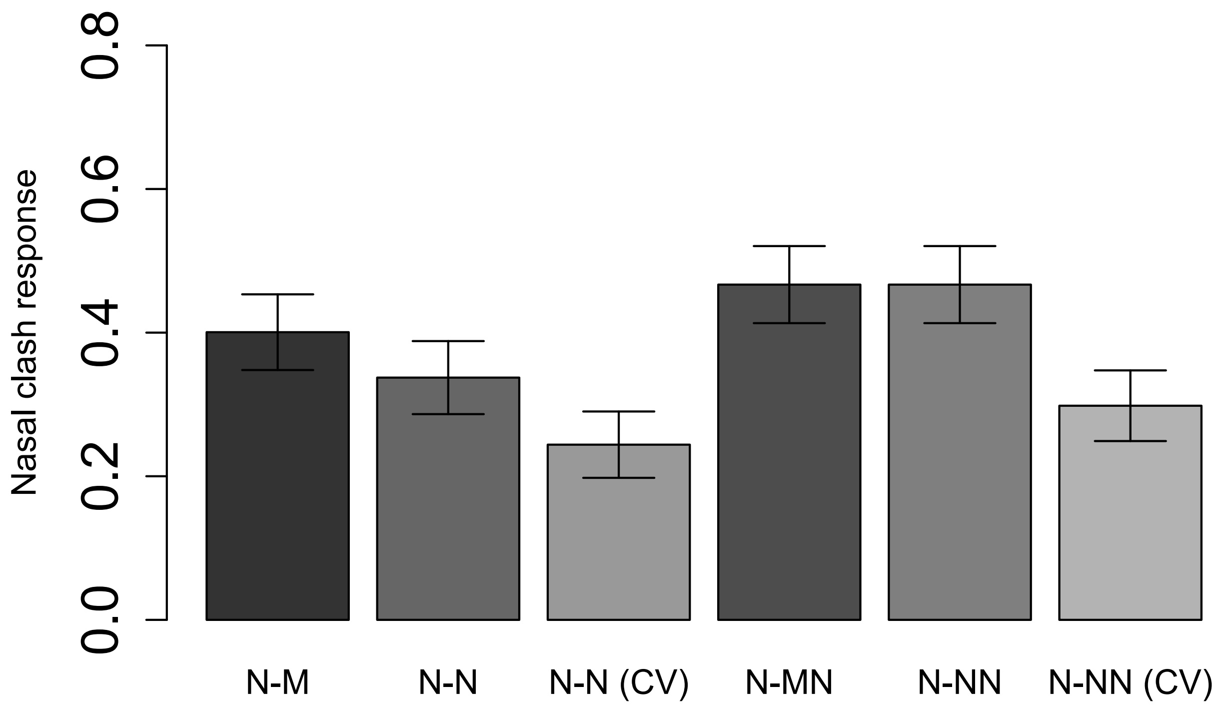

Figure 2: Nasal clash response ratios with 95% confidence intervals. Words that do not begin with a nasal begin with a sonorant.

The first three bars in Figure 2 show the two-nasal condition, in which there were significant differences between N-M and N-N (CV) ($z = -4.663$, $p < .001$) and between N-N and N-N (CV) ($z = -2.944$, $p < .01$) (0.40 vs. 0.34 vs. 0.24). Though the difference between N-M and N-N did not reach significance ($z = -1.852$, *n.s.*), it is in the expected direction. For the three-nasal condition (the rightmost three bars), there were also significant differences between N-MN and N-NN (CV) ($z = -4.919$, $p < .001$) and between N-NN and N-NN (CV) ($z = -4.956$, $p < .001$). However, there were no significant differences between N-MN and N-NN ($z = 0$, *n.s.*) (0.47 vs. 0.47 vs. 0.30); there were no obvious effects of OCP(C) in this context. Surprisingly, there were slightly more nasal clash responses when there were three nasal consonants than when there were only two ($z = 2.087$, $p < .05$) (here but not in Figure 1). We do not have a clear explanation of these unexpected results.

Comparing Figure 1 (the obstruent condition) and Figure 2 (the sonorant condition), the proportion of nasal clash is on average lower under the sonorant condition than the obstruent ($z = 3.189$, $p < .01$). This difference shows that Japanese speakers are more likely to tolerate nasal clash when it results in word-initial obstruents (e.g., ***h****ana-mona*) than liquids (e.g., ***r****ina-mona*). There are two possible reasons for this difference. The first possibility is that /r/ is avoided as a word-initial sound. This hypothesis is possible, as there are few Japanese native words that begin with /r/ (e.g., Labrune 2014), and hence there might be a constraint like *INITIAL-/r/ at work in Japanese phonology (Kawahara 2015). The second possibility is that, as was the case for Parker's (2002) experiments with English speakers, the ordering of the two words was affected by sonority considerations: obstruent-initial words were preferred to come before nasal-initial words, and nasal-initial words were preferred to come before liquid-initial words (see Smith 2002 for related observations). Under this interpretation, while English prefers more sonorous word-initial segments, Japanese prefers less sonorous word-initial segments. In the analysis that follows, we adapt the second explanation because it explains why there were no clear effects of OCP(nasal) in Figure 1.[9] Given this, we can assume that sonority preference and OCP(nasal) canceled each other out, resulting in near-chance performance.

To summarize, the results indicate that when Japanese speakers are asked to make a group name based on two names, various factors affect the ordering; (i) sequences of two identical nasals are avoided; (ii) sequences of identical CV-moras are avoided even more strongly; and (iii) the word with a lower sonority consonant is preferred word-initially. As we shall see, each of these factors can be represented by phonological constraints, and a MaxEnt analysis is suitable for modelling the overall results.

2.5. The MaxEnt analysis

To model the stochastic nature of the Japanese name ordering patterns observed in the experiment above, we used a MaxEnt grammar model (Hayes & Wilson 2008). MaxEnt is similar to Optimality Theory (OT: Prince & Smolensky 2004) in that a set of candidates is evaluated against a set of constraints. Unlike OT, however, the constraints are weighted (rather than ranked), as in Harmonic Grammar (HG: Legendre et al. 1990, 2006; Pater 2009, 2016; Potts et al. 2010). The probabilities of each candidate are assigned based on their constraint violation profiles. More specifically, for each candidate, the weighted constraint violations are summed to give its H(armonic)-score, which is mapped to probabilities in such a way that $P(cand_i)=exp(-H(cand_i))$, relatived to all the other candidates so that their probabilities sum to 1.

The procedure of calculating probabilities is as follows (Hayes 2017; Hayes et al. 2012; Hayes, Zuraw, Siptar & Londe 2009; Hayes & Wilson 2008; and Zuraw & Hayes 2017 in particular):

1) As in HG, for each candidate the harmonic score (H-score) is calculated as the sum of C_i*w_i, where the candidate's violation of each constraint (C_i) is multiplied by its weight (w_i);

2) Each candidate's "bare" probability is calculated as $e^{-(\text{H-score})}$;

3) The $e^{-(\text{H-score})}$ is summed over all candidates;

4) P(x), the predicted probability of candidate x, is its $e^{-(\text{H-score})}$ divided by the sum of $e^{-(\text{H-score})}$ of all of the candidates.

For analysis we used the MaxEnt Grammar Tool (Hayes, Wilson & George 2009), which calculates optimal weights for each constraint from the frequency distributions of the actual outcomes. To implement the MaxEnt analysis, we use the following four constraints. First, *SON(C_2) > SON(C_1) disfavors forms in which the second word begins with a less sonorous consonant than the first word (e.g., /m/ > /h/ in ***m**ona#**h**ana*; /r/ > /m/ in ***r**ina#**m**ona*). Second, OCP(nasal) is a constraint that is violated by two consecutive nasal consonants across a word boundary (e.g., *ha**n**a#**m**ona*; *ri**n**a#**m**ona*).[10] Since the experimental results did not show a substantial difference between sequences of two and three nasals, their violation profiles were not distinguished. Third, OCP(C) is violated if the two nasals across the word boundary are identical (e.g., *ku**m**i#**m**ona*; *ru**m**i#**m**ona*). Fourth, OCP(CV) is violated if there is a pair of adjacent identical CV-moras (e.g., *ha**na**#**na**mi*; *ri**na**#**na**mi*). The violation profiles of these constraints as well as the candidate sets fed to the MaxEnt Grammar Tool are shown in (1) and (2).

Table 6 shows the results for the constraint weights generated by the MaxEnt Grammar Tool. The MaxEnt analyses are given in (1) and (2),[11] and (3) and (4) compare the observed probabilities with those predicted by the MaxEnt Tool. We observe that the two probabilities are highly correlated, indicating the success of the MaxEnt analysis.[12]

Table 6: The constraints and their weights generated by the MaxEnt Grammar Tool

Constraints	Weight
*SON(C_2) > SON(C_1)	0.11
OCP (nasal)	0.082
OCP (C)	0.263
OCP (CV)	0.579

(1) MaxEnt analysis (the obstruent condition)

	*S(C_2) > S(C_1)	OCP (nasal)	OCP (C)	OCP (CV)				
weights	0.11	0.082	0.263	0.579		H-score	$e^{-(\text{H-score})}$	Predicted Prob.
mona + (hana/kana)								
mona # (hana/kana)	−1					−0.11	0.8958	0.493
(hana/kana) # mona		−1				−0.082	0.9213	0.507
mona + (kumi/fumi)								
mona # (kumi/fumi)	−1					−0.11	0.8958	0.5585
(kumi/fumi) # mona		−1	−1			−0.345	0.7082	0.4415
nami + (hana/kana)								
nami # (hana/kana)	−1					−0.11	0.8958	0.693
(hana/kana) # nami		−1	−1	−1		−0.924	0.3969	0.307

(2) MaxEnt analysis (the sonorant condition)

	*S(C_2) > S(C_1)	OCP (nasal)	OCP (C)	OCP (CV)				
weights	0.11	0.082	0.263	0.579		H-score	$e^{-(\text{H-score})}$	Predicted Prob.
mona + (rina/rena)								
mona # (rina/rena)						0	1	0.5479
(rina/rena) # mona	−1	−1				−0.192	0.8253	0.4521
mona + (rumi/remi)								
mona # (rumi/remi)						0	1	0.6118
(rumi/remi) # mona	−1	−1	−1			−0.455	0.6344	0.3882
nami + (rina/rena)								
nami # (rina/rena)						0	1	0.7377
(rina/rena) # nami	−1	−1	−1	−1		−1.034	0.3556	0.2623

(3) Sets 1 & 2

Forms	Observed Prob.	Predicted Prob.
mona # (hana/kana)	0.47	0.49
(hana/kana) # mona	0.53	0.51
mona # (kumi/fumi)	0.57	0.56
(kumi/fumi) # mona	0.43	0.44
nami # (hana/kana)	0.70	0.69
(hana/kana) # nami	0.30	0.31

(4) Sets 3 & 4

Forms	Observed Prob.	Predicted Prob.
mona # (rina/rena)	0.57	0.55
(rina/rena) # mona	0.43	0.45
mona # (rumi/remi)	0.60	0.61
(rumi/remi) # mona	0.40	0.39
nami # (rina/rena)	0.73	0.74
(rina/rena) # nami	0.27	0.26

2.6. Summary

In this section, we examined the group-name formation pattern in Japanese, in which two names are combined to form a group name. We observed that similarity avoidance plays a visible role in this word formation such that similarity at the word boundary is avoided, and the higher the similarity, the more strongly it is disfavored. In particular, sequences of two nasals and sequences of CV-moras with two identical nasals were particularly disfavored. Importantly, however, no phonological constraints were deterministic, i.e., inviolable. They simply reduced the probability of nasal clash. In this sense, identity avoidance constraints stochastically affect the word formation pattern. We modeled these gradient patterns using a MaxEnt grammar as well as different types of OCP constraints. We also found that Japanese speakers may prefer less sonorous consonants word-initially. Although this preference toward lower sonority has been observed cross-linguistically (Smith 2002), we believe that it is a new finding for Japanese.

3. Rendaku as evidence for stochastic phonological knowledge

3.1. Identity avoidance in rendaku

We next turn to the analysis of another word formation pattern, rendaku, that shows stochastic and systematic influences of identity avoidance constraints. This section analyzes the experimental data presented by Kawahara and Sano (2016) to show the generality of the constraints and analysis developed in Section 2. Before delving into the analysis, we first briefly review their experimental design and results.

The purpose of Kawahara and Sano (2016) was to examine whether identity avoidance blocks rendaku application. The set of stimuli in Table 7 was used to test the effects of identity avoidance at the consonantal level (i.e., OCP(C)), and that in Table 8 the effect of identity avoidance at the CV-moraic level (i.e., OCP(CV)). In each set, their stimuli contained four first elements (E1s) and three different second elements (E2s), the latter drawn from the set of consonants /k, t, s, h/ that may undergo rendaku, which yielded 12 E2s for each E1. There were thus 48 combinations in total.

Table 7: The list of the stimuli used in Set 1. All combinations of E1 and E2 (4 * 12 = 48) were tested. E2 were nonce words.

E1		E2	
/iga/	*	/keniro/	/komoke/ /korimo/
/aza/		/seniro/	/somoke/ /sorimo/
/kuda/		/teniro/	/tomoke/ /torimo/
/kaba/		/heniro/	/homoke/ /horimo/

Table 8: The list of the stimuli used in Set 2.

E1		E2	
/iga/	*	/kaniro/	/kamoke/ /karimo/
/aza/		/saniro/	/samoke/ /sarimo/
/kuda/		/taniro/	/tamoke/ /tarimo/
/kaba/		/haniro/	/hamoke/ /harimo/

The participants were 43 native speakers of Japanese who were undergraduate students of a Japanese university. None of them participated in the experiment presented in Section 2. The experiment was conducted online using SurveyMonkey. In the test, they were presented with two elements (E1 and E2) and two forms (rendaku and non-rendaku forms) and asked which was more natural; that is, it was a forced-choice wug test (Berko 1958). The stimuli were presented in hiragana, the standard Japanese orthography for native words (rendaku generally applies only to native words). The order of the stimuli was randomized. See Kawahara and Sano (2016) for further details.

Figure 3 shows the results of the applicability of rendaku under each condition. A significant difference was found between cases that violate CV-moraic identity avoidance and those that did not (0.27 vs. 0.44; $z = 5.32$, $p < .001$). The results also show that there was a significant difference between consonantal identity avoidance and the control group (0.39 vs. 0.45; $z = 2.23$, $p < .05$), as well as between moraic identity avoidance and consonantal identity avoidance ($z = 4.55$; $p < .001$), which suggests that the effect of identity avoidance is stronger at the CV-moraic level (the first bar) than at the consonantal level (the third bar).

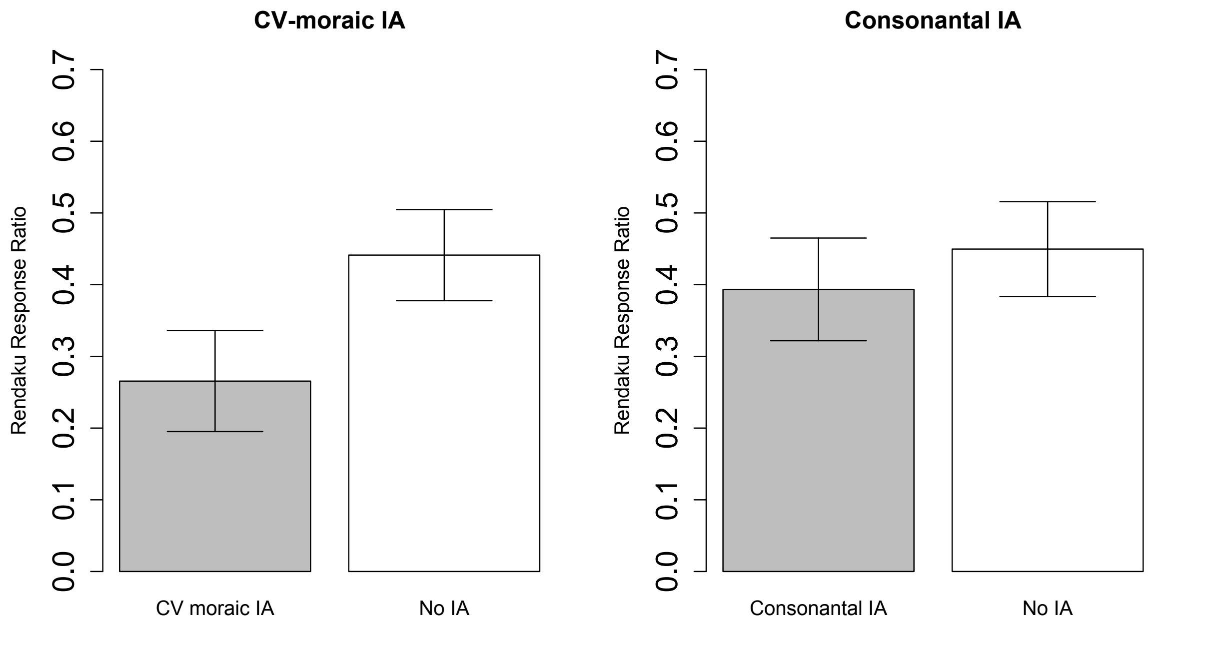

Figure 3: Proportion of rendaku application under each condition (adapted from Kawahara & Sano 2016)

To sum up, Kawahara and Sano (2016) showed that rendaku is less likely to occur when it results in identical consonants in adjacent moras. Furthermore, the applicability of rendaku was even more reduced when rendaku resulted in adjacent identical CV moras. These results show that the greater the similarity of the strings of segments that rendaku creates, the more likely it is to be avoided, which constitutes another case of gradient phonological effects on word-formation patterns. Like the case analyzed in Section 2, the effects of phonological constraints were stochastic; they did not deterministically block rendaku but merely reduced the probability of its application.

3.2. The MaxEnt analysis

For the current MaxEnt analysis of rendaku, we used four constraints. Following the most comprehensive OT analysis of Japanese rendaku, presented by Ito and Mester (2003), we use REALIZE MORPHEME (RM) and IDENT(voice); the former constraint encourages rendaku, assuming that rendaku is the realization of a compound juncture morpheme. IDENT(voice) disfavors rendaku, because rendaku changes the underlying specification of a [voice] feature. We also used OCP(C) and OCP(CV), defined in Section 2.

Like the MaxEnt analysis presented in Section 2, two candidates (rendaku and non-rendaku forms) were evaluated for each input form, with the violation profiles shown in (5). The results appear in Table 9 and (6). The MaxEnt Tool replicated the experimental results successfully with the multiple OCP constraints we posited in Section 2; the predicted probabilities are almost identical to the observed probabilities, as shown in (6).

(5) The MaxEnt analysis of rendaku

	RENDAKU	IDENT (voice)	OCP (C)	OCP (CV)		H-Score	$e^{-(\text{H-score})}$	Predicted Prob.
weights	4.89	5.1	0.24	0.60				
/…pa+ta…/								
…pata…	−1					−4.89	7.52*10^-4	0.55
… pada…		−1				−5.1	6.10*10^-4	0.45
/…ga+ko…/								
…gako…	−1					−4.89	7.52*10^-4	0.61
…gago…		−1	−1			−5.34	4.80*10^-4	0.39
/…ga+ka…/								
… gaka	−1					−4.89	7.52*10^-4	0.74
… gaga…		−1	−1	−1		−5.69	3.38*10^-4	0.26

Table 9. The posited constraints and obtained weights.

Constraints	Weight
RM	4.89
IDENT (voice)	5.1
OCP (C)	0.24
OCP (CV)	0.6

(6) Observed and predicted probabilities.

Forms	Observed Prob.	Predicted Prob.
…pa#ta…	0.55	0.55
…pa#da…	0.45	0.45
…ga#ko…	0.61	0.61
…ga#go…	0.39	0.39
…ga#ka…	0.74	0.74
…ga#ga…	0.26	0.26

4. Conclusions

The current paper explored a stochastic yet systematic aspect of Japanese word formation shown in group name formation and rendaku. In both types of word formation, sequences of two moras with the same nasal consonants are avoided, and sequences of two identical moras are avoided even more strongly. However, it is not that case that a violation of one of these constraints entirely dictates the word formation pattern; the effects of phonological constraints are probabilistic, suggesting that phonological constraints can impose stochastic influences on word formation.[13] We also showed that a MaxEnt grammar is a general, useful tool to model such stochastic patterns. Overall, this research contributes to the growing body of literature showing that phonological knowledge can be both stochastic and systematic.

Let us return to the view anticipated in the introduction that competence is binary and performance is gradient. This view is widely shared among generative grammarians, either implicitly or explicitly (Neeleman 2013; Schütze 1996, 2016; Sprouse 2007a, b).[14] We disagree with this view because the factors contributing to gradiency in nickname formation patterns and rendaku (various types of OCP constraints and sonority-driven ordering constraints) are undoubtedly matters that belong to competence and should not be relegated to performance. If we were to relegate gradiency to performance, we would have to consider the OCP a matter of performance, but it is hard to imagine that anybody is willing to do so.

In addition to the contributions this study makes to understanding the issue of gradiency, we would like to highlight the fact that it is the first to systematically analyze the formation of group names in Japanese from the perspective of formal phonological theory. The results in Section 2 show that this method is useful in revealing some aspects of phonological knowledge that Japanese speakers possess. In particular, we discovered that Japanese speakers may favor less sonorous consonants word-initially. We hope that this methodology will be used to explore the nature of other phonological and morphological patterns in other languages. In particular, since identity avoidance is observed across many languages, it is of interest to test the generality of how identity avoidance may affect the formation of new coordinate compounds like those tested in Section 2 of this paper (see also Shih 2014; Shih & Zuraw 2017).

The current study performed an experiment to examine OCP effects in Japanese group name formation; an interesting question that arises is whether the patterns we observe hold in existing words as well. Unfortunately, to the best of our knowledge, there is no large-scale corpus of Japanese group names like *mana-kana* consisting of two personal names. However, there is an alternative way to address the OCP effects in Japanese: Many Japanese names consist of two disyllabic Sino-Japanese morphemes, such as *kazu-taka,* where *kazu* and *taka* are Sino-Japanese morphemes. If OCP(CV) is an active constraint in Japanese phonology, the prediction is that the order *kazu-taka* is more frequent than *taka-kazu*, as the latter violates OCP(CV). A future study can use corpora of Japanese names to explore whether this prediction is borne out in order to further test the effect of OCP(CV) in Japanese.

Another limitation of this study is that we mainly explored the effects of similarity avoidance in sequences of nasal consonants (in the formation of group names) and voiced obstruents (in the analysis of rendaku).[15] We do not mean to imply that these two classes of sounds are particularly or uniquely susceptible to identity avoidance constraints. Further research should address the question of whether other types of segments (such as stridents and obstruents in general) can cause similar identity avoidance effects.[16] Shih and Zuraw (to appear) show that OCP(nasal) but not other types of OCP affects the variable ordering of adjective nouns in Tagalog. Addressing the matter of the kinds of OCP constraints that exist, how they affect our speech behavior, and why is a very exciting topic for future study.

Notes

[*] Acknowledgments: The present study was presented at the regular meeting of Tokyo Circle of Phonologists (University of Tokyo) and the 165th NINJAL Salon (NINJAL). We would like to thank Takeru Honma, Kohei Nishimura, Clemens Poppe, Stephanie Shih, Tim Vance, and two anonymous reviewers for helpful comments.

[1] One alternative for modeling gradient patterns is Noisy Harmonic Grammar (see Coetzee & Kawahara 2013), which was implemented in Praat (Boersma 2001) as early as 2006. We do not intend to engage in a comparison between MaxEnt analysis and other related frameworks in this paper. See Hayes and Wilson (2008: Section 9.1) and Hayes (2017) for an extended comparison of MaxEnt Grammar and other related constraint-based approaches.

[2] A CV-mora is a unit that plays an active role in Japanese phonology, orthography, and speech production and perception (see, e.g., Ito 1989; Kubozono 1989; Labrune 2012; Otake et al. 1993). Since all the stimuli discussed in this paper are light syllables, CV-moras can be considered identical here to light syllables. We use the term "CV-mora" following Kawahara and Sano (2016).

[3] An anonymous reviewer asked whether MaxEnt Grammar predicts that all phonological patterns must be stochastic. The answer is no. When the weights of constraints are heavily skewed, a certain candidate can reach a very high probability of winning, allowing us to model non-stochastic phonological patterns. For instance, suppose we model a language in which a bilabial click does not exist but instead surfaces as a bilabial stop. Let us set the weight of the markedness constraint prohibiting a click as 10 and the weight of the faithfulness constraint as 1 (for how to read MaxEnt tableaux, see Section 2.5).

/ʘ/	CLICK (w=10)	Faith (w=1)	H-Score	$e^{-(\text{H-score})}$	Predicted Prob.
[ʘ]	−1		−10	4.539993e^{-05}	0.0001
[b]		−1	−1	0.3678794	0.9999

The result is that the probability of observing a click in this language is less than .001, an almost categorical result; indeed, even if it actually occurred, it would be considered a speech error. We can obtain these (near-)categorical results because MaxEnt involves exponentiation. In this way, MaxEnt allows room to accommodate both categorical and stochastic phonological and morphophonological patterns.

[4] We assume that the vowel sequence [iu], with no fall in sonority, is syllabified separately as [i.u]. The onsetless nature of the second syllable is represented by "_" in the text. See Kubozono (2015) for an extensive discussion of Japanese diphthongs and hiatus.

[5] There are several studies of sonority effects on blend formation in other languages. Bat-El (1996) discusses the role of sonority in blend formation in Hebrew. Likewise, Labrune (2006) suggests that similar tendency may be observed in Japanese blending formation.

[6] While Parker (2002) has shown that sonority is one key fact that affects binomial ordering, it is not the only factor that affects word ordering in English binomials. See also Benor and Levy (2006), Mollin (2012), and Lohmann (2014) for recent corpus-based surveys of English binomial orderings.

[7] The disyllabic names used in the stimuli all have initial accent, and hence the stimuli are controlled in this respect. Whether Japanese accent, like English stress (Lohmann 2014), affects binomial ordering in Japanese is an interesting question for future research.

[8] Some consider /h/ a voiceless approximant; i.e., a voiceless sonorant (Chomsky & Halle 1968). We follow other work (Jaeger & Ohala 1984; Lass 1976: 64–68; Parker 2002; Sagey 1986) that has demonstrated that /h/ is a voiceless fricative.

[9] Of course, it is possible to tease apart these two hypotheses empirically by using glide initial words in place of /r/-initial words. In our experiment, however, we found it hard to find a sufficient number of glide-initial disyllabic girls' names in Japanese.

[10] Since two consecutive nasal consonants within a word (e.g., ***m**o**n**a*) are shared by the compared candidates, they can be ignored in our tableaux thanks to the Cancellation Lemma (Prince & Smolensky 1993/2004).

[11] The harmonic scores of candidates can be used to model acceptability judgments as well (e.g.,

Coetzee & Pater 2008); provided that the optimal candidate in each candidate set has the same violation profile, then the lower a candidate's harmonic score is *across* candidate sets, the more unlikely it is to be considered acceptable. To take the analysis in (1) as an example, we can predict that *ha**n**a#**m**ona* (= −0.082) is the most harmonic, *ha**na**#**na**mi* (= −0.924) the least, and *ku**m**i#**m**ona* (= −0.345) in between; as a result, *ha**na**#**na**mi* is judged to be less acceptable than *ku**m**i#**m**ona* and *ku**m**i#**m**ona* less acceptable than *ha**n**a#**m**ona*.

[12] An anonymous reviewer asked whether we could have used partially ordered constraints (Antilla 1997, 2002) or Noisy Harmonic Grammar for the case at hand (Coetzee & Kawahara 2013; Coetzee & Pater 2008). One clear advantage of the MaxEnt model is its ability to predict the probabilities of each candidate, allowing us to compare the predictions with the observed data. For more recent discussions of various stochastic phonological models, see Zuraw and Hayes (2017) as well as Hayes (2017).

[13] As an anonymous reviewer rightfully point out, this paper does not show that all phonological/morphophonological patterns are stochastic; indeed, Japanese does not stochastically use front rounded vowels or clicks. Recall that when weights are sufficiently skewed, MaxEnt can provide deterministic outcomes as well. See Footnote 3.

[14] Pullum (2013b) writes, "Ad Neeleman is a good representative of the large majority of modern theoretical linguists who work with generative-enumerative syntax (GES) theories" (p. 532). Therefore, we are not the only ones who feel that this position is widely assumed in generative linguistics.

[15] One limitation of this study, as an anonymous reviewer pointed out, is that we did not test the effects of OCP on obstruent segments, as we used obstruents that did not form a natural class, [k, t, h]. In part, this was inevitable were we to obtain a sufficient number of stimuli. However, we do not wish to imply that identity avoidance effects among obstruents are uninteresting, but leave this topic for future research.

[16] We thank an anonymous reviewer for pointing out this issue.

References

Adli, Aria (2010) Constraint cumulativity and gradience: Wh-scrambling in Persian. *Lingua* 120(9): 2259–2294.

Alderete, John and Stefan Frisch (2007) Dissimilation in grammar and the lexicon. In: Paul de Lacy (ed.) *The Cambridge Handbook of Phonology*, 379–398. Cambridge: Cambridge University Press.

Anttila, Arto (1997) Deriving variation from grammar. In: Frans Hinskens, Roeland van Hout and W.

Leo Wetzels (eds.) *Variation, change, and phonological theory*, 35–68. Amsterdam: John Benjamins.

Anttila, Arto (2002) Morphologically conditioned phonological alternations. *Natural Language & Linguistic Theory* 20: 1–42.

Baayen, R. H. (2008) *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Bat-El, Outi (1996) Selecting the best of the worst: the grammar of Hebrew blends. *Phonology* 13: 283–328.

Benor, Sarah Bunin and Roger Levy (2006) The chicken or the egg? A probalistic analysis of English binomials. *Language* 82: 233–277.

Berent, Iris, Donca Steriade, Tracy Lennertz and Vered Vaknin (2007) What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104: 591–630.

Berent, Iris and Joseph Shimron (1997) Co-occurrence restrictions on identical consonants in the Hebrew lexicon: Are they due to similarity? *Journal of Linguistics* 39: 31–55.

Berkley, Deborah (1994) The OCP and gradient data. *Studies in the Linguistic Sciences* 24: 59 –72.

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

Boersma, Paul (2001) Praat, a system for doing phonetics by computer. *Glot International* 5(9/10): 341–345.

Boersma, Paul and Bruce Hayes (2001) Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32: 45–86.

Bresnan, Joan and Jennifer Hay (2008) Gradient grammar: An effect of animacy on the syntax of give in New Zealand and American English. *Lingua* 118(2): 245–259.

Chomsky, Noam (1957) *Syntactic structures*. The Hague: Mouton.

Chomsky, Noam (1965) *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam and Morris Halle (1968) *The sound pattern of speech*. Cambridge, MA: MIT Press.

Clements, George N. (1990) The role of the sonority cycle in core syllabification. In: John Kingston and Mary Beckman (eds.) *Papers in laboratory phonology 1: Between the grammar and physics of speech*, 283–333. New York: Cambridge University Press.

Coetzee, Andries W. and Shigeto Kawahara (2013) Frequency biases in phonological variation. *Natural Language and Linguistic Theory* 31(1): 47–89.

Coetzee, Andries W. and Joe Pater (2008) Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language & Linguistic Theory* 26: 289–337.

Cohn, Abigail (2006) Is there gradient phonology? In: Gisbert Fanselow, Caroline Féry, Ralph Vogel, and Matthias Schlesewsky (2006), 25–44.

Colavin, Rebecca, Roger Levy and Sharon Rose (2014) Modeling OCP-place in Amharic with the maximum entropy phonotactic learner. *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 46: 27–41.

Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis and Ingrid Norrmann (2011) Explaining sonority projection effects. *Phonology* 28: 197–234.

Fanselow, Gisbert, Caroline Féry, Ralph Vogel, and Matthias Schlesewsky (2006) (eds.) *Gradience in grammar: Generative perspectives*. Oxford: Oxford University Press.

Frisch, Stefan, Janet Pierrehumbert and Michael Broe (2004) Similarity avoidance and the OCP. *Natural Language & Linguistic Theory* 22: 179–228.

Goldsmith, John (1976) Autosegmental phonology. Unpublished doctoral dissertation, MIT.

Goldwater, Sharon and Mark Johnson (2003) Learning OT constraint rankings using a maximum entropy model. In: Jennifer Spenader, Anders Erikson and Osten Dahl (eds.) *Proceedings of the Stockholm workshop on variation within Optimality Theory*, 111–120. Stockholm: Stockholm University.

Greenberg, Joseph H. and James J. Jenkins (1964) Studies in the psychological correlates of the sound system of American English. *Word* 20: 157–177.

Halle, Morris (1978) Knowledge unlearned and untaught: What speakers know about the sounds of their language. In: Morris Halle, Joan Bresnan and George Miller (eds.) *Linguistic theory and psychological reality*, 294–303. Cambridge, MA: MIT Press.

Hayes, Bruce (2017) Varieties of Noisy Harmonic Grammar. *Proceedings of the 2016 Annual Meeting of Phonology*. Online Publication. The Linguistic Society of America. Washington, DC.

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

Hayes, Bruce, and Zsuzsa Cziráky Londe (2006) Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology* 23: 59–104.

Hayes, Bruce, Kie Zuraw, Peter Siptar and Zsuzsa Londe (2009) Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85: 822–863.

Hayes, Bruce, Colin Wilson and Anne Shisko (2012) Maxent Grammars for the metrics of Shakespeare and Milton. *Language* 88: 691–731.

Hayes, Bruce, Colin Wilson and Benjamin George (2009) Manual for Maxent Grammar Tool. Downloadable at http://www.linguistics.ucla.edu/people/hayes/Maxent-GrammarTool/

Irwin, Mark (2016) A rendaku bibliography. In: Timothy Vance and Mark Irwin (eds.) *Sequential voicing in Japanese compounds: Papers from the NINJAL Rendaku Project*, 235–250.

Ito, Junko (1989) A prosodic theory of epenthesis. *Natural Language & Linguistic Theory* 7: 217–260.

Ito, Junko and Armin Mester (2003) *Japanese morphophonemics: Markedness and word structure*. Cambridge, MA: MIT Press.

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

Jaeger, Jeri and John J. Ohala (1984) On the structure of phonetic categories. *Proceedings of the 10th Annual Meeting of the Berkeley Linguistics Society*: 15–26.

Katamba, Francis (1993) *Morphology*. London: Macmillan Press, LTD.

Kawahara, Shigeto (2013) Testing Japanese loanword devoicing: Addressing task effects. *Linguistics* 51(6): 1271–1299.

Kawahara, Shigeto (2015) Japanese /r/ is not feature-less: A rejoinder to Labrune (2014). *Open Linguistics* 1: 432–443.

Kawahara, Shigeto, Hajime Ono and Kiyoshi Sudo (2006) Consonant co-occurrence restrictions in Yamato Japanese. *Japanese/Korean Linguistics* 14: 27–38. Stanford: CSLI Publications.

Kawahara, Shigeto and Shin-ichiro Sano (2016) Rendaku and identity avoidance: Consonantal identity and moraic identity. In: Timothy J. Vance and Mark Irwin (2016), 47–55.

Keller, Frank (2006) Linear Optimality Theory as a model of gradience in grammar. In: Gisbert Fanselow, Caroline Féry, Ralph Vogel and Matthias Schlesewsky (2006), 270–287.

Kenstowicz, Michael (1994) *Phonology in generative grammar*. Oxford: Blackwell.

Kilbourn-Ceron, Oriana and Morgan Sonderegger (2017) Boundary phenomena and variability in Japanese high vowel devoicing. *Natural Language & Linguistic Theory*.
DOI: https://doi.org/10.1007/s11049-017-9368-x

Kumagai, Gakuji (2017) Cumulative faithfulness effect in Māori loanword adaptation: The case of repair for consonant clusters. *Phonological Studies* 20: 77–84.

Kubozono, Haruo (1989) The mora and syllable structure in Japanese: evidence from speech errors. *Language and Speech* 32: 249–278.

Kubozono, Haruo (2015) (ed.) *The handbook of Japanese language and linguistics: Phonetics and phonology*. Berlin: Mouton de Gruyter.

Kubozono, Haruo (2015) Diphthongs and vowel coalescence. In: Haruo Kubozono (2015), 215–252.

Labrune, Lawrence (2006) Phonemic preferences in Japanese non-headed binary compounds: what waa-puro, mecha-kucha and are-kore have in common. *Gengo Kenkyu* [Journal of the Linguistic Society of Japan] 129: 3–41.

Labrune, Laurence (2012) Questioning the universality of the syllable: Evidence from Japanese. *Phonology* 29: 113–152.

Labrune, Lawrence (2014) The phonology of Japanese /r/: A panchronic account. *Journal of East Asian Linguistics* 23: 1–25.

Lasnik, Howard (2004) Pronouns and non-coreference. *University of Maryland Working Papers in Linguistics* 13: 214–227.

Lasnik, Howard and Mamoru Saito (1984) On the nature of proper government. *Linguistic Inquiry* 15: 235–289.

Lass, Roger (1976) *English phonology and phonological theory*. Cambridge: Cambridge University Press.

Leben, William R. (1973) Suprasegmental phonology. Unpublished doctoral dissertation, MIT.

Legendre, Géraldine, Yoshiro Miyata and Paul Smolensky (1990) Harmonic Grammar—a formal multilevel connectionist theory of linguistic wellformedness: An application. *Proceedings of the 20th annual conference of the Cognitive Science Society*: 884–891. Cambridge: Lawrence Erlbaum.

Legendre, Géraldine, Antonella Sorace and Paul Smolensky (2006) The optimality theory-harmonic grammar connection. In: Paul Smolensky and Géraldine Legendre (eds.) *The harmonic mind: From neural computation to optimality theoretic grammar, vol. 2: Linguistic and philosophical implications*, 339–402. Cambridge, MA: MIT Press.

Lohmann, Arne (2014) *English coordinate constructions: A processing perspective on constituent order*. Cambridge: Cambridge University Press.

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

McCarthy, John J. (1986) OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17: 207–263.

McCawley, James D. (1968) *The phonological component of a grammar of Japanese*. The Hague: Mouton.

McPherson, Laura and Bruce Hayes (2016) Relating application frequency to morphological structure: The case of Tommo So vowel harmony. *Phonology* 33: 125–167.

Mollin, Sandra (2012) Revisiting binomial order in English: ordering constraints and reversibility. *English Language and Linguistics* 16: 81–103.

Moore-Cantwell, Claire and Joe Pater (2016) Gradient exceptionality in Maximum Entropy Grammar with lexically specific constraints. *Catalan Journal of Linguistics* 15: 53–66.

Neeleman, Ad (2013) Comments on Pullum. *Mind & Language* 28(4): 522–531.

Otake, Takashi, Giyoo Hatano, Anne Cutler and Jacques Mehler (1993) Mora or syllable? Speech Segmentation in Japanese. *Journal of Memory and Language* 32: 258–278.

Padgett, Jaye (1992) OCP subsidiary features. *Proceedings of the North East Linguistic Society* 22: 335–346. University of Massachusetts, Amherst. GLSA.

Parker, Steve (2002) Quantifying the sonority hierarchy. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Parker, Steve (2011) Sonority. In: Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice (eds.) *The Blackwell companion to phonology*, 1160–1184. Oxford: Wiley-Blackwell.

Pater, Joe (2009) Weighted constraints in generative linguistics. *Cognitive Science* 33: 999–1035.

Pater, Joe (2016) Universal grammar with weighted constraints. In: John McCarthy and Joe Pater (eds.) *Harmonic grammar and harmonic serialism*, 1–46. London: Equinox Press.

Pierrehumbert, Janet B. (1997) Stochastic phonology. *Glot International* 5(6): 195–207.

Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt and Michael Becker (2010) Harmonic grammar with linear programming: From linear systems to linguistic typology. *Phonology* 27: 77–117.

Prince, Alan (1983) Relating to the grid. *Linguistic Inquiry* 14: 19–100.

Prince, Alan and Paul Smolensky (1993/2004) *Optimality theory: Constraint interaction in generative* *grammar*. Malden, MA & Oxford, UK: Blackwell.

Pullum, Geoffrey K. (2013a) The central question in comparative syntactic metatheory. *Mind & Language* 28(4): 492–521.

Pullum, Geoffrey K. (2013b) Consigning phenomena to performance: A response to Neeleman. *Mind & Language* 28(4): 532–537.

Sagey, Elizabeth (1986) The representation of feature and relations in nonlinear phonology. Unpublished doctoral dissertation, MIT.

Schütze, Carson T. (1996) *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago, IL. University of Chicago Press.

Schütze, Carson T. (2016) *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Berlin: Language Science Press.

Shademan, Shabhame (2007) Grammar and analogy in phonotactic well-formedness judgments. Unpublished doctoral dissertation, University of California, Los Angeles.

Shih, Stephanie S. (2014) *Towards optimal rhyme*. Unpublished doctoral dissertation, Stanford University.

Shih, Stephanie S. (2016) Super additive similarity in Dioula tone harmony. In: Kyeong-min Kim, Pocholo Umbal, Trevor Block, Queenie Chan, Tanie Cheng, Kelli Finney, Mara Katz, Sophie Nickel-Thompson, and Lisa Shorten (eds.) *Proceedings of the 33rd West Coast Conference on Formal Linguistics*, 361–370. Somerville, MA: Cascadilla Proceedings Project.

Shih, Stephanie S. and Sharon Inkelas (2016) Morphologically-conditioned tonotactics in multilevel Maxent Entropy grammar. *Proceedings of the 2015 Annual Meeting of Phonology*. Online Publication. The Linguistic Society of America. Washington, DC.

Shih, Stephanie S. and Kie Zuraw (to appear) Phonological conditions on variable adjective-noun word order in Tagalog. *Phonological Analysis*.

Smith, Jennifer (2002) Phonological augmentation in prominent position. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Sorace, Antonella and Frank Keller (2005) Gradience in linguistic data. *Lingua* 115(11): 1497–1524.

Sprouse, Jon (2007a) A program a program for experimental syntax: Finding the relationship between acceptability and grammatical knowledge. Unpublished doctoral dissertation, University of Maryland, College Park.

Sprouse, Jon (2007b) Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics* 1: 123–134.

Sprouse, Jon (2015) Three open questions in experimental syntax. *Linguistics Vanguard*. Online Publication.

Tanaka, Yu (2017) *The Sound patterns of Japanese surnames*. Unpublished doctoral dissertation, University of California, Los Angeles.

Vance, Timothy J. (1980) The psychological status of a constraint on Japanese consonant alternation.

Linguistics 18: 245–267.

Vance, Timothy J. (1987) *An introduction to Japanese phonology*. New York: SUNY Press.

Vance, Timothy J. (2015) Rendaku. In: Haruo Kubozono (2015), 397–441.

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

White, James (2017) Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a P-map bias. *Language* 93: 1–36.

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

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

Zhang, Jie, Yuwen Lai, and Craig Sailor (2011) Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. *Lingua* 121: 186–206.

Zuraw, Kie (2000) Patterned exceptions in phonology. Unpublished doctoral dissertation. University of California, Los Angeles.

Zuraw, Kie (2010) A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language & Linguistic Theory* 28(2): 417–472.

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

Zuraw, Kie and Yu-An Lu (2009) Diverse repairs for multiple labial consonants. *Natural Language & Linguistic Theory* 17: 197–224.