Consonant Cooccurrence Restrictions in Yamato Japanese

SHIGETO KAWAHARA*
University of Massachusetts, Amherst

HAJIME ONO
University of Maryland, College Park

KIYOSHI SUDO
New York University

1. Introduction

This paper points out previously unnoticed dissimilatory restrictions in Japanese. Often, similar consonants, especially those that have the same place of articulation, are prohibited from cooccurring within a particular domain such as a stem or a word. Such restrictions have been found in many languages, sometimes under the rubric of the **Obligatory Contour Principle** (**OCP**). Languages that have been shown to have such effects include Arabic (McCarthy 1986), English (Berkley 1994), Javanese (Mester 1986), Muna (Coetzee and Pater 2005), Rotuman (McCarthy 2003), Russian (Padgett 1992), and others. The primary aim of this paper is to argue that similar consonant cooccurrence restrictions are found in the native vocabulary of Japanese (Yamato Japanese). We show that homorganic consonants

* We wish to thank Bjarke Frellesvig, Shōsuke Haraguchi, Markus Hiller, Kazu Kurisu, John McCarthy, Joe Pater, Adam Ussishkin, Tim Vance, Cheryl Zoll, the audience at University of Maryland, the members of the UMass Summer Phonology/Phonetics Group, the audience at the Tokyo Circle of Phonologists, and the audience at the 14th Japanese/Korean Conference for very useful comments on this project. Any shortcomings of interpretation and exposition of the data rest solely with us.

are less likely to cooccur than expected within a root, although this tendency is not an absolute prohibition.

The remainder of this paper is structured as follows. §2 presents the method of our survey. §3 discusses several aspects of the consonant cooccurrence restrictions we have found. In §4 we discuss crosslinguistic implications of this study. Throughout, we focus on descriptive aspects of our findings, although theoretical implications are touched upon where appropriate. One final remark before closing this introductory section. Due to space limitations, it is impossible to provide detailed data for all aspects of findings presented in this paper, but more complete data are found in the following website, which can be consulted for further information: http://www.people.umass.edu/kawahara/yamato.htm.

2. Method

The first step of our investigation was to extract all monomorphemic Yamato roots from a large dictionary of Japanese, *Kōjien* (Shinmura 1998). Loanwords, mimetics, affixes, and interjections were excluded.[1] Obsolete words, noted as such in the dictionary, were also excluded for two reasons: this study focused on the synchronic grammar, and the morphological composition of such words was often not clear. In the case of a morphologically related paradigm (e.g. *sadamari* 'law', *sademeru* 'to define', *sadamaru* 'to fix') the root was counted only once. The database contained 4,011 roots at the end of this procedure.

From the collected set of monomorphemic roots, adjacent pairs of consonantal sequence were extracted. For instance, a root that contains {s d m r} yielded three pairs: {s d}, {d m}, and {m r}. This resulted in 4,737 adjacent consonantal pairs. Two notes are in order regarding this procedure. First, coda consonants, which place-assimilate to the following consonant, were systematically ignored. For example, given a word like *tombo* 'dragonfly', only {t b} was counted, but not {m b}; we focused on onset-adjacency. Second, our study was based on surface forms. For example, [ʃi], which is arguably derived from /si/, was counted as [ʃ] (see §4.3 for discussion on this choice).

Then, for all adjacent consonant pairs, the frequency of cooccurrence was counted and summarized, as illustrated in Table 1. Table 1 is a simplified table listing only three consonants, for the sake of illustration. It lists the number of pairs with a particular consonant configuration. For example, there are three pairs in which C_1 is [m] and C_2 is [b]; call these **observed** numbers.

[1] Dissimilatory restrictions on mimetics have been pointed out by Hamano (1998: 41–2).

C_1 \ C_2	m	b	t	Total
m	9	3	6	18
b	2	2	9	13
t	8	7	8	23
Total	19	12	23	54

Table 1. Observed Numbers

C_1 \ C_2	m	b	t	Total
m	6.3	4.0	7.7	18
b	4.6	2.9	5.5	13
t	8.1	5.1	9.8	23
Total	19	12	23	54

Table 2. Expected Numbers

From Table 1 we can calculate the **expected** number of consonant pairs. For example, the probability of [m] occurring as C_1 in Table 1 is $18/54 = .33$. Similarly, the probability of [b] occurring in C_2 is $12/54 = .22$. Therefore, if these two events are independent (i.e. their probabilities do not influence one another), then the probability of both events occurring is $.33 \times .22 = .074$. Since there are 54 pairs in total, we should expect $.074 \times 54 = 4$ pairs that have [m] followed by [b]. This is the **expected** number for the {m b} pair, if the two consonants are combined at random. In general, an expected number for pairs of {x y} can be calculated as $E(C_1 = x, C_2 = y) = P(C_1 = x) \times P(C_2 = y) \times N$ (where N is a total number of pairs).

Expected numbers were calculated for each consonantal pair, as shown in Table 2. Next, so-called O/E values were calculated by dividing each observed number by the corresponding expected number. O/E values smaller than 1 indicate that consonantal pairs are underrepresented (i.e. occur less often than expected); O/E values larger than 1 indicate overrepresentation.

A χ^2 test was used to check the statistical significance of the deviations of O/E values from 1. The χ^2 value for each {x y} pair is the sum of $(O - E)^2/E$ over the four cells, {x, y}, {x, non-y}, {non-x, y}, and {non-x, non-y}. {non-x} and {non-y} are defined as all the sounds (or classes of sounds) other than {x} and {y}, respectively. For example, for χ^2 of two adjacent labials, the relevant cells are {lab, lab}, {lab, non-lab}, {non-lab, lab}, and {non-lab, non-lab}. The degree of freedom is thus $(2 - 1) \times (2 - 1) = 1$.

To investigate the existence of general cooccurrence restrictions based on place homorganicity, consonants were classified according to their major place of articulations (labial, coronal, palatal, dorsal). Further, the coronal class was divided into two classes (coronal sonorants and coronal obstruents), because previous studies (see works cited above) found that such a division is observed. We refer to these classes as “identity classes”, following Yip’s (1989) terminology.

Given the findings of the previous studies, the a priori prediction was that any pair of two consonants from the same identity class should be underrepresented. We set our α-level at .05; if $\chi^2 > 3.84$, then $p < .05$. Since the the multiple applications of the same test were preplanned, no familywise error α-level adjustment was used.

3. Results

3.1. General Patterns

The procedure described above revealed that there are indeed consonant cooccurrence restrictions in Yamato Japanese, as summarized in Table 3, where pairs that are underrepresented to a statistically significant degree are indicated by shading. Note that the linear order is abstracted away; see the website for the table with a preserved linear order.

	Labial p, b, m, ɸ, w	Cor-Obs t, d, ts, s, z, ʃ, ʒ, tʃ, dʒ	Cor-Son n, r	Palatal ç, j	Dorsal k, g
Labial	O = 43 O/E = 0.22 $\chi^2 = 180.11$	O = 404 O/E = 1.35 $\chi^2 = 67.78$	O = 222 O/E = 1.20 $\chi^2 = 11.43$	O = 41.5 O/E = 1.08 $\chi^2 = 0.31$	O = 225 O/E = 1.07 $\chi^2 = 1.72$
Cor-Obs		O = 247 O/E = 0.53 $\chi^2 = 218.75$	O = 295.5 O/E = 1.02 $\chi^2 = 0.27$	O = 62 O/E = 1.03 $\chi^2 = 0.09$	O = 417 O/E = 1.27 $\chi^2 = 45.36$
Cor-Son			O = 69 O/E = 0.39 $\chi^2 = 104.25$	O = 37.5 O/E = 1.01 $\chi^2 = 0.00$	O = 266.5 O/E = 1.31 $\chi^2 = 31.43$
Palatal				O = 3 O/E = 0.39 $\chi^2 = 3.15$	O = 42 O/E = 0.99 $\chi^2 = 0.00$
Dorsal					O = 66 O/E = 0.29 $\chi^2 = 193.31$

Table 3. O/E Values by Identity Classes – Adjacent Pairs (Significant underrepresentation is indicated by shading. If $\chi^2 > 3.84$, $p < .05$. N = 4,737.)

Table 3 clearly shows OCP effects. All pairs of consonants from the same identity class show underrepresentation to a statistically significant degree, with the exception of palatals.[2] For instance, a labial occurs with another labial only about 20 percent as often as expected. Even palatal pairs are underrepresented, but only barely significantly ($p = .075$), due to the small number of data points. Note also that none of the nonhomorganic pairs are

[2] "Palatal" allophones of coronal obstruents (ʃ, ʒ, tʃ, dʒ) pattern with coronal consonants: they are underrepresented with coronal obstruents (O/E = .57, $\chi^2 = 21.89$, $p < .001$). They are prepalatals, and thus are treated as coronals with [−ant] specification in Japanese (see Zoll 1997 for phonological evidence that they are coronals).

underrepresented. These results suggest that, in Yamato Japanese, a pair of adjacent consonants from the same identity class occurs much less likely than expected, quite similar to the effect observed in other languages cited in the introduction. The χ^2 tests show that the probability of these effects occurring by chance is quite small; even for the smallest χ^2 (104.3 for coronal sonorants) the associated probability is smaller than .001.

In Table 3 we only computed values for pairs of adjacent consonants, since it is known that the OCP applies most stringently to adjacent consonants. We also checked whether nonadjacent consonants exhibit any restrictions. The results are given in Table 4, where only weaker (if any) OCP effects hold. Only the shaded cells (coronal sonorants and obstruents) are statistically significant, and the effects are even weaker than those found in Table 3.

	Labial p, b, m, ɸ, w	Cor-Obs t, d, ts, s, z, ʃ, ʒ, tʃ, dʒ	Cor-Son n, r	Palatal ç, j	Dorsal k, g
Labial	O = 45 O/E = 0.87 $\chi^2 = 1.39$	O = 88.5 O/E = 1.01 $\chi^2 = 0.03$	O = 65 O/E = 0.98 $\chi^2 = 0.06$	O = 11 O/E = 0.95 $\chi^2 = 0.03$	O = 74.5 O/E = 1.15 $\chi^2 = 2.15$
Cor-Obs		O = 121 O/E = 0.83 $\chi^2 = 8.76$	O = 140 O/E = 1.25 $\chi^2 = 13.13$	O = 17.5 O/E = 0.91 $\chi^2 = 0.26$	O = 110 O/E = 1.01 $\chi^2 = 0.02$
Cor-Son			O = 49 O/E = 0.57 $\chi^2 = 25.78$	O = 19 O/E = 1.29 $\chi^2 = 1.63$	O = 81 O/E = 0.10 $\chi^2 = 0.10$
Palatal				O = 1 O/E = 0.39 $\chi^2 = 1.03$	O = 16 O/E = 1.11 $\chi^2 = 0.23$
Dorsal					O = 70 O/E = 0.86 $\chi^2 = 2.59$

Table 4. O/E Values by Identity Classes – Nonadjacent Pairs (N = 1679)

One point that merits discussion is the fact that coronal sonorants and coronal obstruents constitute a separate identity class, that is, coronal sonorants and coronal obstruents happily cooccur even in adjacent positions (O/E = 1.02, χ^2 = .27, p = .60). No evidence for such a distinction is found for labials or dorsals. Such a tendency (i.e. a sonority split for coronals) is observed in many languages, including all the languages listed in the introduction. Further evidence is found in an identity-driven cluster reduction in Wintu (McGarrity 1999). The fact that this generalization holds also in Yamato Japanese suggests the robustness of this crosslinguistic tendency, and this generalization might arguably be universal. Here, we have nothing new to say about why there is such a division only for coronals and not for other classes (see Frisch et al. 2004 for a hypothesis; see Coetzee and Pater 2005 and McCarthy 2003 for a counterexample to their claim).

Next, we checked the O/E values of pairs of identical consonants. It is known that totally identical consonants avoid violating the OCP (see the works cited in the introduction). In Yamato Japanese, two voiced obstruents are disallowed within a stem (Lyman's Law; see e.g. Itô and Mester 1986), and even when they are identical, two voiced obstruents cannot cooccur. Besides these cases, however, total identity does seem to provide an escape hatch from the OCP. Each of [p], [ɸ], [w], [t], [ts], [s], [n], [ʃ], [tʃ], [ç], [j], [h] seems to freely occur with itself ([n] and [ʃ] show slight underrepresentation). Cases that are underrepresented to a statistically significant degree are [m], [r], and [k].[3]

Finally, as mentioned above, Yamato Japanese prohibits two voiced obstruents within the same stem, and this restriction holds regardless of the place of articulation. We tested whether two other manner features, [±cont] and [±nasal], are subject to similar restrictions. The results demonstrate that continuancy results in very slight underrepresentation ([+cont] O/E = .80; [−cont] O/E = .90), but [m] and [n] are not underrepresented with one another). Voicing seems to have a special status in Japanese.

3.2. Similarity Correlates with Underrepresentation?

Frisch et al. (2004) argue that the degree of similarity correlates with the degree of underrepresentation: the more similar two consonants are, the less likely they cooccur (see Frisch et al. 2004 on how to measure similarity). Coronal consonants provide a nice testing ground concerning whether this generalization applies in the case of Yamato Japanese or not. We will see that it is (at least partially) supported. Here we present only a subset of the segment-by-segment cooccurrence data (see the aforementioned website for more information).

Starting with [±voice], consider the following observations (the numbers in parentheses represent O/E values).

(1) [±voice]
  [t] is more underrepresented with [s] (.35) than with [z] (.58).
  [ts] is more underrepresented with [s] (.21) than with [z] (.52).
  [ʃ] is more underrepresented with [s] (.29) than with [z] (1.11).

All these suggest that two voiceless consonants are less likely to cooccur

---

[3] With 20 applications of the same χ^2 test, α is adjusted by the Bonferroni method, $\alpha = .05/20 = .0025$. If $\chi^2 > 9.14$, $p < .0025$. See the website for O/E and χ^2 values.

than pairs with different voicing specifications. In other words, within the coronal class, having two voiceless consonants is disfavored. This suggests that agreeing in [±voice] contributes to more similarity, which leads to more underrepresentation. It is important to notice, however, that this assumes place homorganicity; [t] and [k], for example, happily cooccur (O/E = .99 $\chi^2 = .00$).

The only other relevant feature is [±cont], as a difference in [±nas] entails a difference in [±son]. Here, an unexpected pattern is observed:

(2) [±cont]
 [d] is more underrepresented with [s] (.66) than with [t] (1.91).
 [z] is more underrepresented with [t] (.58) than with [s] (1.04).

The generalization appears to be that it is better to agree in continuancy. We do not, however, simply submit to the view that this constitutes a counterexample to the generalization that similarity correlates with the degree of underrepresentation. The reason is that pairs like [d]-[t] and [z]-[s] are not only less underrepresented, but not underrepresented at all.

We suggest that pairs like [d]-[t] and [z]-[s], which differ only in terms of [voice], can be treated as identical for OCP computations. Recall that total identity usually provides an escape hatch from OCP violations. Perhaps [d] and [t], for example, are not different enough to be treated as different, and hence there is no underrepresentation.

This hypothesis points to special properties of voicing in Japanese. On the one hand, although it can affect cooccurrence rates, it can also be ignored in the computation of identity. On the other hand, it is subject to a cooccurrence restriction independent of place homorganicity (see above).[4] Here we suggest two tentative explanations for these peculiar properties of voicing in Japanese. First, it has a special status in Japanese phonology in that it is a diacritic (autosegmental) feature, as Unger (1977) argues. Thus, just as tones can be ignored in the computation of identity, voicing can potentially be ignored, total identity being defined in terms of features other than [±voice]. Further, since it constitutes its own autosegmental tier, it is specifically subject to OCP(voi).

Another plausible explanation is to assume that voicing is perceptually nonsalient. If it is nonsalient, it is easily lost (Steriade 2001), and it can also be ignored in the identity computation. These special properties of voicing in Japanese merit further discussion in future research.

[4] These properties are also found in Muna's prenasalization, as identified by Coetzee and Pater (2005). It might be noteworthy that, historically, Japanese voicing used to be prenasalization (Unger 1977).

3.3. Further Support: Evidence from Rendaku

We have identified a general restriction in Japanese that two adjacent consonants with the same place of articulation are disfavored. We show in this section that such a restriction manifests itself in a morphophonemic alternation as well, namely, Rendaku. Rendaku voices the first consonant of a second member in compound. By this process, [h] alternates with [b], as in *nui* 'sew' + ***h****ari* 'needle' → *nui-****b****ari* 'sewing needle'.

However, stems that begin with [h] followed by [m] do not usually undergo Rendaku, as shown by the examples in (3).[5]

(3) Creation of [b…m] blocked.

suna 'sand' + **h**a**m**a 'shore'	→	suna-**h**a**m**a	*suna-**b**a**m**a
oo 'big' + **h**a**m**o 'fish name'	→	oo-**h**a**m**o	*oo-**b**a**m**o
tema 'trouble' + **h**i**m**a 'boredom'	→	tema-**h**i**m**a	*tema-**b**i**m**a
mai 'dance' + **h**i**m**e 'princess'	→	mai-**h**i**m**e	*mai-**b**i**m**e
kutsu 'shoe' + **h**i**m**o 'lace'	→	kutsu-**h**i**m**o	*kutsu-**b**i**m**o
oo 'big' + **h**e**m**a 'mistake'	→	oo-**h**e**m**a	*oo-**b**e**m**a

This blockage of Rendaku should be compared to a minimally different case in which [h] is followed by [n]; in such a case, Rendaku is not blocked:

(4) Creation of [b…n] not blocked.

ai 'purple' + **h**a**n**a 'flower'	→	ai-**b**a**n**a	*ai-hana
te 'hand' + **h**a**n**e 'wing'	→	te-**b**a**n**e	*te-hane
nagasi 'float' + **h**i**n**a 'doll'	→	nagasi-**b**i**n**a	*nagasi-hina
oo 'big' + **h**u**n**a 'gibel'	→	oo-**b**u**n**a	*oo-huna
oo 'big' + **h**u**n**e 'ship'	→	oo-**b**u**n**e	*oo-hune

The blockage of Rendaku in (3) can be explained in terms of the cooccurrence restriction effects we identified above. Given an underlying /h…m…/ sequence, changing /h/ into [b] would result in two adjacent labials within the same stem, which is avoided. This explanation receives further support from the fact that when /h/ and /m/ are nonadjacent, Rendaku is not blocked (e.g. *ryoori-****b****asa****m****i* 'cooking scissors', *naga-****b****aka****m****a* 'long hakama'). This is predicted because only weaker (if any) OCP restrictions hold on nonadjacent consonantal pairs.

3.4. Restrictions on Verbal Stems

Finally, we present yet another novel finding about the Japanese lexicon. It is well known that in Semitic languages, verbal roots in which the first and

[5] Verbs and deverbal nouns undergo Rendaku even in this configuration; e.g. *musi-****b****a****m****u* 'to corrupt' and *asi-****b****u****m****i* 'stepping'. We do not have a good explanation for this fact besides noting that the applying force of Rendaku might be stronger for verbs than for nouns.

second consonants are identical are banned, while roots where the second and third consonants are identical are permitted (McCarthy 1979). For example, there is a root like *smm* but not a root like *ssm*.

We tested whether a similar restriction holds on Japanese verbal stems. The results revealed a mirror-image restriction: roots with identical consonants in the first and second syllables are common, while verbs with identical consonants in the second and third syllables are highly underrepresented. In other words, roots like those in (5) are rare; out of 1,266 roots, we found only 8 instances (O/E = .1, $\chi^2 = 63.38$, $p < .001$). On the other hand, roots like those in (6) are rather common (60 instances; O/E = .9, $\chi^2 = .68$, $p = .41$).

(5) exceptional (C_i)$C_j$$C_j$ roots (exhaustive)

atatamaru 'to warm up', ononoku 'to be frightened', omomuku 'to visit', kuwawaru 'to add', sitatameru 'to write down', sitataru 'to drip', tumamu 'to pick', wananaku 'to be scared'

(6) $C_i$$C_i$$C_j$ roots (examples)

kakeru 'to pour', kakumau 'to hide', sasuru 'to rub', sosogu 'to pour', tataku 'to hit', tatoeru 'to analogize', nonoshiru 'to curse', mamireru 'to immerse', ninau 'to take on', . . .

This is the mirror image of the Semitic pattern, and a similar pattern is found in Javanese (Mester 1986; though see Yip 1989:355 for an argument that this actually involves prefixing reduplication, since the first two vowels also agree).

These examples show that total identity is not always an escape hatch from OCP restrictions. Another implication is that the Japanese case identified here provides a counterexample to Frisch's (2004) claim that OCP effects are due to processing difficulty in segment linearization, and that the problem of repetition for processing is ameliorated as the loci of repetition get closer to the end of a word. This evidence is stronger than the Javanese case; (6) cannot be treated as reduplication because there are a number of forms that have different vowels in the first and second syllable.

4. Discussion

4.1. Tendencies, Not Prohibitions

One implication of this study is that consonant cooccurrence restrictions are not **categorical prohibitions** but **tendencies**. The cooccurrence restrictions found in Yamato Japanese are much weaker than what is found in Arabic (see Frisch et al. 2004:168; their O/E values for pairs from the same identity class are usually less than 0.1). The O/E values in Japanese are between 0.2 and 0.6 (Table 3). Further, nonadjacent pairs barely show any underrepresentation (Table 4). Thus, it is not that homorganic pairs of consonants are absolutely prohibited; they just occur much less often than expected.

Our study thus shows that such cooccurrence restrictions are general tendencies rather than categorical prohibitions (see, e.g. Berkley 1994, Frisch et al. 2004; see also Moreton et al. 1998 for other aspects of phonological tendencies in the Japanese lexicon). How to capture such tendencies in the current framework of Optimality Theory (Prince and Smolensky 1993) remains an interesting theoretical question. See Coetzee and Pater 2005 for recent discussion.

4.2. The OCP and CV Segregation

It has been thought that OCP effects are tied to CV segregation: in languages that show the OCP, consonants and vowels are organized into different autosegmental tiers, the most famous example being Semitic languages, as proposed by McCarthy (1979). Thus, Yip (1989: 352) notes that "[when OCP effects are observed], at some level adjacency is always involved and that apparent nonadjacent instances *always* involve separation of consonant and vowel melodies underlyingly" (emphasis in the original). Frisch (2004) also claims a connection between the OCP and CV segregation.

However, it is unlikely that Japanese has CV segregation (relative order of consonants and vowels is highly predictable, but not completely: *nata* 'hatchet', *anta* 'you').[6] Thus, OCP restrictions may have nothing to do with CV segregation. See Berkley 1994 for similar arguments.

4.3. Underlying Form or Surface Form?

Our study was based on surface forms, but as Adam Ussishkin reminded us (p.c.), such OCP restrictions used to be captured as Morpheme Structure Conditions, imposed on the lexicon. For example, McCarthy (1979) prohibits two adjacent identical consonants underlyingly by positing that the OCP operates on underlying forms, and roots like *smm* are derived by long distance consonantal spreading of the second consonant.

Since underlying and surface forms are usually similar, it is hard to distinguish at which level consonant cooccurrence restrictions are operative. Yet we maintain that such restrictions are computed on surface forms, as predicted from a surface-oriented theory of phonology such as Optimality Theory. To recap, in building our database of consonantal sequences, surface forms were used. For instance, surface [ɸu], which arguably derives

[6] See McCarthy (1989) for discussion on the relation between predictability of CV order and segregation. One piece of indirect evidence for CV segregation in Japanese is the fact that hip-hop rhymes are computed on the basis of identity of vowels from the end of lines, ignoring consonants (Kawahara 2002). This evidence, however, is weak at best.

from /hu/, was treated as a labial, not as a pharyngeal [h]. This is first of all due to a current claim of Optimality Theory that there is no reason for [ɸu] to have to be underlyingly /hu/. Rather, from the perspective of language learners, it might be more natural to regard [ɸu] as stemming from /ɸu/ (Lexicon Optimization; Prince and Smolensky 1993).

Further, we found empirical evidence that what is at issue might be surface forms rather than underlying forms. Recall that words that have initial [h] followed by [m] generally do not undergo Rendaku to avoid creating [b…m] (§3.3). However, there is one exception: [ɸumi] 'letter' does become [bumi] when compounded, as in *koi-bumi* 'love letter'. This exception makes sense if we assume that [ɸumi] already violates OCP(lab) so that it freely undergoes Rendaku; not undergoing Rendaku does not ameliorate the OCP violation. This explanation assumes that the OCP is enforced on surface forms rather than underlying /humi/, since it distinguishes cases of surface [h…m] (Rendaku blocked) from cases of surface [ɸ…m] (Rendaku not blocked).

Also, [t] and [ts], allophones of /t/, differ slightly in their cooccurrence behavior. And [ç], a surface form of /h/ before [i], is underrepresented with palatal [j], although the effect was only marginally significant ($p = .075$). Finally, [ɸ], the allophone of /h/ before [u], is underrepresented with all other labial consonants [p, b, m, w], even though this pattern does not reach statistical significance because there are not enough words containing this sound. On the other hand, [h], which appears before nonhigh vowels, is underrepresented only with [m, w] (this underrepresentation makes sense because [h] was historically [p]). Therefore, at the very least, we can conclude that, even though /h/ in general can be underrepresented with other labial sounds, the restriction is stronger when /h/ becomes [ɸ] at the surface.

5. Concluding Remarks

This paper has pointed out several previously unnoticed restrictions on consonant cooccurrence in Yamato Japanese. We have shown that a pair of adjacent consonants from the same identity class is strongly disfavored, and Rendaku can be blocked by a prohibition against two adjacent labials. We also pointed out that verbs with identical consonants in the second and third syllables are highly underrepresented. These are rather surprising results in that Yamato Japanese is not historically related to any languages that have been shown to have similar OCP effects. This study thus contributes to the generality of consonant cooccurrence restrictions in natural language.

References

Berkley, D. 1994. The OCP and Gradient Data. *Studies in the Linguistic Sciences* 24:59–72.

Coetzee, A., and J. Pater. 2005. Lexically Gradient Phonotactics in Muna and Optimality Theory. Manuscript, University of Massachusetts, Amherst.

Frisch, S. 2004. Language Processing and Segmental OCP Effects. *Phonetically-Based Phonology*, eds. B. Hayes, R. Kirchner, and D. Steriade, 346–71. Cambridge: Cambridge University Press.

Frisch, S., J. Pierrehumbert, and M. Broe. 2004. Similarity Avoidance and the OCP. *Natural Language and Linguistic Theory* 22:179–228.

Hamano, S. 1998. *The Sound-Symbolic System of Japanese*. Stanford: CSLI.

Itô, J., and A. Mester. 1986. The Phonology of Voicing in Japanese. *Linguistic Inquiry* 17:49–73.

Kawahara, S. 2002. Aspects of Hip-Hop Rhymes in Japanese. Paper presented at the Language Study Workshop at International Christian University, Tokyo.

McCarthy, J. 1979. Formal Problems in Semitic Morphology and Phonology. Doctoral dissertation, MIT.

McCarthy, J. 1989. Linear Order in Phonological Representation. *Linguistic Inquiry* 20:71–99.

McCarthy, J. 2003. Rotuman Consonant Cooccurrence Restrictions. Manuscript, University of Massachusetts, Amherst.

McGarrity, L. 1999. A Sympathy Account of Multiple Opacity in Wintu. *Indiana University Working Papers in Linguistics* 1:93–107.

Mester, A. 1986. Studies in Tier Structure. Doctoral dissertation, University of Massachusetts, Amherst.

Moreton, E, S. Amano, and T. Kondo. 1998. Statistical Phonotactics in Japanese. *Transactions of Technical Committee on Psychological Acoustics*, H-98-120.

Padgett, J. 1992. OCP Subsidiary Features. *Proceedings of North East Linguistic Society* 22:335–46.

Prince, A., and P. Smolensky. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. Manuscript, University of Colorado, Boulder, and Rutgers University.

Shinmura, I., ed. 1998. *Kōjien*. Tokyo: Iwanami Shoten.

Steriade, D. 2001. The Phonology of Perceptibility Effect: The P-map and Its Consequences for Constraint Organization. Manuscript, UCLA.

Unger, J. 1977. *Studies in Early Japanese Morphophonemics*. Bloomington: Indiana University Linguistics Club.

Yip, M. 1989. Feature Geometry and Co-occurrence Restrictions. *Phonology* 6: 349–74.

Zoll, C. 1997. Conflicting Directionality. *Phonology* 14:263–86.