OCP Effects on Japanese Blending

Haruka Tada
The University of Tokyo

1 Introduction

This study takes up blending, a morphophonological phenomenon that involves "merging parts of words into one new word" (Marchand 1969: 29). Blends are classified into two types depending on the process involved: the first one is consciously-formed blends (i.e. blends in word-formation) as in (1), and the second one is spontaneously-formed blends (i.e. blend errors) as in (2) (Kubozono 1990). In all of the following examples, the word before the slash is referred to as the left-hand source word while the one after the slash is the right-hand source word. The resultant blend form is given after the arrow.

(1) Blends in word-formation

a. Japanese
- i. *gorira* 'gorilla' / *kujira* 'whale' → *gojira*
- ii. *poteto* 'potato' / *tomato* 'tomato' → *pomato*
- iii. *piano* 'piano' / *hamonika* 'harmonica' → *pianika*
- iv. *burausu* 'blouse' / *kattosoo* 'top' → *burausoo*
- v. *dojjibooru* 'dodge ball'/ *furisubii* 'frisbee' → *dojjibii*

b. English
- i. *cranberry / apple → cranapple*
- ii. *magazine / book → mook*
- iii. *breakfast* / *lunch → brunch*
- iv. *boxing* / *exercise → boxercise*

(2) Blend errors

a. Japanese
- i. *neko* 'cat'/ *nyanko* 'kitty' → *nenko*
- ii. *tomare* 'stop!' / *sutoppu* 'stop!'→ *tomappu*
- iii. *yuu* 'tie' / *musubu* 'tie' → *yusubu*

b. English
- i. *switched* / *changed → swinged*
- ii. *popular* / *public → poplic*
- iii. *close* / *near → clear*

This study mainly targets Japanese blends in word-formation as in (1a), which are used for naming fictional animals and plants, and new products and phenomena (Kubozono 1995). The fact that there are various linguistic conditions that influence blending is shown by Kubozono (1990, 1995). The detailed information will be provided in the next section.

However, even if all of the conditions are satisfied, there are cases where the resultant blend form cannot be entirely predicted. Taking (1a.iii) as an example, *pimonika* and *pianoka* are also possible resultant blend forms other than *pianika*. Therefore, this study aims to make clear the factors that decide a certain blend form. Especially, I shed light on the influence of the Obligatory Contour Principle (i.e. prohibition of adjacent identical specification; henceforth OCP) in this paper since their effects have been confirmed in other morphophonological phenomena. To the best of my knowledge, the investigation of the effects of the OCP on blending has never been done before. Therefore, the goal of this paper is to investigate the OCP effects on blending by using an experimental approach. The results of two experiments showed that the OCP was concerned with the selection of blend form. The unique characteristics of blending are considered to be reflected to the results.

The organization of this paper is as follows. Section 2 will introduce the constraints on source words and resultant blend forms based on Kubozono (1990) and indicate that a resultant blend form cannot always be narrowed down to one particular form just by following all the constraints. In Section 3, after I provide the basic information of the OCP, previous studies that investigated the effects in Japanese morphophonological phenomena will be reviewed. Section 4 will explain the experimental design and report the results of the two experiments. In Section 5, the results of the two experiments using blending will be compared with those of experiments conducted in previous studies using other morphophonological phenomena, clarifying their similarities and contrasts. Also, the reason for their difference will be discussed by attributing it to the different nature of blending compared to other morphophonological phenomena. Then, I will conclude this paper by addressing some limitations and topics for future research.

2 Background of blending

In this section, I will go into the background of blending. Firstly, the details of various linguistic conditions that influence blending will be provided. Based on Kubozono (1990), the following two types of constraints on blending will be introduced: one concerning the characteristics of source words (i.e. input) and the other

Proceedings of *AJL 4*

concerning how they are actually blended (i.e. output). Then, I will demonstrate that satisfying all the constraints does not always lead to one particular resultant blend form.

2.1 *Constraints on the inputs of blending* Source words are constrained by both syntactic and semantic aspects. Concerning the syntactic aspects, the syntactic category (i.e. part of speech) of the two source words is the same in most cases.[1] For example, in (1a.i), the source words *gorira* and *kujira* are nouns. Noun-noun blending is the most productive pattern.

Concerning the semantic aspects, the two source words are semantically related and generally, involve a similar, but not identical, semantic content. For instance, in (1a.i), *gorira* and *kujira* are both animals.[2]

2.2 *Constraints on the outputs of blending* There are morphological and phonological constraints on resultant blend forms. For the morphological aspects, blends are formed by combining the initial part (A) of the left-hand source word (AB) with the non-initial part (Y) of the right-hand source word (XY) as in (3).

(3) $\underline{A}B$ / $X\underline{Y}$ → AY

Taking (1a.i) as an example, the initial part *go* of the left-hand source word *gorira* and the non-initial part *jira* of the right-hand source word *kujira* form *gojira*.

Concerning the phonological aspects, the following two constraints are presupposed. One is to prohibit resultant blend forms which contain phonotactic structures that are unpermitted in the language. The other is to exclude the blend forms that are identical to one of the two source words. For example, in (1a.i), *gorira* is excluded from a blend form created by *gorira* and *kujira* since it is identical to the left-hand source word *gorira*.

In addition, there are two main phonological constraints. The first constraint concerns the possible switch point from the left-hand source word to the right-hand one. It prohibits switching which causes a split of syllable constituents such as an onset or coda split. This implies that the syllable position where the split occurs within a word is common between two source words. The most preferred switch point in Japanese is mora boundaries.

The second phonological constraint concerns the phonological length, which is measured by mora in Japanese.[3] The length of two source words corresponds to a high degree to that of resultant blend forms. To be specific, the length of the right-hand source word and that of the resultant blend form have a high degree of correspondence (i.e. $|XY| = |AY|$ in (3)). It leads to the fact that there is no length difference in the initial parts between the left-hand and right-hand source word (i.e. $|A| = |X|$ in (3)). For instance, in (1a.i), the right-hand source word *kujira* and the resultant blend form *gojira* have the same phonological length (i.e. three moras), and the initial part of the left-hand word *go* has the same length as that of the right-hand one *ku* (i.e. one mora).

2.3 *Target of this study: area which cannot be covered by constraints* By satisfying the constraints listed above, a resultant blend form is largely determined. However, it cannot be completely narrowed down. Taking (1a.iii) as an example, it becomes clear that the resultant blend form created by *piano* and *hamonika* will be one of *pimonika*, *pianika*, and *pianoka* because they follow the constraints presented in Section 2.2. However, there is a gap between listing the three candidates and deciding the one particular form among them. In short, it is not clear why the resultant blend form from *piano* and *hamonika* becomes *pianika* instead of *pimonika* or *pianoka*.

In the first place, the difference among them is the switch point from the left-hand source word to the right-hand one as summarized in (4). Henceforth, the switch point refers to the point where the left-hand source word switches to the right-hand one.

(4) *piano* / *hamonika*
 a. *pi-monika*: switch between the first and the second mora
 b. *pia-nika*: switch between the second and the third mora = (1a.iii)
 c. *piano-ka*: switch between the third and the fourth mora

---

[1] Very few potential exceptions were found in examples of blend error such as *befirst* which is the resultant blend form of <u>*be*</u>*fore* (conjunction) and <u>*first*</u> (adverb).

[2] In addition, right-hand source words, whose non-initial parts are incorporated in the resultant blend form, plays the role of the head of the whole expression. This is applicable in the case of blend errors whereas the arrangement of the two source words is considered to be arbitrary in case of blends in word-formation.

[3] In English, phonological length is measured by syllable. Strictly speaking, depending on the kind of phenomena, what becomes a counting unit varies. According to Kubozono (1998), in Japanese, the loanword accent rule is based on the syllable whereas the compound accent rule is based on mora. Also, mora is the unit in traditional Japanese poetry such as *haiku* and *tanka*.

The left-hand word switches to the right-hand one after the first mora in (4a), after the second mora in (4b), and after the third mora in (4c). Hence, the unclear point here is why the switch after the second mora is chosen to be employed. More generally, what determines the position of the switch point is vague.

One can think that there is a dominant way of forming blends like compound truncation, in which a truncated form made by the combination of the two initial moras from each member is the dominant output (Kubozono 2002), and the dominance operates blending. However, Irwin (2016) revealed that there is no such dominance in blending: investigation of the English Loanword Clipped Compound Integrated Database showed that in compounds treated as blending, truncation to two moras was only found in 71% of the cases for initial elements (e.g., *mara-nikku* from *mara*son 'marathon' and *pikunikku* 'picnic'), and 54% for final elements (e.g., *japani-an* from *japaniizu* 'Japanese' and *ajian* 'Asian'). Cases where each element was composed of two moras (e.g., *chai-doru* from *chairudo* 'child' and *aidoru* 'idol') accounted for 42% of relevant compounds. Based on these facts, the factors deciding the switch point are considered to be worth investigating.

3 OCP as a possible factor deciding the switch point

This section will introduce the OCP as a possible factor deciding the switch point and review previous studies in which the effects of the OCP were confirmed. The OCP was originally proposed by Leben (1973) for tonal dissimilation (i.e. change of tone when adjacent tone melodies are identical) in Mende, a language spoken in Sierra Leone and Liberia, and other African tone languages. For example, in Shona, a high tone is assigned to *a* in *mbwa* 'dog' while the tone on *a* in *ne-mabawa* 'with dog' changes to a low tone since *e* has a high tone. The naming of this phenomena was proposed by Goldsmith (1976). McCarthy (1979, 1981) expanded the range of the OCP to the segmental feature level, and now the OCP refers to the prohibition of representations in which identical specifications are adjacent.

The OCP effects can be observed in Japanese morphophonological phenomena. In the subsequent sub-sections, I will introduce the OCP effects found in the following five morphophonological phenomena: rendaku (sequential voicing), compound truncation, Japanese new type of nicknaming, *ru*-words, and *ra*-deletion.

3.1 *Rendaku* Rendaku refers to the voicing of the initial obstruent of the second member of a compound (Weijer et al. 2005; Vance 2016; e.g., *shima* 'island' + *kuni* 'country' → *shima-guni* 'island country'). The OCP effects in rendaku were confirmed at four phonological levels: consonant-vowel (henceforth CV), consonantal (henceforth C), [labial], and [voice][4] levels. The confirmation of the effects at the CV and C levels was done by Kawahara and Sano (2016). A two-way forced-choice task was conducted, in which participants chose one from compounds with rendaku and those without rendaku. As a result, the rate in which compounds with rendaku were chosen decreased when rendaku resulted in adjacent identical CV moras across a morpheme boundary (e.g., ***iga-ga**niro* from *iga* + *kaniro*) than when it did not (e.g., *iga-daniro* from *iga* + *taniro*). This tendency was also observed when rendaku caused identical consonants to be in adjacent CV moras (e.g., *i**ga-g**eniro* from *iga* + *keniro*) compared to when it did not (e.g., *iga-deniro* from *iga* + *teniro*).

The effect at the [labial] level was shown by Kumagai (2017a). In a two-way forced-choice task similar to Kawahara and Sano (2016), the declination of the rate in which compounds with rendaku were chosen was found when rendaku resulted in a sequence of consonants with the [labial] feature in adjacent moras (e.g., *nise-**bam**ara* from *nise* + *hamara*) compared to when it did not (e.g., *nise-batara* from *nise* + *hatara*).

Kawahara (2012) demonstrated the effect at the [voice] level, also known as Lyman's Law, which blocks rendaku when the second member of the compound contains an underlyingly voiced obstruent (Lyman 1894). In a naturalness rating task, the rate of compounds with rendaku was lower in those with the OCP (voice) violation (e.g., *nise-**d**aguta* from *nise* + *taguta*) than those without the OCP violation in the second member (e.g., *nise-daruna* from *nise* + *tarura*).

3.2 *Compound truncation* Compound truncation means the shortening of an existing compound phonetically by omitting a part of it while maintaining the recoverability for the original forms (Kubozono 2002; Kageyama and Saito 2016; e.g., *shaapu pensiru* 'mechanial pencil' → *shaa-pen*). Moon (2016) demonstrated CV and C levels of the OCP effects on whether truncated forms become "preserving" forms or "complemented" forms. "Preserving" forms are composed of two moras from the beginning of each members (e.g., *ruu-sute* from *ruusutaa suteetomento* 'rooster statement'). "Complemented" ones are composed of the first independent mora and third mora, which was an independent mora, in the first member, and two moras from the beginning of the second member (e.g., *rusu-sute* from *ruusutaa suteetomento*). Note that the observation of the complemented forms is limited to cases where the first syllable of one member is a heavy syllable. When preserving forms did not involve

[4] The [voice] feature is assumed to be underspecified in Japanese sonorants.

any levels of the OCP violation, they were more likely to be chosen than complemented ones with the OCP violation at the CV mora level such as *ru**su**-**su**te* or at the C level (e.g., *jip**u**-**p**are* from *jiipu pareedo* 'jeep parade') in a two-way forced-choice task.

3.3 *Japanese new type of nicknaming* In addition to the regular nicknaming process where two moras of a family name or a given name are left for truncation, this new type of nicknaming involves the alternation of /h/ at the first mora of given names to [p] (Kumagai 2017b; e.g., *kanno hadami* 'Kanno (family name) Hadami (given name)' → *kam-pada*). Kumagai (2017b) found the OCP effect at [labial] level on these nicknames. In a naturalness rating task, nicknames with the OCP (labial) violation within a single morpheme (e.g., *kasi-**pabi*** from *kasino habiyo*) were evaluated less natural than nicknames either violating the OCP or not violating one (e.g., *kasi-pada* from *kasino hadami).*

3.4 Ru-*words* *Ru*-words refer to turning a noun,[5] especially in loanwords, into a verb whose stem ends in a consonant by attaching the conjugational suffix *-ru* (Kojima 2004; e.g., *kopi* [abbreviated version of *kopii* 'copy'] + *-ru* → *kopir-u* 'make a copy'). Kojima (2004) showed that the formation of *ru*-words was affected by the OCP at CV level. In a form of open question, participants wrote down the forms with the sequence *ru-ru* across a morpheme boundary less often than those without that sequence. Specifically, if a test word ended with *ru*, the suffix tended to be attached to the test word whose last mora was excluded rather than to the entire test word (e.g., *tami**rur-u*** < *tamir-u* from *tamiru* + *-ru*). If the second mora from the last of a test word was *ru*, the suffixation to the entire test word was more likely to occur (e.g., *derutar-u* > *de**rur-u*** from *deruta* + *-ru*).

3.5 Ra-*deletion* *Ra*-deletion is a morphophonological variation (i.e. *ra*-reduced version) of potential forms which comprise a verb with a vowel-final stem and the potential suffix *-rare*[6] (Sano 2019; e.g., *mi-* 'watch'+ *-rare* 'can' → reduced form: *mi-re* 'can see'). Sano (2019) confirmed that the OCP at CV, C, and vowel (henceforth V) levels affected whether vowel-final stem verbs with the potential suffix *-rare* keep the full form or change to *ra*-reduced forms in the analysis of the Meidai Conversation Corpus and the Balanced Corpus of Contemporary Written Japanese. In the case of *re*-final stems in which *ra*-deletion caused adjacent identical CVs (e.g., ***ire-re*** from *ire-rare* **'**can insert'), no application of *ra*-deletion was observed. In the case of *e*-final stems in which vowels in adjacent syllables became identical by *ra*-deletion (e.g., *tab**e**-r**e*** from *tabe-rare* 'can eat'), *ra*-deletion was less likely to be applied compared to *i*-final stems. In the case of *ri*-final stems in which *ra*-deletion resulted in reducing the number of identical consonants in adjacent syllables (e.g., *ka**r**i-**r**e* [two *r*'s] from *ka**r**i-**r**a**r**e* [three *r*'s] 'can borrow'), the rate of applying *ra*-deletion was significantly higher than that of *i*-final stems.

3.6 *Summary* As we have seen so far, various phonological levels of the OCP target the Japanese morphophonological phenomena: (i) CV level in rendaku (Kawahara and Sano 2016), compound truncation (Moon 2016), *ru*-word formation (Kojima 2004), and *ra*-deletion (Sano 2019), (ii) C level in rendaku (Kawahara and Sano 2016), compound truncation (Moon 2016), and *ra*-deletion (Sano 2019), (iii) V level in *ra*-deletion (Sano 2019), and (iv) feature level in rendaku (Kumagai 2017a) and nicknaming (Kumagai 2017b) for [labial], and in rendaku (Kawahara 2012) for [voice]. Based on the facts (i)-(iv), I picked up the OCP as a possible factor that decides the switch point in another morphophonological phenomenon, blending. Thus, the concrete research question of this paper is whether the OCP affects the decision of switch point in Japanese blending.

Before moving on to the experimental section, I would like to mention that the avoidance of the sequence of *ru-ru* in *ru*-words shown by Kojima (2004) can be explained by not only OCP, but also haplology. In haplology, "underlying phonological material of different morphemes merges in the output" (De Lacy 1999: 52). The results of the experiment in Kojima (2004) can be interpreted in this way when using the concept of haplology: the *ru* in test words and the conjugational suffix *-ru* are represented simultaneously by one *ru* in *ru*-words. The reason why this interpretation is also possible is that only the avoidance of a sequence of the identical CV was revealed. Therefore, to answer the research question in this paper certainly, whether a sequence of identical consonants or features is avoided is necessary to be investigated since only OCP can give an explanation for the avoidance.

4 Experiments

In order to answer whether the OCP affects the decision of the switch point, this section will provide the details of two naturalness rating experiments I conducted. Based on the previous studies reviewed in the previous

[5] Word truncation is often involved in the present tense formation in *ru*-words.

[6] In the case of a verb with a consonant-final stem, potential forms are made by attaching the potential suffix *-e* to it (e.g. *ik-* 'go' + *-e* 'can' → *ike* 'can go'). However, this cannot be the target of *ra*-deletion.

section, the OCP effects are predicted to be present in blending. If this is the case, candidates violating the OCP are less natural than candidates not violating the OCP (e.g., ***mama****chi* (the OCP violation) < *maguchi* from *maguro* 'tuna' and *hamachi* 'young yellow tail').

4.1 *Experiment 1* In the first experiment, investigation of whether the OCP at CV, C, and V levels affect the decision of the switch point was conducted. The results showed that the OCP at the CV and C levels affected the decision while the effect of the OCP at the V level was not found.

4.1.1 *Method*

4.1.1.1 *Task* Experiment 1 adopted the method of a rating experiment. Each trial provided two source words and one of two candidates for the resultant blend form. The participants were instructed to evaluate the naturalness of the candidate.

The two source words consisted of names of living creatures, beverages, instruments, countries, or cities, which means that both source words were nouns following the syntactic constraint explained in Section 2.1. Also, satisfying the semantic constraints described in Section 2.1, two source words of a similar semantic content are combined. Whereas source words were both real words with three moras, candidates were novel words with three moras. In accordance with the morphological constraint illustrated in (3) and the two main phonological constraints explained in Section 2.2, two candidates for each resultant blend form were prepared. The concrete way of formation is as follows: Candidate 1 was formed by combining the first mora of the left-hand source word with the second and third moras of the right-hand source word, and Candidate 2 was formed by combining the first and second moras of the left-hand source word with the third mora of the right-hand source word. Thus, two candidates made by *mogura* 'mole' and *nezumi* 'mouse' would be *mo-zumi* for Candidate 1 and *mogu-mi* for Candidate 2. Note that the candidates did not have an unpermitted phonotactic structure of Japanese and were not identical to one of source words.

4.1.1.2 *Stimuli* The stimuli were composed of the following four conditions as presented in (5).

(5) Conditions for Experiment 1

a. Condition1: control (no violation of the OCP at CV, C, or V levels)
mogura 'mole' / *nezumi* 'mouse' → Candidate 1: *mo-zumi*, Candidate 2: *mogu-mi*

b. Condition2: violation of the OCP at CV level
i. *maguro* 'tuna' / *hamachi* 'young yellow tail' → Candidate 1: ***ma-ma****chi*, Candidate 2: *magu-chi*
ii. *unagi* 'eel'/ *iwana* 'char' → Candidate 1: *u-wana*, Candidate 2: *u****na-na***

c. Condition 3: violation of the OCP at C level
i. *bokota* 'Bogota' / *habana* 'Havana' → Candidate 1: ***b****o-****b****ana*, Candidate 2: *boko-na*
ii. *wakame* 'brown seaweed' / *mozuku* 'pickled seaweed' → Candidate 1: *wa-zuku*, Candidate 2: *wa****k****a-****k****u*

d. Condition 4: violation of the OCP at V level
i. *usagi* 'rabbit' / *kitsune* 'fox'→ Candidate 1: ***u****-ts****u****ne*, Candidate 2: *usa-ne*
ii. *otawa* 'Ottawa' / *manira* 'Manila' → Candidate 1: *o-nira*, Candidate 2: *ot****a****-r****a***

Condition 1 was a control condition in which both candidates did not violate any levels of the OCP that I investigated in the current experiment. Condition 2, 3, and 4 tested the effects of the OCP at CV, C, and V levels, respectively. In these conditions, one of the two candidates violated a level of the OCP. Each condition consisted of 10 pairs of source words. From Condition 2 to 4, the OCP violation occurred in Candidate 1 in half of the pairs as in (i) in (5) and in Candidate 2 in the other half of pairs as in (ii) in (5).

4.1.1.3 *Participants* and *procedure* 18 native Japanese speakers between the ages of 20 to 23 from the Kanto region participated in the experiment. Google Forms was used for collecting data. The experiment began with three practice questions so that the participants could understand how two source words can form a candidate. Then, they were asked to evaluate the naturalness of a given candidate using a six-point scale (1 for most unnatural, 6 for most natural). The concrete instruction is as follows: (1 つ目の単語)と(2 つ目の単語)を掛け合わせて新しい XXX を作りました。/ (1 つ目の 単語)出身の人と(2 つ目の単語)出身の人がお笑いコンビを組みました。その名前として「(候補 1 または 2) 」はどのくらい自然ですか？'We made a new XXX by crossing over (the left-hand word) and (the right-hand word). / A person from (the left-hand word) and another person from (the right-hand word) made a new comedy duo. How natural is (Candidate 1 or 2) for the name?' An example is given in Figure 1.

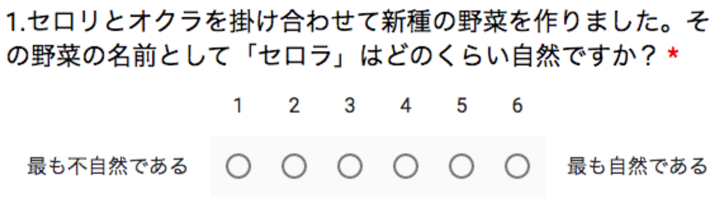

Figure 1: Screenshot of a trial

Two source words and one candidate were written in katakana. Two candidates formed by a pair of source words were presented in different sections so that participants would evaluate both candidates. Four lists were created per section which consisted of 63 pairs of source words (3 practices, 40 targets: 10 pairs of source words × 4 conditions, and 20 fillers). The order of the targets and fillers was randomized per list.

4.1.1.4 *Statistical analysis* Since the responses were categorical but ordered (i.e. ordinal), Cumulative Link Mixed Model (CLMM) was used to compare the ratings of the candidates that had the OCP violation in Conditions 2, 3, and 4 with candidates in Condition 1, which did not have the OCP violation. The independent variable was the presence of the OCP violation, and the dependent variable was the rating. Participants and items were encoded as random factors.

4.1.2 *Results* Figure 2 indicates the average rate of candidates in Condition 1 and candidates with the OCP violation in Condition 2, 3, and 4.

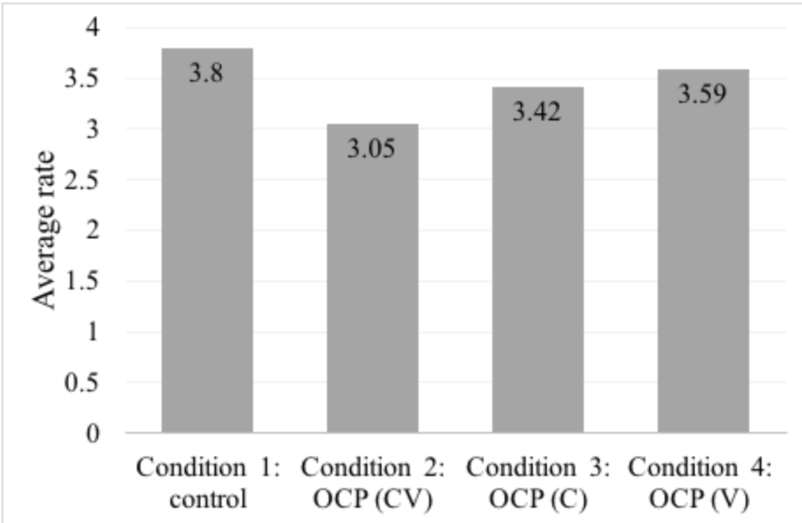

Figure 2: Average rate of candidates in Condition 1 and candidates with the OCP violation in Condition 2, 3, and 4 in Experiment 1

The rate of candidates without the OCP violation in Condition 1 was significantly higher than those with the OCP (CV) violation in Condition 2 (Condition 1: 3.8 vs. Condition 2: 3.05; $z = -5.718$, $p < .001$) and those with the OCP (C) violation in Condition 3 (Condition1: 3.8 vs. Condition 3: 3.42; $z = -2.375$, $p < .05$), respectively. Also, the difference of the rate between candidates without the OCP violation and those with the OCP (V) violation in Condition 4 was not found statistically (Condition1: 3.8 vs. Condition 4: 3.59; $z = -1.261$, *n.s.).*

Note that the rate was not statistically different between Candidate 1 and Candidate 2 in Condition 1 (Candidate 1: 3.89 vs. Candidate 2: 3.71; $z = -0.949$, *n.s.).* Also, the rate of candidates with the OCP violation did not statistically differ depending on the position of the violation (OCP (CV): Candidate 1 with the OCP violation: 3.18 vs. Candidate 2 with the OCP violation: 2.91; $z = -1.725$, *n.s.*; OCP (C): Candidate 1 with the OCP violation: 3.63 vs. Candidate 2 with the OCP violation: 3.2; $z = -1.279$, *n.s.*; OCP (V): Candidate 1 with the OCP violation: 3.7 vs. Candidate 2 with the OCP violation: 3.48; $z = -0.522$, *n.s.*).

4.1.3 *Discussion* The results concerning the OCP at CV and C levels followed the prediction in which candidates violating the OCP are less natural than those not violating one, while the results concerning the OCP at V level did not. Therefore, there are CV and C levels of the OCP effects on the decision of switch point. By contrast, the foundation of the OCP effect at V level could not be accomplished.

4.2 *Experiment 2* The second experiment investigated the effects of the OCP at C and feature levels. This study focused on the features [labial] and [voice]. Following Mester and Ito (1989), the [voice] feature is assumed to be a contrastive one, which means that sonorant voicing is unspecified.[7] The results showed that the OCP at the C level affected the decision whereas the OCP at the feature level did not.

4.2.1 *Method*

4.2.1.1 *Task* Experiment 2 employed the same rating task as Experiment 1.

4.2.1.2 *Stimuli* The stimuli were chosen based on the following four conditions.

(6) Conditions for Experiment 2

a. <u>Condition 1: control (no violation of the OCP at CV, C, V, [labial], or [voice] levels)</u>

b. <u>Condition 2: violation of the OCP at C level</u>

c. <u>Condition 3: violation of the OCP at [labial] level</u>

i. *berona* 'Verona' / *omaha* 'Omaha'→ Candidate 1: ***b**e-**m**aha*, Candidate 2: *bero-ha*

ii. *kabosu* 'kabosu (type of citrus fruit)' / *kurumi* 'walnut'→ Candidate 1: *ka-rumi*, Candidate 2: *ka**b**o-**m**i*

d. <u>Condition 4: violation of the OCP at [voice] level</u>

i. *ginia* 'Guinea' / *tsubaru* 'Tuvalu' → Candidate 1: ***g**i-**b**aru*, Candidate 2: *gini-ru*

ii. *ribia* 'Libya' / *kanada* 'Canada' → Candidate 1: *ri-nada*, Candidate 2: *ri**b**i-**d**a*

In Condition 1, which was the control condition, there was no violation of the OCP at CV, C, V, [labial], or [voice] levels in both candidates. Condition 2 was for replicating the results concerning the OCP effect at C level in Experiment 1. Note that target consonants were limited to labial consonants or voiced obstruents in the current experiment. Whether the OCP at [labial] and [voice] levels affects the decision of the switch point was investigated in Condition 3 and 4 respectively, using a candidate with the corresponding OCP violation. Note that a sequence of the identical consonants was not included in candidates. Each condition contained 10 pairs of source words. Except for Condition1, half of the pairs violated the OCP in Candidate 1 as in (i) in (6), and the other half violated it in Candidate 2 as in (ii) in (6).

4.2.1.3 *Participants* and *procedure* 26 native speakers of Japanese aged from 18 to 27 years old participated. They have spent most of their lives in the Kanto region. The procedure which Experiment 2 employed was the same as Experiment 1 excluding the following two points. One was that the number of fillers per section was 18 pairs of words in Experiment 2. The other was that a simple explanation was provided for pairs of words where the semantic contents were birds, plants, or cities since compared to Experiment 1, unfamiliar words to the participants were used in Experiment 2 to have a certain number of items.

4.2.1.4 *Statistical analysis* The same way of statistical analysis as Experiment 1 was employed in Experiment 2.

4.2.2 *Results* The average rate of candidates in Condition 1 and candidates with the OCP violation in Condition 2, 3, and 4 is shown in Figure 3.

[7] As its evidence, Mester and Ito (1989) showed that blocking of rendaku does not occur even when the second member contains a sonorant as well as a voiceless obstruent. In fact, the results will not change even if I adopt [+voice] here.

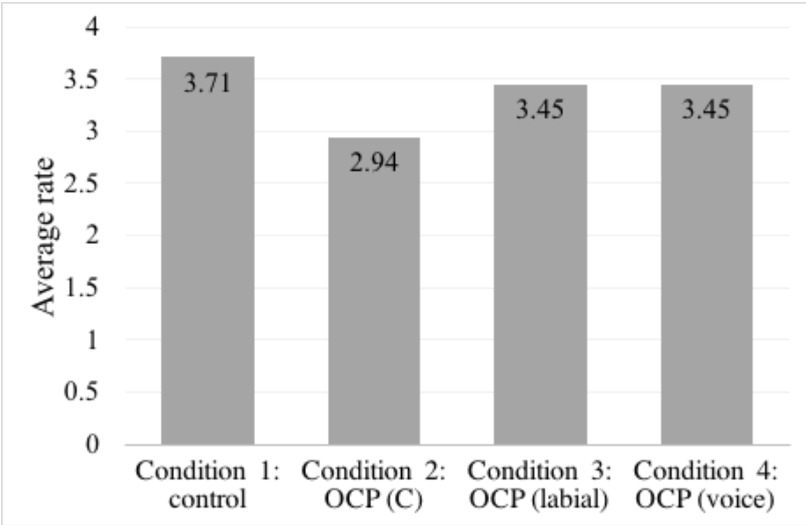

Figure 3: Average rate of candidates in Condition 1 and candidates with the OCP violation in Condition 2, 3, and 4 in Experiment 2

The rate of candidates without the OCP violation in Condition 1 was significantly higher than that of those with the OCP (C) violation in Condition 2 (Condition 1: 3.71 vs. Condition 2: 2.94; $z = -3.8$, $p < .001$). On the other hand, in terms of the rate, candidates without the OCP violation did not statistically differ from candidates with the feature level of the OCP violation in Condition 3 and 4 ([labial]: Condition 1: 3.71 vs. Condition 3: 3.45; $z = -1.649$, *n.s.*; [voice]: Condition 1: 3.71 vs. Condition 4: 3.45; $z = -1.395$, *n.s.*).

Note that the difference in the rate between Candidate 1 and Candidate 2 was not found statistically in Condition 1 (Candidate 1: 3.82 vs. Candidate 2: 3.6; $z = -0.937$, *n.s.*). There was no statistically significant difference between the rate of Candidate 1 with the OCP violation and that of Candidate 2 with the OCP violation (OCP (C): Candidate 1 with the OCP violation: 3.12 vs. Candidate 2 with the OCP violation: 2.75; $z = -1.304$, *n.s.*; OCP (labial): Candidate 1 with the OCP violation: 3.28 vs. Candidate 2 with the OCP violation: 3.62 ; $z = 1.71$, *n.s.*; OCP (voice): Candidate 1 with the OCP violation: 3.63 vs. Candidate 2 with the OCP violation: 3.27; $z = -1.081$, *n.s.*).

4.2.3 *Discussion* The prediction in which candidates with the OCP violation are less natural than those without the violation turned out to be true for the results concerning the OCP at C levels, but did not hold for those concerning the OCP at feature level. Hence, I succeeded in replicating the results in Experiment 1, which suggested that the OCP at the C level gave the effect on determining the switch point, in spite of limiting the target consonants to voiced obstruents and labial consonants. However, the effects of the OCP at the [labial] and [voice] levels could not be found.

5 General discussion

Through the two experiments explained in the previous section, the OCP was found to affect the decision of the switch point. To be precise, among various phonological levels of the OCP, the effects of the OCP at CV and C levels were confirmed whereas those of the OCP at V and feature levels were not. Based on them, this section will discuss two main points. Firstly, the similarities will be discussed between the results of the two experiments in blending and the results of the previous experiments in other morphophonological phenomena. Secondly, I will discuss the difference in results among these morphophonological phenomena and the reason for the difference.

The observation of the OCP effects at both CV and C levels in Japanese blending suggests that the OCP effects truly exist in blending as well as rendaku (Kawahara 2012; Kawahara and Sano 2016: Kumagai 2017a), compound truncation (Moon 2016), nicknaming (Kumagai 2017b), and *ra*-deletion (Sano 2019). The presence of the OCP effect at the C level prevents from explaining the avoidance phenomenon by haplology. In addition, in the two experiments, I excluded the combination of two words where the second moras or consonants were identical in both source words (e.g., *i**wa**na* 'char'and *sa**wa**ra* 'Spanish mackerel'), so that there is no ambiguity in the origin of the mora or consonant in candidates. Thus, the existence of genuine OCP effects is supported.

The OCP effects have been found at V level on *ra*-deletion (Sano 2019) and at feature level on rendaku (Kawahara 2012; Kumagai 2017a) and nicknaming (Kumagai 2017b). However, the two experiments from this study did not find these influences. The fact that the absence of the OCP effects at the V and feature levels in blending may come from the unique status of this morphophonological phenomenon.[8] Before I explain in detail, let me make clear the presupposition I made deriving from the similarity effects. This states that "[t]he more similar the two elements are, the stronger the identity avoidance effect is" (Suzuki 1998: 17). The presence of this effect was confirmed between CV and C levels by Kawahara and Sano (2016) in rendaku: the rate in which compounds with rendaku were chosen was significantly lower when rendaku caused adjacent identical CV moras than when it caused identical consonant to be in adjacent CV moras. Also, the confirmation of the effect was done in compound truncation by Moon (2016): compared to complemented forms with the OCP violation at the C level, those with violation of the OCP at the CV level were chosen significantly less often. Moreover, Kumagai (2017a) found the effect between the C and feature levels in rendaku: the more similar the two adjacent labial consonants were, the lower the rate in which compounds with rendaku were chosen. By considering these results above, we can construct the strongness hierarchy of the OCP effects among phonological levels: CV level > C level > feature level. The position of the V level in this hierarchy cannot be decided straightforwardly. However, based on its non-commonality[9] and the presence of a remnant of old Japanese in which vowel harmony was present in native words (e.g., ***m**a**n**a**ko*** 'eye' from *ma-no-ko* 'eye-GEN-child' and ***k**u**n**u**gi*** 'sawtooth oak' from *ku-no-ki* 'eat-GEN-tree'), it is plausible to think that the position is below the C level. As for the precise hierarchical relation between the V and feature levels, I will leave this as a subject for further research. The probability of the occurrence of the sequence of identical moras, segments, or features might be related to the hierarchy such that the stronger the level of the OCP effects is, the lower the emergence of the identity at the level is. CV level, whose OCP effect is the strongest, has the lowest probability, followed by C levels, and V and feature levels. When proceeding the discussion, I assume that the level to which OCP effects extend depends on the morphophonological phenomenon.

Therefore, I will first discuss what unique points in blending lead to the decision of the range of subject of the OCP effects from CV to C levels. Three unique characteristics of blending will be pointed out by comparing it with other morphophonological phenomena. Blending will be compared mainly with compound truncation since they are similar in that the processes involve taking out parts of two words. After the comparison between them, I will add some comments on rendaku and *ra*-deletion about the difference from blending.

The first unique point is the paradigmatic relation between two source words in blends whereas two members which will be truncated are syntagmatically related (Kubozono 1990). The following three constraints function as evidence. First, two source words in blending must belong to the same syntactic category and have a semantic relation. However, it is not necessarily in truncation. For instance, in *pasokon*, the truncated form of *paasonal konpyuutaa* 'personal computer', the parts of speech of the two members are different, and they do not have a semantic relation. Second, blends consist of the initial part of the left-hand source word and the non-initial part of the right-hand word while truncated forms generally consist of the two initial moras from each member of the compound. Third, the common switch point between the two source words is related to the same phonological length of the right-hand word as the resultant form in blending. On the other hand, there is no such length constraint on the relationship between each member and the truncated form. In rendaku, two members of a compound are syntagmatically related, and the same relation is hold between a verb with a vowel-final stem and the potential suffix *-rare,* which must be sticked before *ra*-deletion. As the evidences, in rendaku, the parts of speech of the two members do not need to be the same nor do the members need to be semantically related. The process involved is the voicing of the initial obstruent of the second member, and rendaku is not constrained by the phonological length. In *ra*-deletion, the vowel-final stem verb and the potential suffix *-rare* obviously belong to a different syntactic category and have different meanings. *Ra* in the potential suffix *-rare* is deleted after the suffix is attached to a vowel-final stem verb. The phonological length does not constrain the *ra*-deletion.

The second unique point is the relationship between the two source words and the resultant blend form. By blending two words, new entities or concepts are created. This can be supported by the first unique point (i.e. syntagmatic aspect between two source words) because one slot is filled as a result of merging two source words, which are in a competitive relationship. Hence, the resultant blend form indicates the new entities or concepts. However, it is not the case that truncation creates something new from a compound. There is no difference in meaning between a compound and the truncated form. Rendaku and *ra*-deletion are basically under the same situation as compound truncation.

The third unique point is the low productivity of blends in word-formation. As I explained at the beginning of Section 1 and led by the second unique point (i.e. the creative aspect of new entities or concepts), the use is

[8] Note that based on the fact that confirmed the OCP effect at the feature level were not across a morpheme boundary (Kawahara 2012; Kumagai 2017a; Kumagai 2017b), the reason for the non-foundation in blending might be related to the morpheme boundary. The discussion about the relationship between the presence or absence of OCP effects and morpheme boundary is left for future study.

[9] According to Bye (2011), all the known examples of vowel dissimilation involve vowel height. However, there are no examples involving backness and rounding. By contrast, wide variety of features such as [labial], [coronal], [lateral], [rhotic], [voice], [spread glottis], [nasal], [continuant], which are mainly used in consonants, are confirmed to participate in dissimilation.

limited to naming new entities or concepts such as fictional animals and plants or new products and phenomena. Thus, blend forms often disappear without being established as general vocabulary (Kubozono 1995). The infrequency of blending was shown by Irwin (2016) using the English Loanword Clipped Compound Integrated Database: the rate of the occurrence of blending was only 5.6% (24 out of 432 compounds). By contrast, truncation is productive, and truncated forms are often used to reduce the production cost in our daily life without recognizing that the forms are a truncated version (Ito 2011). The productivity of rendaku and *ra*-deletion is also considered high because we apply them unconsciously (younger generation in particular for the application of *ra*-deletion).

Next, I will explain how these unique characteristics of blending are connected to the absence of the OCP effects at V and feature levels. The reason for the absence might be related to the interaction between the unique characteristics of blending and the probability of occurrence of the sequence of identical moras, segments, or features. Since the degree of freedom in the formation of blends is low (i.e. a few candidates for a resultant blend form can be created), the probability has a direct impact on blending, and the difference in the probability is reflected as the presence or absence of the OCP effects. Rendaku and *ra*-deletion always have a choice regarding whether or not they are applied. Therefore, even under the situation where the identity might emerge, the emergence can be easily prevented by stopping the application. On the other hand, once compound truncation and blending are decided to be carried out, the choice of stopping their processes will disappear. The identity which might emerge in the process of compound truncation can be avoided by changing the way of taking moras from each member since there are several ways of making truncated forms. By contrast, the first unique point of blending based on the constraints in Section 2.1 and 2.2 does not give such freedom in candidates for a resultant blend form when avoiding the identity without stopping the process itself. Thus, blending has to make a compromise with the avoidance of the sequence of identical V or features, whose probability of emergence is high.

6 Conclusion

The ultimate goal of this study was to discover the factors that decide the switch point in Japanese blending. Concretely, this paper has investigated the effects of the OCP as a possible factor that determines the switch point. As the result, there seems to be variation among phonological levels: the effects were confirmed at CV and C levels, but not in V and feature levels. What these results suggest is that blending is affected by the OCP at the CV and C levels just like other morphophonological phenomena. By contrast, blending differs in that the OCP effects at the V and feature levels are absent, which derives from the unique status of blending.

The following three points should be kept in mind as limitations for this study. The first point is that real words were used as source words in the two experiments. This caused phonotactics in candidates and the number / kind of segments which were the target of the OCP violation not to be controlled. In other words, I could not take the following two things into consideration. One is blocking effects (i.e. "non-occurrence of one form due to the simple existence of another" (Aronoff 1976: 43)) on candidates. For example, one candidate formed by *suzuki* 'sea bass' and *iwashi* 'sardine' was *suzushi*. Since Japanese has the adjective *suzushii* 'cool', the naturalness for *suzushi* might be reduced. The other is the possibility that some segments or combinations are more likely to become a target of the OCP effects, but others are not. For instance, the sequence of *ta* (e.g., ***tata****chi* from *tanuki* 'racoon dog' and *itachi* 'weasel') is not preferred and more avoidable than that of *ma* (e.g., ***mama****chi* from *maguro* 'tuna' and *hamachi* 'young yellow tail'). Also, the etymology of source words could not be unified even though permitted sound patterns vary by "lexical stratum" in Japanese (i.e. Yamato, Sino-Japanese, and foreign words; Ito and Mester 1999). Taking the singleton-p as an example, it is permitted in foreign words, but not in Yamato and Sino-Japanese (Ito and Mester 1999). Nevertheless, it was prioritized to have a certain number of items instead of unifying the etymology of source words.

The second limitation is the uncertainty of how participants recognized the etymological morpheme boundary of the source words even though they are synchronically simplex words. For example, although *mejiro* 'white-eye' is a synchronically simplex word, there is an etymological morpheme boundary after the first mora, and it can be divided into *me* 'eyes' and *shiro* 'white.' Such a morpheme boundary might influence the naturalness rating. Taking two candidates for a resultant blend form formed by *mejiro* and *karasu* 'crow' as examples, the candidate following the boundary *me-rasu* might be evaluated more natural than the one not following the boundary *meji-su*. However, I could not control the influence in the experiments.

The third limitation is that the placement of word accent by participants on source words was unclear. This led to the inability to rule out the possibility that accents on source words affect the naturalness rating. Since Labrune (2002) suggested that in compound truncation, the position of the accent nucleus in the input determines the truncated form, it would not be surprising if accents give the influence on blending.

Although there are some limitations, this study can contribute to a deeper understanding of the characteristics of blending itself, the uniqueness of blending compared to other morphophonological phenomena, and various phonological levels of the OCP. In order to heighten the degree of contribution regarding the OCP effects on blending, I have come up with three topics for future research. The first topic is the OCP effects at other levels.

In addition to other feature levels such as [coronal] and [dorsal], phonemic and orthographic levels are possible levels. If the former is the case, *ho-kuso* [ho-kuso] from *hokene* and *mikuso* will be evaluated more natural than *ho-fuso* [**h**o-**ɸ**uso] from *hokene* and *mifuso*. If the latter is the case, the naturalness of *ha-sake* はさけ from *hasoro* and *masake* will be higher than that of *ha-pake* はぱけ from *hasoro* and *mapake*.

The second topic is the domain where the OCP operates in blending. In rendaku, although the OCP (voice) operates regardless of whether a sequence of consonants with the [voice] feature in adjacent moras or not (Kawahara 2012; Kawahara and Sano 2014a), the OCP (labial) operates only when a sequence of consonants with the [labial] feature is in adjacent moras (Kumagai 2017a). Thus, depending on the level of the OCP, the domain can be changed in blending.

The third topic is whether lexically-violated OCP (i.e. source words with the OCP violation; e.g., *ba**nana*** 'banana') changes the location of the switch point. This idea derives from Kawahara and Sano (2014b). They showed the inclination of the rate of rendaku application when two members, which will constitute a compound, have the sequence of identical CV mora. For example, the sequence of identical CV mora, which consists of the last mora of the first member and the first mora of the second mora (e.g., *ika-ganiro* from ***ika*** + ***ka**niro*), increased the rate of rendaku application. In case of blending, the effect of the lexically-violated OCP is considered to be present when the naturalness of candidates resolving the lexical OCP violation is rated higher than those in which the violation is remained. For instance, as a resultant blend form made by *ichigo* 'strawberry' and *ba**nana***, *ichi-na* has higher naturalness than *i-**nana***.

Furthermore, I would like to shed light on another factor concerning the decision of the switch point. A possible factor is the identical segment located on the same position in two source words. This idea is based on the results of Kubozono's (1990) survey, in which the exception to the phonological length constraint on blending (i.e. the right-hand source word and the resultant blend form have the same phonological length) emerged when there was a phonemic overlapping between the two source words. This overlapping segment will be integrated into resultant blend forms, so that the number of segments taken from each source word which are retained in the blend forms will be maximized. For instance, *pia**n**o* and *hamo**n**ika* have *n* at the third mora. As a possible reason why *pianika* is chosen as the blend form, the switch after the second mora is the position which results in the sufficient use of the identical segment *n.* The investigation of these effects listed so far can come closer to reaching the ultimate goal of this study, that is to discover the factors that decide the switch point in Japanese blending.

References

Aronoff, Mark (1976) *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

Bye, Patrik (2011) Dissimilation. In van Oostendorp, Marc, Ewen, Colin J., and Hume, Elizabeth (eds.) *The Blackwell Companion to Phonology. Volume III: Segmental Processes.* Oxford: Wiley-Blackwell. 1408-1433.

De Lacy, Paul (1999) Morphological Haplology and Correspondence. In De Lacy, Paul and Nowak, Anita (eds.) *University of Massachusetts Occasional Papers: Papers from the 25th Reunion.* Amherst, MA: GLSA. 51-88.

Goldsmith, John (1976) *Autosegmental Phonology*. Doctoral dissertation, MIT.

Irwin, Mark (2016) The Morphology of English Loanwords. In Kageyama, Taro and Kishimoto, Hideki (eds.) *Handbook of Japanese Lexicon and Word Formation.* Berlin: Mouton de Gruyter. 161-200.

Ito, Junko and Mester, Armin (1999) The structure of the phonological lexicon. In Tsujimura, Natsuko (ed.) *The Handbook of Japanese Linguistics*. Malden, MA, and Oxford, U.K.: Blackwell Publishers. 62-100.

Ito, Mitsu (2011) About Blending. *Studies of liberal arts* 17 (3), 1-15.

Kageyama, Taro and Saito, Michiaki (2016) Vocabulary strata and word formation processes. In Kageyama, Taro and Kishimoto, Hideki (eds.) *Handbook of Japanese Lexicon and Word Formation.* Berlin: Mouton de Gruyter. 11-50.

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

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

Kawahara, Shigeto and Shin-ichiro, Sano (2014b) Identity Avoidance and rendaku. *Proceedings of Phonology 2013*.

Kawahara, Shigeto and Shin-ichiro, Sano (2016) Rendaku and identity avoidance: Consonantal identity and moraic identity. In Vance, Timothy and Irwin, Mark (eds.) *Sequential Voicing in Japanese: Papers from the NINJAL Rendaku Project*. Amsterdam: John Benjamins. 47-55.

Kojima, Chisato (2004) OCP in the Derived Environment. *Phonological studies* 7, 1-8.

Kubozono, Haruo (1990) Phonological constraints on blending in English as a case for phonology-morphology interface. *Yearbook of Morphology* 3, 1-20.

Kubozono, Haruo (1995) *Gokeisei to oninkouzou.* [Word formation and phonological structure.] Tokyo: Kuroshio Publishers.

Kubozono, Haruo (1998) On the Universality of Mora and Syllable. *Journal of Phonetic Society of Japan 2* (1), 5-15.

Kubozono, Haruo (2002) *Shingo-wa koushite tsukurareru.* [This is how new words are created.] Tokyo: Iwanami Shoten.

Kumagai, Gakuji (2017a) Testing the OCP-labial effect on Japanese rendaku. Unpublished Manuscript. NINJAL. http://ling.auf.net/lingbuzz/003290.

Kumagai, Gakuji (2017b) The orthographic Lyman's Law and OCP effects in Japanese nicknames. *The 25th Japanese/Korean Linguistics Conference.* University of Hawai'i at Manoa.

Labrune, Laurence (2002) The Prosodic Structure of Simple Abbreviated Loanwords in Japanese: A Constraint-based Account. *Journal of the Phonetic Society of Japan* 6 (1), 98–120.
Leben, William (1973) *Suprasegmental Phonology*. Doctoral dissertation, MIT.
Lyman, Benjamin S. (1894) Change from surd to sonant in Japanese compounds. *Oriental Studies of the Oriental Club of Philadelphia*, 1-17.
Marchand, Hans (1969) *The categories and types of present-day English word-formation: A synchronic-diachronic approach*. Beck.
McCarthy, John J. (1979) *Formal Problems in Semitic Phonology and Morphology*. Doctoral dissertation, MIT.
McCarthy, John J. (1981) A Prosodic Theory of Nonconcatenative Morphology. *Linguistic Inquiry* 12, 373-418.
Mester, Armin and Ito, Junko (1989) Feature Predictability and Feature Specification. *Language* 65, 258-293.
Moon, Changyun (2016) The Influence of OCP on Word Truncation. *Tsukuba Japanese linguistics* 20, 109-126.
Sano, Shin-ichiro (2019) Patterns of Variable *Ranuki* in Japanese: Identity Avoidance and Register. *Phonological Studies* 22, 75-82.
Suzuki, Keiichiro (1998) *A Typological Investigation of Diissimilation*. Doctoral dissertation, University of Arizona.
Vance, Timothy (2016) Introduction. In Vance, Timothy and Irwin, Mark (eds.) *Sequential Voicing in Japanese: Papers from the NINJAL Rendaku Project*. Amsterdam: John Benjamins. 1-12.
Weijer, Jeroen van de, Nanjo, Kensuke, and Nishihara, Tetsuo (2005) Voicing in Japanese. In Weijer, Jeroen van de, Nanjo, Kensuke, and Nishihara, Tetsuo (eds.) *Voicing in Japanese.* Berlin: Mouton de Gruyter. 1-4.