Identity Avoidance and Lyman’s Law*

Shigeto Kawahara and Shin-ichiro Sano

Paper to appear in Lingua

Abstract

Rendaku is a morphophonological process whereby the initial consonant of the second member of compounds becomes voiced. One famous factor that blocks rendaku is Lyman’s Law: when a second element already contains a voiced obstruent, rendaku is blocked. This blockage of rendaku due to Lyman’s Law is almost exception-less in the contemporary Japanese, and thus has been treated as if it applies uniformly to all forms that fit the structural description. However, our current experiment shows that this uniformity assumption does not hold. Concretely, the experiment reveals a hitherto unnoticed generalization: among those structures that violate Lyman’s Law, there are some that are more disfavored than others. More specifically, Japanese speakers disfavor structures with two adjacent identical CV moras with a voiced obstruent onset (e.g. /dadana/) more than structures that merely contain two voiced obstruent onsets (e.g. /dogara/). In addition to this new descriptive discovery in Japanese phonology, this paper makes three contributions to general linguistic theory: (i) the importance of experimentation in linguistic research; (ii) the role of grammar that cannot be deduced from lexical patterns; (iii) parametrization of the locality of dissimilatory effects within a single language.

*This paper was presented as a portion of our talk at the Keio phonetics-phonology study group in July 2013, Tokyo Circle of Phonologists in October 2013 and International Conference on Phonetics and Phonology in December 2013. We received much valuable advice from the audiences at these occasions. We are also grateful to Donna Erickson, Kazutaka Kurisu, Jeff Moore, Yukiko Sugiyama and three anonymous reviewers for their comments on previous versions of the paper. The editorial advice from the associate editor, Ann-Michelle Tessier, was very helpful in finishing this work. This project is supported by JSPS KAKENHI Grant #26770147 to the first author and #25770157 and #25280482 to the second author, and #26284059 to both of the authors. All remaining errors are ours.
1 Introduction

1.1 Synopsis

Rendaku is a morphophonological process whereby the initial consonant of the second member of compounds becomes voiced (e.g. /tako/ ‘octopus’ → /oo-dako/ ‘big octopus’).\(^1\) One famous factor that blocks rendaku is Lyman’s Law (Lyman, 1894): when a second element already contains a voiced obstruent, rendaku is blocked (/tokage/ ‘lizard’ → /oo-tokage/ ‘big lizard’). This law itself has been studied from several different perspectives: theoretical (Ito & Mester, 1986, 2003a; Kawahara & Zamma, to appear; Mester & Ito, 1989), historical (Unger, 1975; Vance, 2005, 2007), and experimental (Ihara et al., 2009; Kawahara, 2012, to appear; Vance, 1980). This blockage of rendaku has only a few lexical exceptions, and has been formalized as the OCP effect on the [+voice] feature within a morpheme (Ito & Mester, 1986).

This blockage of rendaku due to Lyman’s Law is almost exception-less in contemporary Japanese and thus has been treated as if it applies uniformly to all forms that fit the structural description (“two voiced obstruents within a morpheme”). However, our current experiment shows that this uniformity assumption does not hold. Concretely, the experiment reveals a hitherto unnoticed generalization: among those structures that violate Lyman’s Law, there are some that are more disfavored than others. More specifically, Japanese speakers disfavor structures with two adjacent identical CV moras with a voiced obstruent onset (e.g. /dadana/) more than structures that merely contain two voiced obstruent onsets (e.g. /dogara/). In other words, among those structures that violate Lyman’s Law, those that also violate Identity Avoidance (Yip, 1998) are especially dispreferred by Japanese speakers. We conclude that this finding instantiates a case in which experiments reveal a preferential hierarchy between two sets of structures which cannot be learned from a pattern in the lexicon.

Going beyond our new descriptive discovery about rendaku and Lyman’s Law, this paper con-

---

\(^1\)Here and throughout, when we refer to Japanese words and nonce words, we deploy the standard phonemic transcription system rather than the IPA transcriptions, because phonetic details are not relevant in this paper. See Vance (2008) for the correspondence between the standard phonemic transcription system and the actual phonetic realizations in Japanese.
tributes to general linguistic theorization in three ways: (i) the effect of Identity Avoidance on items that already violate Lyman’s Law has been unknown despite the fact that rendaku is very well studied in the past literature, and this new finding thus suggests that experimentation is important in linguistic research; (ii) since in the contemporary Japanese lexicon there are a few lexical items that violate Lyman’s Law, the distinction between mere Lyman’s Law violation and simultaneous violation of Lyman’s Law and Identity Avoidance cannot be deduced from the lexical patterns, suggesting the role of grammar that goes beyond lexical patterns; (iii) Identity Avoidance is effective only between adjacent positions, whereas Lyman’s Law does not show this locality restriction, despite the fact that they are both dissimilatory forces—this result supports the idea that the domain of dissimilarity forces can be parameterized (Odden, 1994; Pulleyblank, 2002; Suzuki, 1998), even within a single language.

1.2 Background and the current experiment

One factor that has been claimed to affect rendaku application is the avoidance of identical CV mora sequences (i.e. Identity Avoidance: Yip 1998): rendaku is less likely to apply when rendaku results in two adjacent identical moras (e.g., schematically, */iga+ganiro/ from /iga+/kaniro/). This effect of identity avoidance was first hinted at by Sato (1988) and reiterated by Labrune (2012). Irwin (2014) on the other hand denies that there is such a effect based on the lack of statistical evidence in the lexicon. However, in a recent experimental study using nonce words, Kawahara & Sano (2014) show that avoidance of two adjacent identical CV moras does reduce the applicability of rendaku. The nonce-word experiment shows that rendaku is less likely to apply when it would result in two identical adjacent moras across a morpheme boundary (e.g.


3Avoidance of identical structures in proximity, which has been formalized in the theoretical literature as the OCP effects, is observed at various phonological levels in many different languages (e.g. Alderete & Frisch 2007; Berent 2013; Frisch 2004; Frisch et al. 2004; Leben 1973; McCarthy 1979, 1986; Odden 1994; Suzuki 1998; Yip 1988, 1998).
/iga+ganiro/) than when it would not (e.g. /iga+daniro/). That study, however, is limited in that it tested the effect of Identity Avoidance only in environments where Lyman’s Law is irrelevant. The Identity Avoidance effect identified by Kawahara & Sano (2014) can thus be used to test the uniformity assumption about Lyman’s Law—whether all Lyman’s Law violations are equally disfavored by Japanese speakers.

Therefore, extending on Kawahara & Sano (2014) (and, to a lesser extent, other previous experimental studies on rendaku and Lyman’s Law (Ihara et al., 2009; Kawahara, 2012; Vance, 1980)), the current experiment tests the effects of Identity Avoidance on those words that already violate Lyman’s Law. The primary question is whether, when rendaku violates both Identity Avoidance and Lyman’s Law, it would be judged to be worse than when it violates only Lyman’s Law. The experiment shows that Japanese speakers judge the simultaneous violation of Identity Avoidance and Lyman’s Law as worse than a mere violation of Lyman’s Law. The uniformity assumption is therefore shown not to hold.

The current experiment was a two-way forced-choice wug-test (Berko, 1958) on rendaku, testing the effect of Identity Avoidance at the CV moraic level on those words that violate Lyman’s Law. Identity Avoidance at the moraic level—not at the consonantal level—was tested, because CV sequences constitute an important phonological unit in Japanese; i.e. the mora (Ito, 1989; Kubozono, 1989; Labrune, 2012).\(^4\) The current experiment also tests the locality of Identity Avoidance and Lyman’s Law. The experiment built on a naturalness judgment study by Kawahara (2012), itself being a follow-up study of Ihara et al. (2009), and Vance (1980) in particular (see also Kawahara to appear for a recent review of the experimental studies of rendaku).

\(^4\)Two anonymous reviewers asked if Identity Avoidance effects at the consonant level should be tested in a future study. Indeed, our new experiment in progress attempts to tease apart the effect of Identity Avoidance at the consonantal level and the effect of Identity Avoidance at the moraic level. The preliminary analysis shows that the effect of Identity Avoidance at the moraic level is clear, whereas the effect of Identity Avoidance at the consonant level is tangible but less substantial.
2 Method

2.1 Stimuli

The stimulus structure of the experiment is illustrated in Table 1. There were in total five conditions in the current experiment: (i) no Lyman’s Law violation, (ii) local Lyman’s Law violation without Identity Avoidance violation, (iii) local Lyman’s Law violation with Identity Avoidance violation, and (iv) non-local Lyman’s Law violation without Identity Avoidance violation, and (v) non-local Lyman’s Law violation with Identity Avoidance violation. In the first control condition, where there are no Lyman’s Law violations, there were 6 nonce items for each type of rendaku-undergoing consonant. Among those structures that would result in a violation of Lyman’s Law after rendaku (conditions (ii)-(v)), there were 3 items for each rendaku-undergoing consonant.

The stimuli of the current experiment are listed in Table 2. The experiment included all the four consonants that can potentially undergo rendaku in Japanese (/t, k, s, h/). All the stimuli consist of only CV light syllables, which coincide with one mora in the phonology of Japanese (Ito, 1989; Kubozono, 1989). Some stimuli were taken from Vance (1980) and Kawahara (2012) (shown by **) or from Kawahara (2012) alone (shown by *). There were a total of 72 stimuli (4 consonants × (6 items × 1 condition + 3 × 4 conditions)).

Table 1: Illustration of each condition. LL=Lyman’s Law; IA=Identity Avoidance.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) No-violation:</td>
<td>/tamuma/ → /damuma/</td>
</tr>
<tr>
<td>(ii) Local LL:</td>
<td>/taguta/ → /daguta/</td>
</tr>
<tr>
<td>(iii) Local LL &amp; IA:</td>
<td>/tadaru/ → /dadrar/</td>
</tr>
<tr>
<td>(iv) Non Local LL:</td>
<td>/tatuga/ → /datuga/</td>
</tr>
<tr>
<td>(v) Non Local LL &amp; IA:</td>
<td>/takuda/ → /dakuda/</td>
</tr>
</tbody>
</table>

/h/ becomes /b/ after rendaku, and therefore, for example, /hibiro/ would become /bibiro/, violating Identity Avoidance.
Table 2: The stimulus table. Some stimuli were taken from Vance (1980) and Kawahara (2012) (shown by **) or from Kawahara (2012) alone (shown by *)

<table>
<thead>
<tr>
<th>No-violation</th>
<th>Local LL</th>
<th>Non Local LL</th>
<th>No-violation</th>
<th>Local LL</th>
<th>Non Local LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>/t/</td>
<td></td>
<td></td>
<td>/s/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tamuma**</td>
<td>taguta**</td>
<td>tatsuga**</td>
<td>semaro**</td>
<td>sebato**</td>
<td>sekabo**</td>
</tr>
<tr>
<td>tatsuka**</td>
<td>tozumi*</td>
<td>tenago*</td>
<td>sekato**</td>
<td>segeha*</td>
<td>soyoga*</td>
</tr>
<tr>
<td>taruna*</td>
<td>tegura*</td>
<td>tomiba*</td>
<td>sutane*</td>
<td>sobumo</td>
<td>sukaza*</td>
</tr>
<tr>
<td>tonime</td>
<td>tadanu (IA)</td>
<td>takuda (IA)</td>
<td>samohe</td>
<td>sazanu (IA)</td>
<td>suhozu (IA)</td>
</tr>
<tr>
<td>tekeha</td>
<td>tedesa (IA)</td>
<td>tokodo (IA)</td>
<td>sorise</td>
<td>sozoka (IA)</td>
<td>setaze (IA)</td>
</tr>
<tr>
<td>tokeho</td>
<td>todohu (IA)</td>
<td>temide (IA)</td>
<td>sateme</td>
<td>suzuhu (IA)</td>
<td>sokezo (IA)</td>
</tr>
<tr>
<td>/k/</td>
<td></td>
<td></td>
<td>/h/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kimane**</td>
<td>kidake**</td>
<td>kitage**</td>
<td>honara**</td>
<td>hobasa**</td>
<td>hokada**</td>
</tr>
<tr>
<td>kikake**</td>
<td>kobono*</td>
<td>koriga*</td>
<td>hinumi*</td>
<td>haboke*</td>
<td>hekazu*</td>
</tr>
<tr>
<td>katoni*</td>
<td>kabomo*</td>
<td>kamagi*</td>
<td>honiko</td>
<td>hogore*</td>
<td>hemiga*</td>
</tr>
<tr>
<td>kumisa</td>
<td>kegere (IA)</td>
<td>kurgi (IA)</td>
<td>hakisa</td>
<td>hibiro (IA)</td>
<td>haheba (IA)</td>
</tr>
<tr>
<td>konihe</td>
<td>kuguha (IA)</td>
<td>kesoge (IA)</td>
<td>heraho</td>
<td>hubumo (IA)</td>
<td>hisubi (IA)</td>
</tr>
<tr>
<td>keharo</td>
<td>kogoko (IA)</td>
<td>kanuga (IA)</td>
<td>hihonu</td>
<td>hebeno (IA)</td>
<td>huhobu (IA)</td>
</tr>
</tbody>
</table>

2.2 Task

The current experiment used a two-way forced-choice wug-test (Berko, 1958). Within each trial, the participants were given /nise/ ‘fake’ as a first element (E1) and a nonce noun as a second element (E2), and were provided with a compound form with rendaku and one without rendaku. They were then asked to choose the better resulting compounding form. For example, they were asked: “given /nise/ ‘fake’ and /tamuma/ (nonce), what would be the better outcome, /nisetamuma/ or /nisedtamuma/? Please choose the one that sounds more natural to you.” They were told to treat N2s as nonce words, to minimize the effects of analogy with existing words.7

---

6Our previous experiments showed that using nonce nouns for both E1 and E2 can impose too much psycholinguistic burden on native speakers during wug-tests about rendaku. A real word was used for E1 for this reason.

7Kawahara (2012) did not find a substantial difference between when the participants were told to treat the stimuli as nonce words and when they were told to treat them as old native words, as was done in Vance (1980).
2.3 Procedure and participants

The experiment was run online using SurveyMonkey. During the initial instructions, the participants were first told what rendaku is with actual Japanese examples. The participants then went through three practice questions with /nise/ as E1 and three real lexical items as E2, in order to familiarize themselves with rendaku and the task of the current experiment.

In the main session, all the target stimuli were presented. Although all the stimuli were presented in Japanese orthography, the participants were asked to choose an option that sounds more natural. The order of the stimuli was randomized per participant. After the main part of the experiment, they were asked if they were familiar with Lyman’s Law: we excluded those who explicitly knew about Lyman’s Law from the analysis, because there are a few participants who have taken introductory linguistic courses, and we needed to exclude potential bias that these participants may have. As a result, 38 native speakers of Japanese completed the study. The participants voluntarily participated in the experiment, and there were no monetary compensations.

2.4 Statistics

Since the response was binary (yes-rendaku or no-rendaku), logistic linear-mixed model analyses were run to analyze the results (Jaeger, 2008). Subjects and items (both E1 and E2) were encoded as random factors. Both slopes and intercepts of random effects were included in the model to have the maximal random structure (Barr, 2013; Barr et al., 2013).

---

As an anonymous reviewer pointed out, the use of orthography may be of some concern, as “participants may disfavor orthographic forms with two identical letters”. We fully acknowledge that this possibility is an interesting issue, and indeed have entertained this idea elsewhere—Lyman’s Law (and Identity Avoidance) may be orthotactics rather than phonotactics (Fukazawa et al., 2013). Testing this possibility requires a set of new experiments, however, which is beyond the scope of the current investigation. For a comparison between orthography stimuli and audio stimuli in a phonological judgment experiment using Lyman’s Law-driven devoicing patterns in Japanese loanwords, see Kawahara (2013).
Figure 1: The proportions of rendaku application for each condition. Loc=Local. LL=Lyman’s Law. IA=Identity Avoidance. The error bars represent binomial 95% confidence intervals. Rendaku was most likely when it does not violate Lyman’s Law (the leftmost bar). Rendaku is least likely when it violates local Lyman’s Law and Identity Avoidance (the middle bar).

3 Result

Figure 1 shows the proportions of rendaku application for each condition, with error bars representing binomial 95% confidence intervals. The first bar is a case which does not violate Lyman’s Law. The second bar violates local Lyman’s Law but not Identity Avoidance. The third bar is a case which violates both local Lyman’s Law and Identity Avoidance. The fourth bar violates non-local Lyman’s Law, but not Identity Avoidance. The fifth bar violates non-local Lyman’s Law and Identity Avoidance. See Table 1.

The results first of all show that Lyman’s Law is active in blocking rendaku: the difference between the leftmost bar (no Lyman’s Law violation) and the other four bars (Lyman’s Law violations) is statistically significant (0.58 vs. 0.39, 0.27, 0.36, 0.37: $z = 5.032, p < .001$). This result
replicates the previous studies in that Lyman’s Law is psychologically active in Japanese speakers’ minds (Ihara et al., 2009; Kawahara, 2012; Vance, 1979).

The comparison between the second and the third bar shows that the simultaneous violation of Identity Avoidance and Lyman’s Law decreases the probability of rendaku, as compared to the mere violation of Lyman’s Law, when the violation is local (0.39 vs. 0.27: \( z = 4.77, p < .001 \)). The comparison between the fourth and fifth condition shows that the effect of Identity Avoidance is not tangible when the violation is not-local (0.36 vs. 0.37: \( z = 0.35, n.s. \)).

4 Discussion

4.1 Identity Avoidance and Lyman’s Law

Our results first of all show that the uniformity assumption about Lyman’s Law—that all Lyman’s Law violating items are equally unacceptable—is false. Those items that violate both Lyman’s Law and local Identity Avoidance are considered to be worse than those that merely violate Lyman’s Law. To the best of our knowledge, this is a new finding about rendaku and Lyman’s Law. The results overall show that the prohibition against adjacent identical structures (Identity Avoidance), which is known to be active in several other languages (e.g. Alderete & Frisch 2007; Berent 2013; Frisch 2004; Frisch et al. 2004; Leben 1973; McCarthy 1979, 1986; Suzuki 1998; Yip 1988, 1998), is also active in Japanese in blocking rendaku, further supporting the recent experimental findings by Kawahara & Sano (2014) (also claimed by Sato 1988 and Labrune 2012).

An anonymous reviewer raised an interesting issue about how to interpret this core result, raising two possible interpretations: (i) Lyman’s Law and IA are additive, or (ii) Lyman’s Law and IA interact in some way so that IA enhances Lyman’s Law. The current experiment was not designed to tease apart these possibilities, but we suspect that the null hypothesis would be the first possibility, especially given that the effect of Identity Avoidance on rendaku is already established in other contexts (Kawahara & Sano, 2014). This issue however needs to be resolved empirically with further experimentation.
4.2 Locality of dissimilatory forces

The fact that Identity Avoidance exerts its effect only in a local fashion is interesting, and compatible with a cross-linguistic observation about similarity avoidance: languages disfavor similar structures in proximity, and its effect is stronger when the two structures are closer (Frisch, 2004; Frisch et al., 2004; Odden, 1994; Pulleyblank, 2002; Suzuki, 1998).

On the other hand, the comparison between the second and the fourth conditions in Figure 1 shows that, controlling for the effect of Identity Avoidance, there are no obvious differences between the local violation of Lyman’s Law and the non-local violation of Lyman’s Law (0.39 vs. 0.36: \(z = 0.39, n.s.\)). This result contrasts with those of Vance (1980) and Ihara et al. (2009) who have found the locality effect of Lyman’s Law, but is compatible with that of Kawahara (2012) who has not. It may be the case, as conjectured by Kawahara (2012), that young speakers have lost the difference between local Lyman’s Law and non-local Lyman’s Law, a trend which was already observed in the experiment by Ihara et al. (2009). In short, Identity Avoidance is sensitive to locality, whereas Lyman’s Law is not, at least in contemporary Japanese.

One general theoretical implication that follows from the current study is therefore that Identity Avoidance is active only between two adjacent moras, whereas Lyman’s Law does not show such a locality condition. This difference supports the idea that different dissimilation forces can have different domains, i.e. the domain of dissimilation can be parameterized (Odden, 1994; Pulleyblank, 2002; Suzuki, 1998), even within a single language.

4.3 The role of grammar beyond the lexicon

The second theoretical implication of the current results is the role of grammatical constraints that go beyond lexical patterns. In the Japanese lexicon, there are only a handful of exceptions that violate Lyman’s Law: /hun-zibaru/ ‘tie tightly’, /X-zaburoo/ (personal names), and /nawa-basigo/ ‘rope ladder’ (Kindaichi, 1976; Otsu, 1980; Vance, to appear).\(^9\) None of these actual

\(^9\)There are obsolete items that are not in use in Modern Japanese which violate Lyman’s Law: /waka-ziraga/ ‘young white hair’ and /rei-degami/ ‘thank you letter’. We also note that Lyman’s Law also functions as a phonotactic constraint in the Japanese native lexicon, and therefore those lexical items that violate Lyman’s Law are rare at best in
examples violate local Identity Avoidance, and therefore it is impossible to learn the effect of Identity Avoidance from the existing data about rendaku in the current Japanese lexicon. We thus conclude that this finding instantiates a case in which experiments reveal a preferential hierarchy between two sets of structures, which cannot be learned from a lexical pattern (see e.g. Berent et al. 2007; Kager & Pater 2012; Moreton 2002). Japanese speakers know that violating local Identity Avoidance and Lyman’s Law at the same time is worse—or more marked—than merely violating Lyman’s Law, and this knowledge cannot be learned from the existing lexical pattern about rendaku.

4.4 Final remark: Experimentation in phonology

Finally, we would like to conclude with a general remark about experimentation in linguistic research, as a third theoretical contribution of this study. Rendaku has been extensively studied in the past literature (see the list in footnote 2). Nevertheless, this effect of Identity Avoidance on Lyman’s Law-violating items has not been known. We thus conclude that a linguistic experiment is a useful methodology that complements the traditional approach to phonology based on dictionaries and introspection, in that it can reveal aspects of phonological knowledge that are difficult to access otherwise.

References


