EXAMINING LEXICAL AND PHONOLOGICAL FACTORS ON RENDAKU IN SPONTANEOUS SPEECH*

SHIN-ICHIRO SANO  
Okayama Prefectural University

1 Introduction

By studying the Corpus of Spontaneous Japanese, this paper offers new findings on the patterns of rendaku in actual usage. In particular, the study confirms the effect of place of articulation on rendaku application (Ohala 1983; Hayes 1999; Hayes and Steriade 2004), and Rosen’s Rule (Rosen 2001, 2003). However, as for the effect of lexical strata (Itô and Mester 1986), and Strong Lyman’s Law (Unger 1975; Vance 2005), the study reveals some previously unknown aspects of rendaku patterns.

1.1 Rendaku

Rendaku (aka. sequential voicing), one of the well-known and well-studied morphophonological processes in Japanese refers to the voicing of initial voiceless obstruents of the second member of morphologically derived words (most likely compounds) (Vance 1979, 1980). As example (1) shows, when the two nouns /hosi/ ‘star’ and /sora/ ‘sky’ are morphologically concatenated, the initial consonant /s/ of the second noun becomes voiced, producing /hosizora/ ‘starry sky.’

(1) hos + sora ‘star + sky’  =>  hosizora ‘starry sky’

* I would like to thank Shigeto Kawahara, Kevin Heffernan, the members of the Keio phonetics-phonology study group and the participants of FAJL7 for their invaluable comments. I am also grateful to the members of ICU linguistics and NINJAL for their help organizing the conference. This project is supported by JSPS KAKENHI Grant #26770147, #25770157, and #26284059. Any remaining faults are, of course, mine.

1 The transcription system adopted in this paper is the standard phonemic transcription of Japanese, rather than the narrow phonetic transcription, as phonetic details are outside the scope of this paper.
Not all compounds, however, undergo this voicing process; that is, rendaku shows variability. Prior work has identified various kinds of lexical, phonological and morphological factors that affect the likelihood of rendaku application. Some representative factors are shown in (2) through (4).

(2) Lyman’s Law (OCP(voice), Itô and Mester 1986)
Voiced obstruent(s) in the second member blocks rendaku.
 e.g. hoshi + kuzu ‘star dust’ => hosikuzu (*hosiguzu)

(3) Right branch condition (*left), (Otsu 1980)
Nouns on the right branch undergo rendaku.
 e.g. nise+tanuki+siru ‘fake raccoon soup’ => [nise [tanuki jiru] ‘raccoon soup that is fake’
=> [[nise danuki] jiru] ‘soup of fake raccoon’

(4) Lexical strata (Itô and Mester 1986)
Rendaku applies only to Yamato words, not to Sino-Japanese words, and Foreign words.
 e.g. gengo + kenkyuu ‘language study’ => gengokenkyuu (*gengokenkyuu)
 waarudo + kappu ‘world cup’ => waarudokappu (*waarudogappu)

The most famous and presumably the most influential factor is Lyman’s Law (also known as OCP(voice)) in (2), where rendaku is blocked when the second member of a compound already contains a voiced obstruent. /hosi/ and /kuzu/, for example, do not form */hosiguzu/ and the voicing of /k/ remains intact, because the second noun contains a voiced obstruent /z/, and this /z/ functions as a blockage of rendaku.

The right branch condition in (3) refers to the structure of compounding, and it specifies that only (the second) members on the right branch can undergo rendaku. Thus, even though both of the two examples consist of exactly the same members /nise/ ‘fake,’ /tanuki/ ‘raccoon,’ and /jiru/ ‘soup,’ only the compound with the structure [[nise danuki] jiru], where the second member is on the right branch, can undergo rendaku.

The final example of rendaku-affecting factors is lexical strata in (4). The Japanese lexicon consists of four kinds of strata: Yamato (native) words, Sino-Japanese words (old borrowings from Chinese, henceforth SJ words), foreign words (recent borrowings mostly from English), and onomatopoeia or sound symbolic/mimetic words (Itô and Mester, 1986). Among these strata, rendaku applies only to Yamato words, and not to words on other strata. Therefore, SJ words, such as /gengo+kenkyuu/ and foreign words, such as /waarudo+kappu/ do not undergo rendaku. As these factors and others come into play, the distribution of rendaku is not uniform.

1.2 Problems and Goals

The previous studies on rendaku were conducted from various perspectives, including the formal analysis such as Optimality-Theoretic approaches (Itô and Mester 1996, 2003), and experimental approaches (Vance 1979 et seq.; Ihara et al. 2009, 2011; Kawahara 2012; Kawahara and Sano 2014). In particular, building upon traditional (i.e., intuition-based and dictionary-based) studies,
the recent experimental approaches have conducted experiments using naturalness judgment, wug-test, nonce words or novel compound formation, and confirmed the psychological reality of various factors that have been claimed to affect rendaku application. These studies, either traditional or recent approaches, have made substantial contributions to phonological theory and to many theoretical debates. However, the aspects of actual rendaku production, especially in spontaneous speech, have yet to be explored.

With this background, the goals of this study are summarized as follows: (i) focusing on the production in spontaneous utterances, to address the patterns of rendaku (non)-application in actual speech, (ii) to examine the following factors: lexical strata, place of articulation, Rosen’s Rule that refers to the length of the first members of compounds, Lyman’s Law that is about the presence/absence of voiced obstruents in the second members, and Strong Lyman’s Law (henceforth SLL) that refers to the presence/absence of voiced obstruents in the first members. Specifically, the examinations of Rosen’s Rule and of SLL are motivated by the fact that many works have focused on the effect of the second members (i.e., the rendaku undergoers), but not on the effect of the first members. Accordingly, by testing the effects of Rosen’s Rule and of SLL, this study sheds light on the (potential) roles of the first members.

2 Method

This study employs a Japanese speech corpus as a source of the data. This section presents how the data were retrieved from the corpus.

2.1 Corpus

The corpus employed is the CSJ-RDB, (the Corpus of Spontaneous Japanese – Relational Database, Kokuritsu Kokugo Kenkyuujo 2012). The CSJ-RDB is a part of the CSJ, called Core with rich annotation, and consists of 201 speech samples, amounting to 45 hours of speech. The main feature of the CSJ-RDB is that various kinds of annotations are linked together. For example, phonetic/phonological information and morphological information is at different linguistic levels, and is stored in separate files. However, because these files are linked together, different kinds of information are simultaneously accessible in the data retrieval. This feature makes it possible to conduct the detailed data retrieval; namely, we can specify the target from various aspects, and can get more data without also extracting unwanted portions. For example, the program can handle such detailed instruction as “Retrieve every initial (accent-bearing) /ga/ of the second member in two consecutive nouns.” This feature is especially important for the study of rendaku, because in order to retrieve rendaku data by batch processing, we need to specify at least the following information: segment (voiced/voiceless obstruents), position in each word (the initial position), and position in each noun-noun sequence (the second noun).3 The segmental information is in the phonetic transcription files, but the positional and the part-of-speech information is in the morphological transcription files. The CSJ-RDB provides a useful way to link these kinds of information in different files that existing corpora did not provide. As is commonly faced with in the study of phonological variation, this kind of issue

---

3 This study focuses only on noun compounds (see Section 2.2).
regarding the data retrieval is a part of the reasons why rendaku in spontaneous speech is understudied.

2.2 Data Retrieval

The target items were retrieved from the CSJ-RDB in the following manner: (i) all sub-corpora of the CSJ-RDB were examined; (ii) Navicat Premium (http://www.navicat.jp/premium/) that implements the programming language SQL was used; (iii) phonetic/phonological and morphological information was employed; (iv) the target items are compound nouns consisting of more than two nouns. To examine the effect of preceding and following contexts, each of the items is the trigram that includes potential rendaku undergoers (the second noun of compounds) with the following structure.

(5) [compound 1st Noun + 2nd Noun] + following context
e.g. hosoi + sora + da ‘star + sky + copula’

The first nouns correspond to the preceding contexts, and the second nouns to rendaku undergoers.

Figure 1. Search formula

4 Compared to compound nouns, compound verbs are rarely observed, as they are subject to various restrictions (Kindaichi 1976), and hence this study did not focus on compound verbs.
5 If the trigrams are compounds consisting of three nouns, they would be subject to the right branch condition. Those items that were subject to this condition and did not undergo rendaku were removed from the dataset.
The instruction is given to the program by using the search formula with regular expressions presented in Figure 1. Subsequently, the retrieved data were subjected to data screening, as described in the next section.

2.3 Data Screening

The CSJ-RDB and the SQL produced the data that is close to our instruction. However, some tokens that are outside our scope of investigation are also included in the dataset. Therefore, the retrieved data was subjected to data screening, and tokens that did not match the purpose of this study were excluded from the dataset: (i) I focus only on the tokens, where the initial consonants of the second nouns (i.e., rendaku undergoers) are one of the following obstruents \([h, t, k, s, b, d, g, z]\), otherwise the tokens were excluded from the dataset. If rendaku does not apply, and consonants are faithfully realized, then voiceless obstruents will result; on the other hand, if rendaku applies, then voiced obstruents will result; (ii) I excluded tokens where the initial consonants of the second nouns are underlingly voiced, because these examples cannot undergo rendaku; (iii) I excluded tokens with a word boundary between the first and the second nouns, because these are not compounds; (iv) I excluded tokens where rendaku is categorically blocked due to lexical restrictions as shown in (6) through (11).

(6) Numerals e.g. sanjuu hachi ‘thirty-eight’

(7) Proper nouns e.g. hayasiya kompee \((a\ personal\ name)\)
    nakano simbasi \((the\ name\ of\ a\ train\ station)\)

(8) Nominalizing suffix -sa e.g. ooki sa ‘largeness’

(9) Diminutives and titles e.g. baa chan ‘grandma,’ Kindaichi sensei ‘Prof. Kindaichi’

(10) Nouns preceded by the honorific prefix -o e.g. o kasi ‘snack’

(11) Nouns preceded by native readings \((kunyomi)\) of numerals e.g. hito heya ‘one-room’
    huta toori ‘two ways’

In numerals such as example (6) /sanjuu hachi/, and in proper nouns such as example (7) /hayasiya kompee/, and /nakano simbasi/, the second nouns does not undergo rendaku. The nominalizing suffix -sa (8) does not undergo rendaku, as in /ooki sa/. Additionally, Diminutives and titles (9) did not undergo rendaku, as in /baa chan/, and /Kindaichi sensei/. Similar lexical restrictions are also imposed by the first nouns. Nouns preceded by the honorific prefix -o - (10) did not undergo rendaku, as in /o kasi/, supporting the observation in Sato (1989). Also, nouns preceded by native readings of numerals (11) did not undergo rendaku, as in /hito heya/, and /huta toori/. This supports Nakagawa’s (1966) observation. All these tokens were excluded from the dataset.
3 Dataset

After screening the tokens found in the CSJ-RDB, 1,193 tokens remained. Each token was then classified for the application/non-application of rendaku. This primary classification constitutes the basis of the following analysis, and every probability of rendaku is calculated based on this variable. The overall distribution is summarized in Table 1.

Table 1. The distribution of (non-)rendaku in the CSJ-RDB

<table>
<thead>
<tr>
<th></th>
<th>frequency</th>
<th>ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>+rendaku</td>
<td>709</td>
<td>0.59</td>
</tr>
<tr>
<td>-rendaku (faithful)</td>
<td>484</td>
<td>0.41</td>
</tr>
<tr>
<td>total</td>
<td>1,193</td>
<td>1</td>
</tr>
</tbody>
</table>

As shown in Table 1, rendaku applies about 60% of the time in spontaneous speech style represented in this corpus. Each of the tokens is subjected to the analysis with respect to the following factors: lexical strata, place of articulation, Rosen’s Rule, Lyman’s Law, and SLL.

4 Analysis

This section illustrates the analyses of the effects of the factors that have been claimed to affect rendaku applicability. Each of the analyses is presented with the result and the discussion.

4.1 Lexical Strata

Firstly, I will show the examination of the effect of lexical strata. As mentioned above, it is widely recognized that the Japanese lexicon is divided into the following strata: Yamato (native) words, SJ words that are old borrowings from Chinese, foreign words, and onomatopoeia (the sound-symbolic or mimetic words). Among these strata, the application of rendaku is generally believed to be limited to Yamato words (Itô and Mester 1986). Thus, the question to be addressed here is whether the effect of lexical strata is active in spontaneous speech, more specifically, rendaku is indeed limited to Yamato words or is observed also in other strata. The result of the analysis is shown in Figure 2.

Although no foreign loanwords nor any onomatopoeia with rendaku were observed in the current dataset, as Figure 2 shows, more than 44% of SJ words underwent rendaku. The result shows that the lexical strata affects the likelihood of rendaku application. However, the distribution of rendaku is slightly different from the previous assumption, that is, in spontaneous speech, SJ words are more likely to undergo rendaku than previously supposed, though not to the extent of Yamato words. Furthermore, the result implies the reorganization of the Japanese

---

6 SJ words that underwent rendaku are limited to small number of items, such as kaisya ‘company,’ tokei ‘watch, sho ‘institute,’ and hon ‘book.’ This leads to the question that the result is nothing more than a reflection of
lexicon. Given that SJ words can undergo rendaku that is characteristic of Yamato words, we can argue that the nativization of SJ words is developing.

Figure 2. The effect of lexical strata ($\chi^2=54.49$, df=1, $p < .01$)

4.2 Place of Articulation

Universally, voiced stops with backer place of articulation are more marked as represented by the following hierarchy: *g >> *d >> *b (Ohala 1983; Hayes 1999; Hayes and Steriade 2004). This is due to the physiological, aerodynamic reason, that is, the voicing induced by the glottal vibration is difficult to be maintained in backer place of articulation. The nonce word experiment (Ihara et al. 2011) shows that the markedness hierarchy manifests itself in the pattern of rendaku, that is, the more back the place of articulation, the lower the likelihood of rendaku. Thus, the question to be addressed here is whether rendaku follows the universal markedness hierarchy in spontaneous speech.

Figure 3. The effect of place of articulation ($\chi^2=51.55$, df=2, $p < .01$)

As Figure 3 shows, from bilabial to velar, the probability of rendaku becomes lower in backer place of articulation. The result shows that the place of articulation affects the likelihood of rendaku.

---

exceptional behaviors of these items. However, even if the effect is examined by type frequency, the similar result is obtained (rendaku likelihood: Yamato words: 82%, SJ words: 71.8%).
of rendaku application in spontaneous speech. Specifically, the applicability of rendaku follows the universal markedness hierarchy, that is, the backer the place, the less rendaku is applied. In addition, the result replicates the nonce word experiment. This suggests that the native speakers’ judgment is consistent with their production.

4.3 Rosen’s Rule

As mentioned above, Rosen’s Rule (Rosen 2001, 2003) refers to the length of the first nouns. In short, the rule states that when the first noun is three moras and longer, rendaku is more likely to apply, while when the first noun is two moras and shorter, rendaku is less likely to apply. That is, there is a gap between two-mora first nouns and three-mora first nouns, and longer first nouns promote the likelihood of rendaku application. Thus, the question to be addressed here is whether Rosen’s Rule is active in spontaneous speech. If the rule is active, the probability of rendaku should be higher in longer first nouns.

Figure 4 shows that the probability of rendaku is higher when the first noun is three, four, or five moras than when it is one, or two moras. There is a significant difference between the first nouns with two moras and shorter and those with three moras and longer. The result is consistent with what Rosen’s Rule states. This suggests that Rosen’s Rule affects the likelihood of rendaku application in spontaneous speech, in support of the psychological reality of Rosen’s Rule.

4.4 Lyman’s Law

As described in Section 1.1, Lyman’s Law (Itô and Mester 1986) refers to the presence/absence of voiced obstruents in the second noun. Specifically, rendaku is blocked when the second noun of a compound already contains a voiced obstruent. This is regarded as an OCP effect where multiple voiced obstruents are not permitted in a single stem. Lyman’s Law in rendaku has been found to be active in several experimental studies (Vance 1979; Ihara et al. 2009; Kawahara 2012). Based on these empirical findings, I addressed the question whether Lyman’s Law is active in spontaneous speech. In the current dataset, however, no examples were attested where there are voiced obstruents other than the rendaku undergoers in the second nouns. In other
words, Lyman’s Law categorically blocks rendaku. This suggests that Lyman’s Law has a stronger impact in spontaneous speech than in experimental settings.

4.5 Strong Lyman’s Law (SLL)

SLL refers to the presence/absence of voiced obstruents in the first noun. Specifically, rendaku is blocked when the first noun of a compound contains a voiced obstruent. For example, among the personal names /shima+də/ undergoes rendaku, because the first noun does not involve voiced obstruents, while /shiba+tə does not undergo rendaku, because the first noun involves a voiced obstruent (Sugito 1965; Zamma 2005). SLL is named after its effect; namely, a voiced obstruent in the first noun has an effect on the 2nd noun across the morphological boundary, unlike Lyman’s Law that exerts its effect within a single morpheme. This law has been shown to hold in Old Japanese (Unger 1975; Vance 2005), but its synchronic status has been much debated (Sugito 1965; Itô and Mester 1996, 2003; Vance to appear). Furthermore, there are many lexical exceptions, such as /kage+bəsɨ/ ‘dry in shade.’ Thus, the question to be addressed here is if SLL is active in spontaneous speech.

![Figure 5. The effect of SLL (χ²=19.54, df=1, p < .01)](image)

As shown in Figure 5, the probability of rendaku is higher when a voiced obstruent is present in the first noun than when it is absent. If SLL is active, voiced obstruent in the first nouns should block rendaku, and the probability of rendaku should be lower in “presence” than in “absence” The “presence,” however, promoted the rendaku applicability. The tendency does not change, even if we restrict our attention to Yamato words (presence: 75%, absence 63%). This is the pattern opposite of what was previously claimed. Furthermore, Lymans’ Law monitors the presence/absence of voiced obstruents within the second nouns, and it was active; on the other hand, SLL that has an effect over the entire compound was not active in a way previously supposed. This suggests that the effect of OCP(voice) is stem-bound.

4.5.1 SLL – Locality

Next, let us take a closer look at the effect of SLL. Firstly, I examine the effect of SLL in terms of the locality, that is, the distance between voiced obstruents (blocker) and rendaku undergoers
(target). For Lyman’s Law, the locality effect was shown to be active (Vance 1979; Ihara 2009). Thus, the question to be addressed here is whether the locality effect of SLL is active in spontaneous speech. To specify the position of the blocker (voiced obstruent), I focused only on tokens where a single voiced obstruent is present in the first nouns, and the distance is based on the number of moras intervening between the blocker and the target.

In Figure 6, the probability of rendaku is higher when the distance is two or three moras than when it is one mora or adjacent. The result shows that rendaku is less likely to apply when the distance between the blocker and the target is one mora and closer than when it is two moras and farther. In other words, the closer the blocker, the stronger the blocking effect is. Thus, we can argue that the locality effect of SLL affects the rendaku likelihood in spontaneous speech.

4.5.2 SLL – Number of Blockers

Finally, I will present the examination of the effect of SLL in terms of the number of blockers. We can hypothesize that if there are more blockers, the blocking effect would be stronger. Thus, the question to be addressed is whether SLL is sensitive to the number of blockers in spontaneous speech. I focus only on tokens where a voiced obstruent is present in the first nouns.

![Figure 6. The effect of SLL and locality ($\chi^2=17.36$, df=3, $p < .01$)](image)

![Figure 7. The effect of SLL and number of blockers ($\chi^2=4.24$, df=1, $p < .05$)](image)
As Figure 7 shows, the probability of rendaku is higher when there is one blocker, than when there are two blockers. The result shows that rendaku is less likely to apply when there are two blockers in the first nouns than when there is one blocker. In other words, the more voiced obstruents there are, the stronger the blocking effect is. Thus, we can argue that SLL is sensitive to the number of blockers, and it affects the rendaku likelihood in spontaneous speech. To summarize the results of the examinations of SLL, the effect of SLL itself was not confirmed, rather it promoted the rendaku applicability. However, if we focus on the distribution of blockers, the effects of the locality and of the number were confirmed. Therefore, as far as the locality and the number are concerned, we can argue that the cross-linguistically common dissimilating force is working in rendaku that is confirmed in the previous literature (Cross-linguistic pattern: Itô and Mester 2003; Frisch 2004; Frisch et al. 2004; Ihara et al. 2009 (locality), Tesar 2007; Kawahara 2011 (number), Devoicing of voiced geminates in Japanese: Sano 2013; Kawahara and Sano 2013 (corpus-based study)).

5 Conclusion

This corpus-based study presented new findings on the patterns of rendaku, especially in actual production. In particular, this study (i) confirmed some factors that have been claimed to affect rendaku applicability, such as place markedness, Rosen’s Rule, and Lyman’s Law, (ii) revealed some previously unknown aspects of rendaku patterns, such as lexical strata, and SLL. Furthermore, the effects of place markedness and of the locality and the number in SLL were shown to follow the cross-linguistically common pattern.

References


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


