Whiskey or Bhiskey?: Influence of First Element and Dialect Region on Sequential Voicing of shoochuu

Katsuo Tamaoka  
Nagoya University

Fumiko Ikeda  
Yamaguchi Prefectural University

Abstract: The present study investigated how five first elements (i.e., imo, soba, kome, mugi and kokuto) affect sequential voicing (rendaku) in the second element shoochuu across six dialect regions (i.e., Kagoshima, Oita, Fukuoka, Yamaguchi, Hiroshima and Shizuoka). A decision tree analysis of questionnaire data obtained from 405 participants was conducted to predict voiced (/z/) or voiceless (/s/) decisions based on the two variables of shoochuu ingredient and dialect region. Results indicated that the type of shoochuu ingredient as the first-element was a significant factor for voiced-or-voiceless decisions, dividing them into four groups. (1) 'Imo+shoochuu' showed the highest frequency of voicing at 93.83%. (2) 'Kome+shoochuu' (88.89%) and 'Soba+shoochuu' (84.69%) showed similar percentages. (3) 'Mugi+shoochuu' (72.59%) was significantly lower than imo, kome or soba. (4) 'Kokuto+shoochuu' (56.44%) was the lowest. The present study also demonstrated that the six dialect regions appeared to have no influence on voiced-or-voiceless decisions. Dialect seems to have no influence on rendaku occurrences, at least in the case of shoochuu.*

Keywords: sequential voicing, rendaku, dialect region, voiced-or-voiceless decision

1. Introduction

Shoochuu (焼酎) is a traditional Japanese alcoholic drink which is usually made from one of various ingredients such as barley, rice, buckwheat or sweet potato in much the same way as whiskey is usually distilled from either rye or wheat. The word shoochuu is combined with the name of the particular main ingredient to

*The authors would like to express their sincere gratitude to the former Editor-in-Chief, Taro Kageyama, and the present Editor-in-Chief, Haruo Kubozono, and two anonymous reviewers for their insightful and constructive comments to complete the present paper. The authors also greatly appreciate the contributions of Professor Nobuko Kibe of Kagoshima University for collecting data in Kagoshima Prefecture, Associate Professor Mika Matsuda of Beppu University for collecting data in Oita Prefecture, Associate Professor Yayoi Miyaoka of Hiroshima University of Economics for collecting data in Hiroshima Prefecture, and Professor Yasushi Terao of the University of Shizuoka for collecting data in Shizuoka Prefecture.
produce a compound word. As a result, the pronunciation of the initial consonant of /s/ in *shoochuu* frequently becomes voiced /z/. This is called *rendaku*, or sequential voicing, which refers to the voicing of the initial voiceless obstruent of a second element in the case of compounds and prefix-plus-base combinations (Vance, 1987).

An in-flight video program shown on All Nippon Airways (ANA) flights included an advertisement for 'ANA Shoochuu Kurabu [Shoochuu Club]'\(^1\) (ANA, 2007). Four kinds of *shoochuu* were introduced in this program: (1) *imo-joochuu* /imo-zyoRtyuR/ (‘R’ represents vowel length), (2) *soba-joochuu* /soba-zyoRtyuR/, (3) *mugi-shoochuu* /mugi-zyoRtyuR/, and (4) *kome-shoochuu* /kome-zyoRtyuR/, referring to spirits distilled from sweet potato, buckwheat, barley and rice, respectively. Whereas the initial consonant of the second-element in the names of the first two drinks was voiced as /z/, illustrating *rendaku*, this initial consonant in the latter two kinds of *shoochuu* was unvoiced, remaining unchanged.

As observed in this example, *rendaku* does not occur consistently. Individual preferences seem to influence *rendaku* occurrences. For example, *mugi-shoochuu* can be pronounced as either voiceless *mugi-shoochuu* /mugi-zyoRtyuR/ or voiced *mugi-joochuu* /mugi-zyoRtyuR/. Likewise, native Japanese speakers may pronounce the combination of *kome* ‘rice’ and *shoochuu* as either *kome-shoochuu* /kome-zyoRtyuR/ or *kome-joochuu* /kome-zyoRtyuR/. Thus, sequential voicing, or *rendaku*, fluctuates even for the same word among native Japanese speakers. The purpose of the present paper, then, is to predict how voiced (/z/) or voiceless (/s/) decisions are affected by the two variables of ingredient name and dialect. Specifically, we investigated how five first elements (i.e., *imo*, *soba*, *kome*, *mugi* and *kokuto*) affect *rendaku* in the second element *shoochuu* across six dialect regions (i.e., Kagoshima, Oita, Fukuoka, Yamaguchi, Hiroshima and Shizuoka).

2. Assumptions and Questionnaire Studies on Rendaku
Each of the five first elements (i.e., *imo*, *soba*, *kome*, *mugi* and *kokuto*) has a different set of compound combinations. For instance, the element *imo* can be compounded with *kayu* ‘porridge’ and pronounced as *imogayu* with a voiced consonant, meaning ‘potato porridge’. Because each of these five first elements produces a different number of compound words, the possibility of *rendaku* cannot be held constant across all five first elements when considering all possible compounds together. Therefore, in order to directly compare the five first elements with each other, the present study investigated voicing judgments by native Japanese speakers by giving each participant a choice of either voiced (*rendaku*) or unvoiced (non-*rendaku*) for all the five first elements combined with the same second element (i.e., *shoochuu*).

*Rendaku* fundamentally occurs among native Japanese lexical items, called *Wago*, but not very frequently among non-native lexical items (e.g., Itô and Mester,

\(^1\) This program was shown as an ANA in-flight video program in May, 2007. *Auwamori*, which is also classified as a type of *shoochuu* in Okinawa, was also introduced in the in-flight program, but *auwamori* is excluded due to the purpose of the present *rendaku* study.
1995, 2003, 2008; Ohno, 2001; Tanaka, 2009; Vance, 1987, 1996). However, Sino-Japanese words, referred to as Kango, exhibit rendaku in some compounds. For instance, the Kango compound word 株式会社 (‘joint-stock company’), which combines /kabusiki/ (株式) and /kaisya/ (会社), is pronounced /kabusiki-gaisya/ with the initial consonant /k/ of the second-element becoming voiced /g/. Likewise, with the Kango compound word 夫婦喧嘩 (‘conjugal quarrel’), the combination of /huRu/ and /kenka/ is voiced as /huRu-genka/. A typical explanation for rendaku occurrences in Kango is that non-native lexical items undergo rendaku once they come to be used frequently and become very familiar to native Japanese speakers (e.g., Otsu, 1980; Irwin, 2005; Itô and Mester, 2003; Ohno, 2001; Takayama, 1999).

However, as Ohno (2001) pointed out, the definition of ‘familiarity’ is quite ambiguous. Ohno (2001) provides some examples of frequently-used familiar Kango which are never voiced, such as kyaku ‘a guest’ in the compound *romarikyaku* ‘a guest who stays overnight’, and *seesseki* ‘result’ as in *kooseseski* ‘a good result’. In contrast, some less-frequently-used Kango undergo rendaku. Examples include *bookoo* ‘service’ and *ooroo* ‘lantern’ in the compounds *dettibookoo* ‘apprenticeship’ and *isidooroo* ‘stone lantern’. As seen in these examples, when ‘familiarity’ is defined as frequent use, the occurrence of rendaku cannot be clearly explained. Taking this contrastive tendency among Kango items into account, Itô and Mester (2008) took a new approach by dividing Sino-Japanese (Kango) into two subcategories depending on whether or not they show compound voicing.

To investigate the ‘familiarity’ factor, the present study included samples from different dialect regions. The Kagoshima region is known to have over 100 *shoochuu* breweries, all of which use *imo* (‘sweet potato’) in production (Nihon Shuzoo Kenkyuukai, 2004). Among these, *Satsuma Shiranami* is perhaps the most well-known ‘sweet potato’ *shoochuu* brand. Extending the idea of familiarity, labeled as ‘Japanized’ by Otsu (1980) or as ‘native look-alikes’ by Itô and Mester (2003), it is assumed that for *shoochuu*, being Kango, if frequency of use for an *imo*-and-*shoochuu* compound word affects rendaku, people in Kagoshima will show higher rendaku frequency (pronouncing it as /imo-zyoRtyuR/) than people from other regions due to their familiarity with the product. Likewise, the Ōita region has two famous brands of 100% pure *mugi* (‘barley’) *shoochuu*, called *Nikaidoo* and *Iibiko* (Nihon Shuzoo Kenkyuukai, 2004). Again, it is expected that people in Ōita will probably pronounce the compound with *mugi* with a voiced sound, i.e., as /mugi-zyoRtyuR/, more frequently than people from other regions. Thus, as one aspect of individual differences, dialect may be a good candidate to for investigating the determination of rendaku by lexical ‘familiarity’.

The questionnaire-based approach to sequential voicing, or rendaku, employed in the present study was initiated by Vance (1979). Vance asked 14 native Japanese speakers with the Tokyo *Yamanote* accent to determine whether or not 645 compound words exhibit rendaku. Murata (1984), and Ihara and Murata (2006) also employed a similar approach, but expanded the sample size and employed carefully selected and elaborated test items involving specific aspects of rendaku.
Murata (1984) tested voiced-or-voiceless decisions using two pseudo-Japanese-origin (pseudo-Wago) words as the second elements (i.e., basuri and hukari) and found a significant influence of first elements. After an interval of 21 years, Ihara and Murata (2006, Experiment 1) repeated similar voiced-or-voiceless questions as part of their study. The studies of both Murata (1984) and Ihara and Murata (2006) indicated the importance of the first element in determining the voicing of the second element. However, since the study by Murata and Ihara utilized nonwords for testing rendaku, it was necessary to re-examine the question of using existing (i.e., real) compound words.

Likewise, Ohno (2001) conducted a questionnaire study on existing compound elements using rendaku decisions, and he suggested that native Japanese speakers are likely to refer to existing compounds and semantically parallel forms to determine their voiced-or-voiceless decisions. He did not, however, conduct a systematic rendaku survey in order to support his argument by statistical analysis for the data he collected. As described in this paper, since compounds of shoobuu referring to various ingredients can have both voiced and voiceless forms, occurrences of rendaku in these compounds cannot be explained by a single factor of estimation from existing lexical items as Ohno (2001) proposed. Therefore, the present study investigated voiced-or-voiceless decisions for shoobuu compound words in order to clarify rendaku frequencies from the perspectives of within compounds (whether sequential voicing of shoobuu occurs randomly in each compound), between compounds (whether sequential voicing of shoobuu differs among compounds), and two other potential factors: dialect regions and Internet frequency.

3. Participants and Procedure
A total of 405 undergraduate students from the six different dialect regions participated in the study: 61 participants from Kagoshima, 69 participants from Oita, 60 participants from Fukuoka, 60 participants from Yamaguchi, 63 participants from Hiroshima and 92 participants from Shizuoka. All these native Japanese speakers were born and raised in one of the six dialect regions. Three of the regions are on Kyushu, two regions are in the Chugoku Area, and the Shizuoka region is located relatively far from the other five regions. All participants were asked to choose between voiced and voiceless options, as indicated in the following examples.

- 芋焼酎 (いも じょうちゅう) (/imo-zyoRtyuR/; voiced)
- 芋焼酎 (いも しょうちゅう) (/imo-syoRtyuR/; voiceless)

Participants were required to check either the voiced or voiceless choice written in kanji with their respective pronunciations. Five different types of shoobuu were selected, with the first elements of imo, soba, kome, mugi and kokuto, to investigate frequencies of rendaku occurrence.

4. Rendaku Frequencies for Five First Elements and Six Dialect Regions
Frequencies for the voiced-or-voiceless choices of the initial consonant of shoobuu by participants are reported in Table 1 as a function of the six dialect regions.
Voiced/voiceless frequencies were examined by a series of Chi-square tests of goodness-of-fit, with the expected value set at equal frequency (50%, i.e., random chance).

As shown in the results of the Chi-square tests, of goodness-of-fit, the 61 participants in Kagoshima exhibited a significantly high rendaku frequency of 91.80% for imo-shoochuu [$\chi^2(1)=42.639$, $p<.001$]. Similarly, kome-shoochuu also had a rendaku frequency higher than 90% [$\chi^2(1)=39.361$, $p<.001$]. Both soba-shoochuu (85.25%) and mugi-shoochuu (75.41%) were lower in rendaku frequency than imo- and kome-shoochuu, but nevertheless showed significantly higher occurrences than the random chance frequency of 50% [$\chi^2(1)=30.311$, $p<.001$ for soba-shoochuu; $\chi^2(1)=15.754$, $p<.001$ for mugi-shoochuu]. Unlike the other shoochuu types, the responses for kokutoo-shoochuu were at the random chance level of rendaku frequency [$\chi^2(1)=0.148$, n.s.]. Furthermore, a Chi-square test of independence was conducted to compare rendaku frequencies for the five types of shoochuu. The result was significant, suggesting that the voiced/voiceless frequency patterns differ among the five types of shoochuu [$\chi^2(4)=47.167$, $p<.001$].

The same trend of rendaku frequencies observed among participants in Kagoshima was also seen in the cases of the 69 participants in Oita, the 60 participants in Fukuoka and the 63 participants in Hiroshima (see the results of Chi-square tests of goodness-of-fit in Table 1). All four of these dialect regions shared the same ranking of rendaku ratios of occurrence in the descending order of imo-, kome-, soba-, mugi- and kokutoo-shoochuu. In addition, the results of Chi-square tests of independence indicated significant among the four regions of Kagoshima, Oita, Fukuoka and Hiroshima (see the results of Chi-square tests of independence in Table 1).

However, the voiced/voiceless frequency patterns for the two dialect regions of Yamaguchi and Shizuoka showed significance for kokutoo-shoochuu [$\chi^2(1)=8.067$, $p<.05$ for Yamaguchi; $\chi^2(1)=4.348$, $p<.05$ for Shizuoka], suggesting that kokutoo-shoochuu was pronounced with voiced /a/ more frequently than mere random chance in the dialect regions of Yamaguchi and Shizuoka. The other types of shoochuu exhibited the same trend across all regions. The overall influence of dialect regions is examined in the following section.

5. Decision Tree Analysis for Type of Shoochuu and Dialect Region

A decision tree analysis using SPSS 15.0J, 2006, Classification Trees was conducted to predict the frequencies of voiced-or-voiceless decisions as determined by the two variables of dialect region and shoochuu type. Decision tree analysis attempts to select a useful subset of predictors from a larger set of variables. The technique automatically detects significant interaction effects among independent (predictor) variables, by repeating Chi-square tests at each step for categorical variables. It determines the pair of categories of each independent variable (predictor) that is least significantly different with respect to the dependent variable. In the tree-growing process, each parent node splits into child nodes only if some significant interaction is found among independent variables. Every step for
<table>
<thead>
<tr>
<th>Dialect region</th>
<th>Type of <em>shochohu</em> in kanji</th>
<th>First-element with <em>shochohu</em></th>
<th><em>Rendaku</em> frequency</th>
<th>Chi-square test of goodness-of-fit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Voiced</td>
<td>Voiceless</td>
<td>Ratio</td>
</tr>
<tr>
<td>Kagoshima</td>
<td>蕃焼酎</td>
<td>ino + syoRtyuR</td>
<td>56</td>
<td>5</td>
</tr>
<tr>
<td>(n=61)</td>
<td>蕃焼酎</td>
<td>kome + syoRtyuR</td>
<td>55</td>
<td>6</td>
</tr>
<tr>
<td>Oita</td>
<td>蕃焼酎</td>
<td>ino + syoRtyuR</td>
<td>56</td>
<td>13</td>
</tr>
<tr>
<td>(n=69)</td>
<td>蕃焼酎</td>
<td>kome + syoRtyuR</td>
<td>54</td>
<td>6</td>
</tr>
<tr>
<td>Fukuoka</td>
<td>蕃焼酎</td>
<td>ino + syoRtyuR</td>
<td>54</td>
<td>6</td>
</tr>
<tr>
<td>(n=60)</td>
<td>蕃焼酎</td>
<td>kome + syoRtyuR</td>
<td>54</td>
<td>6</td>
</tr>
<tr>
<td>Yamaguchi</td>
<td>蕃焼酎</td>
<td>ino + syoRtyuR</td>
<td>54</td>
<td>6</td>
</tr>
<tr>
<td>(n=60)</td>
<td>蕃焼酎</td>
<td>kome + syoRtyuR</td>
<td>54</td>
<td>6</td>
</tr>
<tr>
<td>Hiroshima</td>
<td>蕃焼酎</td>
<td>ino + syoRtyuR</td>
<td>54</td>
<td>6</td>
</tr>
<tr>
<td>(n=63)</td>
<td>蕃焼酎</td>
<td>kome + syoRtyuR</td>
<td>54</td>
<td>6</td>
</tr>
<tr>
<td>Shizuoka</td>
<td>蕃焼酎</td>
<td>ino + syoRtyuR</td>
<td>54</td>
<td>6</td>
</tr>
<tr>
<td>(n=92)</td>
<td>蕃焼酎</td>
<td>kome + syoRtyuR</td>
<td>54</td>
<td>6</td>
</tr>
</tbody>
</table>

Note: There was one missing value in the voiced and voiceless frequencies of 'kokuto+schochu' in Hiroshima.

splitting nodes uses Bonferroni's adjusted $p$ values to avoid Type I Error, i.e., the error of rejecting the null hypothesis when it is actually true. The results of decision tree analysis are a hierarchy drawn in a dendrogram: stronger predictors go to the higher nodes while weaker predictors appear at the ends of the branches. Non-significant predictors are not included in the dendrogram. Branches grow when significant interactions are found in the data.

The results of the analysis are depicted as a dendrogram in Figure 1. Dialect region (i.e., Kagoshima, Oita, Fukuoka, Yamaguchi, Hiroshima and Shizuoka)
was not a significant factor for voiced-or-voiceless decisions, so this factor was automatically excluded from the dendrogram. Although the Chi-square tests of goodness-of-fit reported in Table 1 showed a slight difference in the rendaku ratio of kokutoo-shoochu among the six dialect regions, this difference was negligible from an overall perspective. In contrast, the type of shoochu was a significant factor for voiced-or-voiceless decisions \( \chi^2(3)=219.503, p<.001 \). The overall rendaku ratio of 'Imo+shoochu' showed the highest frequency at 93.83%. 'Kome+shoochu' (88.89%) and 'Soba+shoochu' (84.69%) showed similar percentages and are thus classified in the same frequency category. The frequency of voiced-or-voiceless decisions for 'Mugi+shoochu' (72.59%), however, was significantly lower than imo, kome and soba. 'Kokutoo+shoochu' (56.44%) was the lowest.

6. Printed-Frequency in Internet Frequency Search and Rendaku Ratio

The present study used a lexical corpus of the Asabi Newspaper printed from 1985 to 1998, produced by Amano and Kondo (2000). This corpus contains morpheme frequencies of 41,771 types and 287,792,797 tokens. Shoochu in kanji (i.e., 焼酎) was found 562 times, while it appeared only 8 times in the compound form 芋焼酎. The hiragana form いもじょうちゅう (i.e., /imo-zyoRtuR/) appeared 3 times, suggesting all voiced occurrences. This number is insufficient for arguing whether shoochu is voiced or voiceless.

An Internet frequency search using the Google search engine was conducted on April 28, 2008, for the kanji, hiragana and katakana script forms of the compounds of shoochu examined in the present study. Frequencies of occurrences are reported in Table 2. In contrast to the newspaper corpus, kokutoo-shoochu showed the highest frequency at 2,050,000 times in kanji among the five shoochu compounds, whereas soba-shoochu was the lowest at 385,000 in kanji. Hiragana presentations of soba-shoochu showed a total frequency of 725,800, of which 675,000 were voiceless. There were 50,800 voiced incidences, so rendaku occurred in only 7.00% of the examples. A decision tree analysis was conducted to predict voiced and voiceless frequencies for the five different types of shoochu. In the hiragana script, frequencies of voiced-or-voiceless decisions were significantly predicted by shoochu compounds \( \chi^2(3)=223153.732, p<.001 \), indicating a descending rendaku ratio order of 'imo > mugi = kokutoo > kome > soba'. Likewise, in the katakana script, frequencies of voiced-or-voiceless decisions were also significantly predicted by shoochu compounds \( \chi^2(4)=1930.141, p<.001 \), indicating a descending rendaku ratio order of 'kome > mugi > soba > imo > kokutoo'.

It is readily apparent that the order of rendaku ratios as determined by Internet search does not match for the two kana scripts. Furthermore, the rendaku ratio order of neither the hiragana nor the katakana script was congruent with the descending order of 'imo > soba = kome > mugi > kokutoo' which was indicated by the human participants in the present study. The results of the Google frequency search for shoochu compounds seem to reflect neither actual frequencies of appearances nor rendaku ratio. Apart from the actual human propensity toward rendaku, the Internet results were doubtless strongly influenced by commercial popularity at the
Figure 1  Dendrogram of decision tree analysis for voiced-or-voiceless decisions depending on type of *shoochuu*

*Note:* Since dialect region (i.e., Kagoshima, Oita, Fukuoka, Yamaguchi, Hiroshima and Shizuoka) was not a significant factor for voiced-or-voiceless decisions on *shoochuu* compounds, this factor was automatically excluded from the present dendrogram.
Table 2  Frequencies of *shoohuu* in Internet search and *rendaku* ratios

<table>
<thead>
<tr>
<th>Type of <em>shoohuu</em></th>
<th><em>Shoohuu</em> in kanji</th>
<th>Frequency in kanji</th>
<th>Frequency in hiragana</th>
<th>Frequency in katakana</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Voiced</td>
<td>Voiceless</td>
<td>Ratio</td>
</tr>
<tr>
<td>imo + shoohuu</td>
<td>芋焼酎</td>
<td>1,950,000</td>
<td>23,100</td>
<td>638</td>
</tr>
<tr>
<td>kome + shoohuu</td>
<td>米焼酎</td>
<td>1,510,000</td>
<td>33,700</td>
<td>204,000</td>
</tr>
<tr>
<td>soba + shoohuu</td>
<td>蕎麦焼酎</td>
<td>385,000</td>
<td>50,800</td>
<td>675,000</td>
</tr>
<tr>
<td>mugi + shoohuu</td>
<td>麦焼酎</td>
<td>993,000</td>
<td>1,440</td>
<td>644</td>
</tr>
<tr>
<td>kokutoo + shoohuu</td>
<td>黒糖焼酎</td>
<td>2,050,000</td>
<td>6,440</td>
<td>3,070</td>
</tr>
</tbody>
</table>

| Total             | 6,888,000          | 1,950,000 | 23,100   | 638 | 97.31% | 3,050 | 2,740   | 52.68% |
|                   | 1,510,000          | 33,700   | 204,000 | 14.18% | 2,920 | 260   | 91.82% |
|                   | 385,000            | 50,800   | 675,000 | 7.00% | 1,300 | 570   | 69.52% |
|                   | 993,000            | 1,440 | 644 | 69.10% | 1,770 | 389   | 81.98% |
|                   | 2,050,000          | 6,440   | 3,070 | 67.72% | 380 | 531   | 41.71% |

Result of 'decision tree' analysis<sup>(2)</sup>  
$\chi^2 (3) = 223153.732, p < .001$  
$\chi^2 (4) = 1930.141, p < .001$

*Note 1:* Internet (Google) search was conducted on April 28, 2008.

*Note 2:* Decision tree analysis was conducted to predict voiced-or-voiceless decisions by type of *shoohuu*.

*Note 3:* *Rendaku* ratios of *shoohuu* compound words were compared using decision tree analysis.
time when the lexical search was conducted. It should also be noted that the Google search engine is not designed for finding usage frequencies, but for finding particular information. As seen in Table 2, the Google search figures for frequently-used expressions were rounded up to the nearest ten thousand (to 1,950,000 for kanji '芋焼酎', and to 675,000 for hiragana 'そばしようちゅう'), providing just a rough frequency estimation. That the resources of a Google search cannot be identified is just one crucial problem, and a Google search is not an ideal tool for frequency calculation of specific lexical or phrasal expressions.

7. General Discussion
Although various studies have investigated rendaku thoroughly (e.g., Haraguchi, 2001; Itō and Mester, 1986, 2003; McCawley, 1968; Otsu, 1980; Rosen, 2001; Vance, 1979, 1987, 2006), the effects of individual differences on rendaku have not received much attention. The questionnaire-based approach should be able to illuminate subtle differences in rendaku frequencies caused by individual preferences, especially when occurrences of rendaku are not quite certain. Thus, the present study utilized this approach with a large number residents \( n=405 \) who had lived their whole lives in the six different dialect regions (i.e., Kagoshima, Oita, Fukuoka, Yamaguchi, Hiroshima and Shizuoka), focusing only on voiced-or-voiceless decisions for shoobuu compounds with five different first elements (i.e., imo, soba, kome, mugi and kokutoo). The following sections discuss how the results were influenced by (1) the first elements imo, soba, kome, mugi and kokutoo, (2) the second element shoobuu, and (3) other factors of dialect regions and Internet frequency.

7.1 The effect of the first-elements—imo, soba, kome, mugi and kokutoo
The present study indicated that the first elements of shoobuu compound words indeed affect the voicing of the initial consonant of the second element. The results of the Chi-square tests of goodness-of-fit (see Table 1) indicated that all four different Wago first elements (excluding kokutoo) yielded higher voicing frequencies of the shoobuu element than the random chance (50 percent). However, the Sino-Japanese (Kango) first element of kokutoo showed just a chance level of voicing frequency, except for residents of Shizuoka. Furthermore, as depicted in the dendrogram (see Figure 1) produced by the decision tree analysis, the various patterns of shoobuu voicing showed that voicing ratios based on first elements differ, resulting in the descending order of 'imo > soba > kome > mugi > kokutoo'. Even among the four different Wago (excluding kokutoo), rendaku ratios varied from 93.83% for imo to 72.59% for mugi. The present study demonstrated that the four Japanese-origin (Wago) first elements have an influence on the likelihood of voicing in the second element.

Unlike the Wago cases of imo, kome, soba and mugi, the Sino-Japanese (Kango) first element kokutoo displayed a chance rendaku ratio in Kagoshima, Oita, Fukuoka and Hiroshima (see Table 1). These differences in rendaku frequencies were also confirmed by the decision tree analysis (see Figure 1). Although the data of the present study only dealt with the single case of kokutoo, as suggested by previous
studies (e.g., Itō and Mester, 2003, 2008; Ohno, 2001; Tanaka, 2009; Vance, 1996), the frequency data support the idea that rendaku is fundamentally observed among the core stratum of Wago. Regarding this finding, Tamaoka, Ihara, Murata and Lim (2009) also indicated that Wago first elements resulted in rendaku in second elements more frequently than Kango first elements. Furthermore, they also found that two-mora first elements led to rendaku in second elements more frequently than three/four-mora first elements did. Because the word kokutoo contains four morae and is classified as Kango, this word is ‘doubly negative’ in regard to both word length and etymological type, resulting in a lower rendaku ratio. As such, for both etymological and phonological reasons, the first element kokutoo was expected to result in less frequent rendaku than the four two-mora Wago cases (imo, kome, soba and mugi). In this sense, although the present study treated only the single case of shoobuu, the influence of the first element on rendaku supports the notion of ‘lexical strata’.

The Obligatory Contour Principle (OCP) is a constraint that prevents similar or identical phonological features from being repeated (e.g., Itō and Mester, 1986; Kubozono, 1999, 2005; Kubozono and Ota, 1998; Tamaoka, Makioka and Murata, 2004; Tamaoka and Murata, 2001). Lyman’s Law is a particular case of the OCP, blocking rendaku when the second element of a compound word contains a voice obstruent. Sugito (1965), in her article ‘Shibata-san to Imada-san [Shibata and Imada]’, proposed to extend Lyman’s Law to first elements ending with a mora containing a voice obstruent. For example, the family name Shibata, containing the two single-kanji morphemes shiba (‘firewood’) and ta (‘rice field’) does not have rendaku, while the family name Imada, containing the two morphemes of ima (‘now’) and ta, does have rendaku: /ima-da/. Since shiba already contains the voiced obstruent /b/ in its last mora, the initial consonant of the second element, /t/, remains unchanged. In fact, Ramsey and Unger (1972) suggested that the first or second element of a compound word with a voiced obstruent prohibited sequential voicing in the Japanese language used in the eighth century (the Nara period). This idea is called the ‘strong version’ of Lyman’s Law. The present study, however, found the same rendaku ratios for soba and kome (see Figure 1), both of which showed higher ratios than random chance (see Table 1). Since soba contains a voiced obstruent in its last mora, it is expected to block voicing of any following consonant. Despite this expectation, the second element shoobuu was not voiced more often after kome, which does not contain a voiced obstruent. Both soba and kome have the same CVCV structure, so this example casts doubt on Sugito’s generalization that a voiced obstruent consistently blocks rendaku across morpheme boundaries. Thus, the ‘strong version’ of Lyman’s Law is not supported by the compounds of shoobuu observed in this study.

One might argue that the significantly lower frequency of rendaku after mugi than after soba indicates the involvement of Lyman’s Law at the prosodic word level. However, the same argument is also applicable to the significantly lower frequency of rendaku after kome than after imo. This argument would be relevant only if the frequency of rendaku after both mugi and soba were significantly lower
than after kome or imo (ideally, ’mugi = soba < kome = imo’). Therefore, the argument of the difference of voiced frequencies between mugi and soba affected by Lyman’s Law in the first elements does not hold.

7.2 The effect of the second element—shoochuu
Some researchers (e.g., Itô and Mester, 1995, 2003; Tanaka, 2009; Vance, 1987, 1996) suggest an overall generalization that, with some exceptions, rendaku fundamentally occurs among native Japanese words (Wago), but not among Sino-Japanese words (Kango) or other loanwords (Gairaigo). In a wider perspective, this tendency is usually explained in the framework of etymological ‘lexical strata’.
As for etymological types, Japanese lexical items (i.e., morphemes) are classified into four categories, Wago, Kango, Gairaigo and onomatopoeia (Gitaigo/Giseigo). In linguistic studies, it is proposed that etymological types define ‘lexical strata’ in the Japanese lexicon, each having its own unique constraints. In this conceptual framework, there is a core of Wago lexical items surrounded by the two increasingly peripheral layers of Kango and Gairaigo, each defined by its own phonological constraints. Rendaku is originally observed among the core stratum of Wago.

Shoochuu (焼酎) is also a Kango word, involving the On-readings /syôR/ (焼) and /tûR/ (酎) of the two kanji. Since examples of rendaku are easily found among Kango (e.g., Irwin, 2005; Itô and Mester, 2003; Ohno, 2001; Takayama, 1999), rendaku generalizations based on etymological category simply provide an overall phonological likelihood for ‘lexical strata’ in modern Japanese. Establishing two subcategories of Kango depending on whether or not they undergo compound voicing (Itô and Mester, 2008) is one approach to describing the phonological nature of Kango lexical items.

7.3 The effect of other factors—dialect regions and internet frequency
The familiarity and usage-frequency of shoochuu is expected to affect rendaku ratios. As mentioned previously, the Kagoshima dialect region is well known for shoochuu made from sweet potatoes, while the Oita dialect region is famous for shoochuu made from barley (Nihon Shuzoo Kenkyuuai, 2004). Thus, the present study assumed that people in Kagoshima would display a higher rendaku ratio for shoochuu when compounded with imo than would people of other regions. Similarly, people in Oita could be expected to pronounce shoochuu voiced when compounded with mugi. Despite this expectation, dialect regions appeared to have no influence on voiced-or-voiceless decisions.
In addition, in relation to familiarity, results of a Google search showed no resemblance to the voiced-or-voiceless decisions by human participants observed in the present study. Google findings might reflect miscellaneous factors including market trends, and one particular occurrence of an item may be overemphasized due to multiple publications. Since a Google search has no transparent resources for frequencies, it is not recommended for frequency indexes.
8. Limitations and Further Studies
The present study shed light on voiced-or-voiceless choices and individual preferences. However, despite the large number of participants involved in the data collection (over 400 participants from six regions), the scope of the experiment was limited in having investigated only the single case of shochohu, representing the voicing pattern of /s/ to /z/. As Kubozono (1999) and Ohno (2001) explain, there are four patterns in all: /k/ to /g/, /s/ to /z/, /t/ to /d/ and /h/ to /b/ (labial weakening altering the sound of /p/ to /b/, resulting in a pattern of /h/ to /b/). All these rendaku patterns should be investigated using multiple examples in future studies.

References


---

**Author's contact information:**

Graduate School of Languages and Cultures
Nagoya University
Furocho, Chikusaku, Nagoya, Japan, 464–8601
e-mail: ktamaoka@lang.nagoya-u.ac.jp

[Received 2 December 2008; Accepted 1 August 2009]
【要 旨】
「しょうちゅう」か「じょうちゅうか」？
—焼酎の連濁に対する第一要素と方言の影響—

玉岡篤津雄
名古屋大学

池田史子
山口県立大学

本研究では、芋、穀麦、米、麦、黒糖の5種類の第1要素が第2要素である焼酎の連濁に対してどのように影響するかを鹿児島、大分、福岡、山口、広島、静岡の6つの方言地域で調査した。405名に対して質問紙調査を行い、5種類の第1要素と6種類の方言地域の2つの変数が連濁の有無（/s/または/z/）を予測するかどうかについて、決定木（decision tree）分析によって検討した。その結果、焼酎の連濁の予測には、第1要素の焼酎の種類が有意な影響力を持っていった。焼酎の連濁頻度は、決定木分析のアンドログラムに示したように（Fig.1を参照）、4つのグループに分かれた。まず、第1要素焼酎の連濁が93.83%で最も多かった。次に、第2要素焼酎が88.89%、穀麦焼酎が84.69%と同程度の連濁頻度であった。さらに、第3要素焼酎が72.59%とやや低かった。最後に、第4要素焼酎が56.44%であり、ランダムの50%に近い連濁頻度であった。一方、6つの方言地域は、焼酎の連濁頻度に影響しないことが分かった。少なくとも、焼酎の連濁については、地域差がないようである。