Acquisition of Yo-on (Japanese contracted sounds) in L1 and L2 phonology

拗音習得過程に見られる第一，第二言語の音韻構造の影響

Chiharu TSURUTANI

Griffith University

Abstract

This study observes the prosodic constraints that govern the pattern of simplification of Japanese contracted sounds in the acquisition processes of first (L1) and second language (L2) learners: Japanese children and English speaking learners of Japanese.

Japanese language has a group of palatalized consonants, known to scholars of Japanese language as the ‘contracted’ sounds, [CjV]. In accordance with the prosodic constraints of Japanese, which is mora-timed, a CjV mora has to be the same length as a CV mora. However, in addition to typical moraic timing errors, learners of Japanese whose L1 is English appear to treat Japanese contracted sounds initially as consonant + glide clusters, where there is an equivalent [Cj] cluster in English, or otherwise tend to insert an epenthetic vowel [CijV]. The phonological status of the palatal glide [j] is still controversial in both English and Japanese. The acquisition process of the Japanese “CjV” unit by English speaking learners could offer the clue to clarify this issue.

Given the similarity between palatalized consonants and consonant clusters in their phonetic status, the acquisition literature of consonant clusters is considered initially. In general, L1 learners at an early stage of acquisition employ a reduction strategy to cope with the production difficulties associated with the clusters. Sonority appears to be the causal factor referred to most frequently. If the sonority hypothesis governs the reduction pattern, syllable initial clusters reduce to whichever consonant in the cluster creates a maximal sonority rise, and syllable final clusters reduce to whichever consonant in the cluster creates a minimal sonority descent; e.g. Onsets “pl” → “p” Codas “lp” → “l”(Ohala,1999). The former would possibly be the case if the “CjV” unit in Japanese were treated as a consonant cluster. The results of this study show a contrastive reduction pattern between L1 and L2 learners. Due to the moraic timing constraint, Japanese children would not epenthesize vowels in producing contracted sounds, while L2 learners violated moraic timing by lengthening a vowel or inserting gemination as well as an epenthetic vowel. The reduced consonant also differed between the two groups, and the result shed light on the clarification of the phonological status of /Cj/ sequences in Japanese and English. The findings of the study also point to Japanese children’s early acquisition of palato-alveolars, which accounts for the prevalence of the recognition that children are good at producing the contracted sounds.
1 Introduction

Japanese language has a group of palatalized consonants, known to scholars of Japanese language as the ‘contracted’ sounds /CjV/. In Japanese, the basic form of a mora is a CV syllable that corresponds with one kana character. In accordance with the prosodic constraints of Japanese, which is mora-timed, a CjV mora has to be the same length as a CV mora. However, learners of Japanese who are native speakers of English appear to treat Japanese contracted sounds initially as consonant + glide clusters, where there is an equivalent /Cj/ cluster in English, and they tend to violate the moraic timing frame by inserting an epenthetic vowel otherwise. For example, “CijV” is an error pattern commonly heard in second language classrooms. Since the phonological status of a /Cj/ cluster differs between Japanese and English, we expect that Japanese children will present a different error pattern from that of L2 learners of Japanese.

This study aims to observe the acquisition of Japanese palatalized consonants by first language (L1) and second language (L2) learners and the prosodic constraints that govern the pattern of simplification errors in this process. First, the phonetic and phonological status of contracted sounds and consonant clusters will be discussed. Following that, two experiments for L1 and L2 learners are presented separately.
2 Japanese contracted sounds

Japanese language contrasts palatalized and plain consonants at every place of articulation and for all manners except glides. Some contracted sounds are depicted with a palato-alveolar consonant without using diacritics to indicate palatalization. However, they all require palatalization gestures - raising the tongue toward the hard palate when certain consonants are pronounced - and form a phonological contrast with their plain counterpart.

Table 1. Japanese Plain and Palatalized Consonants

<table>
<thead>
<tr>
<th>Labial</th>
<th>Coronal</th>
<th>Dorsal</th>
</tr>
</thead>
<tbody>
<tr>
<td>plain</td>
<td>palatalized</td>
<td>plain</td>
</tr>
<tr>
<td>p</td>
<td>pʲ</td>
<td>t~ts</td>
</tr>
<tr>
<td>b</td>
<td>bʲ</td>
<td>d</td>
</tr>
<tr>
<td>m</td>
<td>mʲ</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>s</td>
<td>ŋ</td>
</tr>
<tr>
<td>z~dz</td>
<td>ʒ</td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>rʲ</td>
<td></td>
</tr>
</tbody>
</table>

*Glides [w] and [j] are not included in this table.

Japanese contracted sounds are orthographically expressed by two kana, one kana Ci and a half-sized kana yV, as illustrated below, but they are pronounced in one-mora length. In Japanese phonology, the basic form of a mora, a CV syllable, corresponds with one kana character. The fact that two kana characters are contracted into one mora length is the origin of the term “contracted sound”.

The glide “j” functions to palatalize the preceding consonant, and the palatalization produces a CjV unit of the same length as a CV mora in accordance with the prosodic constraint of Japanese. It is generally accepted that /Cj/ in Japanese is a palatalized consonant (Vance, 1982; Ito and Mester, 1995) and the existence of /j/ is not regarded as a consonant cluster in Japanese, although some researchers, such as Broselow and Finer (1991) and Matthews (2002), define a CjV syllable in Japanese as a consonant cluster. Matthews claims that consonant and glide sequences constitute genuine onset clusters based on the evidence of vowel / glide alternations induced by morphological affixation.

The formation of hyperpolite adjectives requires the suffix “u” to a stem with a high front vowel as shown below.

---
Matthews interprets that the stem final segment [i] is syllabified as a syllable onset as a result of vowel / glide alternation. Therefore, its mora becomes free to associate to the suffixal melody, resulting in lengthening of the suffix vowel /u/ to [u:]. Matthews further argues that the lengthening of the suffix /u/ would not occur if the high front vowel [i] were syllabified as part of the nucleus.

This is a phonologically plausible explanation. However, by having the segment [i] as part of a syllable onset, [k] is clearly palatalized. The alternation process does not exclude the possibility that /Cj/ is a palatalized consonant. In addition, if /Cj/ is a consonant cluster, it violates sonority distance constraint, which disallows consonantal sequences that do not differ by two points on the sonority scale1 (Selkirk, 1984; Clements, 1990) (e.g. /kj/ is acceptable, but /rj/ is not). Matthews explains this point from his view that Japanese does not impose as severe a minimum sonority distance as English does. However, since Japanese does not have any other consonant clusters, as its predominantly CV oriented syllable structure shows, it is difficult to validate Matthews’s argument.

Although Japanese has five vowels, palatalized consonants are associated only with three vowels [a], [u] and [o], as is the case for a Japanese singleton onset glide. This is recognized as a linguistically natural phenomenon. The contrast between a j-vowel sequence and the same vowel alone, based on the movement of the second vowel formant, creates the following ranking order:

\[
\text{ji-i} > \text{je-e} > \text{ja-a} > \text{jo-o} > \text{ju-u} \quad \text{(Padgett, 2001)}
\]

The contrast between jV and V becomes more distinct as the backness of the vowel increases, which corresponds with markedness prediction. Thus, there should not be languages that allow the “ji-i” contrast while ruling out the other four contrasts. English is one language that has all five contrasts, such as “east” [iːst] and “yeast” [jɪːst], or “ale” [eːl] and “Yale” [jeːl], whereas the contrast is neutralized before front vowels, [i] and [e] in Japanese. The restriction on vowels that follow a palatal glide in the Japanese /Cj/ cluster is due to a historical change (Vance, 1987) and, thus does not attest to the co-moraic nature of the palatal glide in the /CjV/ sequence. We have set out the two hypotheses regarding the /Cj/ sequences. Clearly the palatal glide in the Japanese /Cj/ units belongs to a subconstituent of the consonant. Whether the /Cj/ unit is a palatalized consonant or an onset cluster will be determined later, when considering the outcome of the experiments.

3 The nature of the “Cj” unit

The phonological status of /j/ in the “Cj” unit is controversial in English as well as in Japanese. The “consonant + palatal glide” structure in English has phonologically been

---

1 Here I use sonority scale as discussed by Clements (1990): obstruent < nasal < liquid < glide < vowel.
regarded as a consonant cluster (Gimson, 1994). However, while the glide /j/ can be followed by any vowel, consonant clusters “Cj” allows only [u] to follow (Giegerich, 1992). Besides, the preceding consonant position of /j/ has no restriction and even sonorants can occur. This violates English phonotactics that disallows two sonorant consonants in the onset. Based on this restricted distribution of the palatal glide, Davis and Hammond (1995) proposed treating the glide /j/ in CjV clusters as co-moraic with the following vowel. Anderson (1986) similarly held the view that [ju:] could be considered as a rising diphthong /iu/, as /j/ in CjV combinations only occurs before [u:].

Despite having taken into account the association of /j/ with the following vowel, it is hard to deny the association of /j/ with the preceding consonant, since /j/ palatalizes the adjacent consonant. Differences in the treatment of the pre-nuclear palatal glide are evident in speakers’ judgment as well as in linguistic analysis. In their study mentioned above, Davis and Hammond found some dialectal discrepancy in word games among native speakers about whether /j/ should be treated as part of onset or as part of rhyme. Using the same language game, Pig Latin, Barlow (2001) also reported speaker variation in treating ambiguous consonant sequences. These studies suggest the possibility that all speakers of a given language do not always share the same grammar (Yip, 2003). This issue appears to require further study.

The ambiguous status of /j/ is also reflected in phonetic transcription. For example, in order to transcribe “cute” phonetically, three versions are considered; [k'ut], [kjut] and [kiut]. The vowel [i] and the semi-vowel [j] involve an identical type of articulatory stricture, distinguished only by the fact that [i] has a noticeable duration, whereas [j] is a rapid glide (Catford, 1977). The most widely accepted transcription is [kjut]. However, in support of [kiut], Ladefoged (1982) states that the sequence /ju/ is more like a diphthong than a consonant followed by a vowel. Considering that diphthongs are called gliding vowels (Gimson, 1994) and that /ju/ is a rapid glide from [i] to [u], all three forms mentioned above are more or less accurate transcriptions.

The ambiguity of /j/ discussed above suggests the possibility of observing the typical error pattern of consonant clusters in the production of contracted sounds. To my
knowledge, no study has reported on the acquisition of Japanese contracted sounds in light of the phonological status of the /Cj/ cluster. The study on which I report below contrasts L1 with L2 acquisition of the “Cj” unit. These acquisition data would be able to shed light on the phonological ambiguity of /Cj/ sequences in both Japanese and English. In the following section the acquisition literature of the “Cj” unit will be reviewed.

4 Acquisition of the “Cj” unit

The phonetic transcription of the “Cj” unit in English is identical in consonant clusters and palatalized consonants, as discussed in the previous section. Therefore, the acquisition process can be the same in these two phonologically different sounds.

The acquisition process of consonant clusters in first (L1) and second language phonology has been widely discussed. In general, L1 learners at an early stage of language acquisition employ a strategy of reducing the consonant cluster to cope with the production difficulties associated with the clusters (Ingram, 1989), e.g. stop → top. Apart from reduction, null-onsets (total omission of contracted cluster), coalescence (fusion) and vowel epenthesis are also known to be learners’ strategies (Chin & Dinnsen, 1992). Vowel epenthesis is a particularly common strategy used by L2 learners whose L1 has a dominant CV syllable structure, such as Japanese and Vietnamese (Sato, 1984), e.g. stop → *sutopu*.

How L1 learners (children) simplify consonant clusters has been investigated from various perspectives. Ingram (1989) suggests that the more marked element is omitted, and the lesser marked element remains (deletion of the marked element). Ohala (1999) proposed that children’s consonant cluster reduction is driven by sonority. She asked 16 English children aged from 1.1 to 3.2 years to repeat 28 monosyllabic nonsense words CCVC or CVCC after the examiner and observed their consonant cluster reduction at word initial and final position. The results supported her hypothesis: children reduce syllable initial clusters to whichever consonant in the cluster creates a maximal sonority rise and they reduce syllable final clusters to whichever consonant in the cluster creates a minimal sonority descent.

\[ \text{e.g. Onsets “plV” → “pV” (maximal sonority rise)} \]
\[ \text{Codas “Vlp” → “VI” (minimal sonority descent)} \]

However, when children were confronted with clusters that were impossible in native clusters (e.g. fw-, tm-, -fp), this was not the case. Children clearly did not understand clusters with no English counterpart and often interpreted these clusters as having two syllables. Hence Ohala’s study observed more incorrect responses and vowel epenthesis. Her study provides evidence that English learners whose L1 does not have a dominant CV structure do make vowel epenthesis errors. The study also adds new insight into the occurrence of vowel epenthesis in L2 learners’ acquisition.

L1 learners’ acquisition of palatalized consonants is usually later than their acquisition of other plain consonants in most languages except Slavic languages that have many palatalized consonants (Timm, 1977; Locke, 1983). Japanese is not an exception in this
respect, since the frequency of occurrence of palatalized consonants is relatively low compared with Slavic languages. The following is the developmental sequence for Japanese, quoted from a table in Yasuda (1970 in Locke, 1983) that summarizes two studies from the mid-1960s which remain informative today.

Table 2. Developmental Schedule for Japanese Phonology according to two studies cited in Yasuda

<table>
<thead>
<tr>
<th>Age*</th>
<th>Initial or medial word position</th>
<th>Both positions</th>
<th>Initial or medial word position</th>
<th>Both positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before 3:5</td>
<td>m b t fj d3 \h F c n t \d k k kj gj \j w N pj bj \ç mj j</td>
<td>m b w fj d3 F c n t d k g kj n N bj pj çj mj</td>
<td>m b w j n N n t d k g fj d3 (mj) (bj) (Pj) h F ç (çj) kj gj</td>
<td>m b w j n N n t d k g fj d3 kj gj</td>
</tr>
<tr>
<td>3:6-3:11</td>
<td>r fj</td>
<td>qj</td>
<td>fj r (qj)</td>
<td>h F</td>
</tr>
<tr>
<td>4:0-4:5</td>
<td>dz ts</td>
<td>fj r h</td>
<td>After 4:0, s ts dz</td>
<td>s ts dz</td>
</tr>
<tr>
<td>4:6-5:11</td>
<td>s</td>
<td>dz ts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5:0-5:5</td>
<td>S r</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5:6-5:11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Age given is earliest age at which 75% of all subjects correctly articulated the sounds listed.  
**Parentheses indicate that only the word-initial position was tested.  

Among all palatalized consonants, /rj/ is listed as one of the late-acquired sounds. The early acquisition of [tʃ] and late acquisition of [s] is also characteristic of Japanese. The data in Table 2 resemble the cross-linguistically recognized order of acquisition of consonants, that is, “labials and alveolars first” (Jakobson, 1968; Locke, 1983; Ingram, 1989; Vihman, 1992; Bernhardt and Stemberger, 1998). It is thus assumed that the acquisition order in one language is determined by the development of children’s articulatory system and the frequency of occurrence of the sound. Based on the counting of mora in nouns and adjectives from the NTT telecommunications database (Yoneyama, personal communication, 2002), the following frequency order of palatalized consonants was reported. Numbers in parentheses indicate the number of times these consonants appeared in the data.

[s] (20,753) >>> [ʃ] (10,254) > [tʃ] (6,187) > [kj] (2,309) > [rj] (1,261) > [çj] (646)

Although the numbers are much smaller, the data collected by Maekawa (1993) on the frequency of mora from the text of two conversations and one narrative present the same order. In Japanese the frequency of [tʃ] is not high, while the frequency of [s] is high. Even though children’s active vocabulary is slightly different from adults’ lexicon, the acquisition order does not seem to correspond with the frequency of occurrence in Japanese. It may be useful to explore further how the early acquisition of [tʃ] in children
relates to their articulatory capacity.

5 Experiment 1

Having discussed the acquisition process of the /Cj/ unit in both English and Japanese, let us now turn to the experiment I have conducted to help determine the phonological status of the /Cj/ unit, by observing how English speakers treat unfamiliar /Cj/ sequences. This experiment examined the types of errors made by L2 learners on contracted sounds. Typically L2 learners (English speaking learners of Japanese) violate the moraic timing constraint on a CV syllable when it occurs at the penultimate position. They try to put stress on the CV syllable by lengthening a vowel or by inserting gemination after the vowel before the following consonant so that the stressed syllable forms a heavy syllable (Tsurutani, 2000). Although English does not possess geminate consonants within a word, learners’ conscious efforts to avoid lengthening the stressed short vowel often induce a glottal stop and result in gemination on the following consonant. It is suspected that the difficulty L2 learners have with contracted sounds arises from the unfamiliar phonetic environment where a consonant in the /Cj/ cluster is /r/, or where the /Cj/ cluster occurs before the vowels [a] and [o], since English does not have these phonetic combinations. The prosodic interference around contracted sounds needs to be examined, together with the vowels [a] and [o].

5.1 Hypotheses
Apart from the common timing errors, vowel lengthening (CjVV), and insertion of gemination (CjVC), I predicted that I would find the following two types of errors for L2 learners attempting to cope with unfamiliar sounds.

1) Vowel epenthesis (CijV)
2) Consonant cluster reduction (CV or jV)

If the learners treat /j/ as part of an onset, the reduction pattern of consonant clusters will follow the sonority-driven hypothesis and create a maximal sonority rise (e.g. CjV → CV), since the Cj unit occurs only in the syllable initial position in Japanese.

5.2 Subjects
The recording was conducted in April 2003 at Griffith University in Brisbane, Australia where the researcher is a lecturer in Japanese language. The subjects of this experiment were 17 beginner students of Japanese language (all females), enrolled in the first year course at Griffith University. They had studied Japanese for several months at high school, but their proficiency was still at beginners’ level. All were native speakers of English. At the time of conducting this experiment, the students had received 30 hours of classroom instruction at university.

A short perception test was conducted along with this experiment. The subjects were

--- 34 ---

2 Only a liquid-glide cluster, [jiu] in “lewd”, is tolerated in pronouncing approximants in British English and in some speakers of Australian English.
asked to choose either CV or CjV written in Kana after listening to one member of a minimal pair, such as "kyaku" or "kaku". The result of the test confirmed that all the participants could accurately distinguish CV from CjV syllables, although three could not perceive the difference between "ryuu" and "ruu". The misproductions made by these subjects thus can be considered as attempts involving correctly stored input that is incorrectly realized through their first language phonology in production.

5.3 Materials
It was attempted to select stimuli from words familiar to beginners, taking into consideration the following criterion:

The vowel in the target sound is short.
In order to observe the violation of English phonotactic constraints, the syllable structure has to be "Cj + V" instead of "Cj +VV" except in "kyu", "gyu" and "ryu" which have no light syllables in Japanese.

Since the vocabulary of subjects was limited, most of words with /rjV/ and /gjV/ were new to subjects. Before the recording, the researcher went through the words with subjects and taught them the words they did not know.

![Table 3. Stimulus words used for Experiment 1](image)

<table>
<thead>
<tr>
<th>kya</th>
<th>kyu</th>
<th>kyo</th>
</tr>
</thead>
<tbody>
<tr>
<td>okyakusan</td>
<td>kyu</td>
<td>yuubinkyoku</td>
</tr>
<tr>
<td>(guest)</td>
<td>(nine)</td>
<td>(post office)</td>
</tr>
<tr>
<td>gya</td>
<td>gyu</td>
<td>gyo</td>
</tr>
<tr>
<td>gyaku</td>
<td>gyuunyuu</td>
<td>kingyo</td>
</tr>
<tr>
<td>(opposite)</td>
<td>(milk)</td>
<td>(gold fish)</td>
</tr>
<tr>
<td>rya</td>
<td>ryu</td>
<td>ryo</td>
</tr>
<tr>
<td>ryaku</td>
<td>ryuugaku</td>
<td>ryokoo</td>
</tr>
<tr>
<td>(abbreviation)</td>
<td>(studying abroad)</td>
<td>(trip)</td>
</tr>
</tbody>
</table>

5.4 Test format
Picture cards without writing were used to extract the stimulus words from subjects, who were asked to say the word three times. When they forgot the word, the researcher whispered it. Their pronunciation was recorded and transcribed by the researcher.

5.5 Evaluation
It is difficult to judge L2 learners’ pronunciation as their acoustic target could fall between two phonemes. This is particularly so with beginners whose production is not consistent. The recorded data were first analyzed aurally by three native listeners who have taught Japanese to English learners for five to ten years. When the two listeners
agreed, this was accepted as a correct form. All three judges agreed on the transcription of most items (80 percent). The acoustic analysis was used to clarify ambiguous sounds when the native listeners produced three different transcriptions, which happened only on two items.

5.6 Findings
In total, 459 tokens were collected. The results are presented in Figure 1 below.

![Pattern of L2 Learners' Mistakes](image)

**Table 4. Actual Number of Errors**

<table>
<thead>
<tr>
<th>Correct form</th>
<th>Error types</th>
<th>CjVV</th>
<th>CjVQ</th>
<th>CVCV</th>
<th>CV</th>
<th>jV</th>
</tr>
</thead>
<tbody>
<tr>
<td>kya</td>
<td>42</td>
<td>6</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kyu</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kyo</td>
<td>38</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rya</td>
<td>31</td>
<td>2</td>
<td>10</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ryu</td>
<td>12</td>
<td>15</td>
<td>5</td>
<td>3</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>ryo</td>
<td>12</td>
<td>46</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gya</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gyo</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>323</td>
<td>27</td>
<td>19</td>
<td>15</td>
<td>61</td>
<td>14</td>
</tr>
</tbody>
</table>

Errors were mostly observed in “ryV” in the form of consonant cluster reduction. Some subjects substituted the approximant [i] for the flap [r]. It appears that “kyV” and “gyV” were easier for adult learners as demonstrated by the low error rate. Since most subjects did not know the stimulus words that started with /gi/ and /ri/, they had to repeat these
after the researcher.\(^3\) As a result, vowel epenthesis errors that are common in L2 learners’ production of consonant clusters were fewer than expected. It is notable that most simplification errors were observed in the pattern of “CV” in “ryV”. “ryu” was assumed to be the subject of L1 transfer and to have a different error pattern from “rya” and “ryo”. In fact no simplification error of /ju/ was observed in /rju/. This satisfies the principle of sonority-driven reduction of consonant clusters, that is, a cluster is reduced to create maximal sonority rise at syllable initial position. Thus, the result indicates that subjects treated the /Cj/ unit as a consonant cluster. All the subjects produced “kju” and “gyu” correctly as English has an equivalent sound (e.g. cute, argue).

The English phonotactics appears in timing errors as well. The other frequent error patterns were the insertion of gemination and vowel lengthening. English speakers tend to produce a heavy syllable in a stressed syllable either by lengthening a vowel or inserting gemination. The insertion of gemination was observed in four contracted sounds:

\[
\begin{align*}
  & \text{kya in okyakusan} \\
  & \text{kyo in yuubinkyoku} \\
  & \text{rya in ryaku} \\
  & \text{ryo in ryokoo}
\end{align*}
\]

Except for “kya”, the syllable positions of the contracted sounds were all in the penultimate position. However, “kya” could have been treated as a penultimate syllable as a result of vowel devoicing between [k] and [s]; o-kyaku-san → o-kyak-san. This is the same condition as for timing errors occurring in the production of CV morae\(^4\). The insertion of gemination is considered to be a result of an effort by learners to avoid lengthening a vowel in a penultimate position, as well as an attempt to create a heavy syllable on a stressed syllable.

6 Experiment 2

It is generally believed that for Japanese children the secondary articulation is more difficult, and is acquired later, than plain consonants (See Table 2). On the other hand, we often hear comments by mothers that Japanese children’s speech i.e. “yooji-go” (kids’ talk) contains lots of contracted sounds. In this experiment, the wider range of contracted sounds were used as stimuli to observe Japanese monolingual children’s production

---

\(^3\) The reviewer pointed out a possible task effect by repetition. Although repetition generally increases accuracy, the subject made more errors on [gj] and [rj] than on [kj]. By the stage when learners can produce the words that have /rj/ sounds only by English cues, they would usually have mastered the pronunciation as well. I therefore judged that the data reflected the closest indication of the learners’ actual linguistic behavior that I could obtain under the circumstances and I decided to accept the data as valid.

\(^4\) It was found that timing errors commonly occur on the penultimate syllable with the HL pitch pattern, as the result of transfer of English prosody (Tsurutani, 2000). The penultimate position is the canonical location of stress in a left-headed metrical foot in English.
ability in contracted sounds, as well as how their error pattern differs from that of L2 learners.

6.1 Hypothesis
A number of researchers argued that children have early sensitivity to the prosodic structure of words (Demuth, 1996; Fee, 1992; Boysson-Bardies, Vihman, Roug-Hellichius, Durand, and Arao, 1992). Due to the moraic timing constraint, Japanese children will not epenthesize vowels in any contracted sounds.

If the palatalized consonants are simplified, different error patterns are predicted for the flap by the following two accounts:

\[
\begin{align*}
\text{rjV} & \rightarrow \text{rV} \\
& \rightarrow \text{jV}
\end{align*}
\]

The simplification error of a glide and vowel means that the pattern of reduction is not governed by sonority. In the case of consonants that are less marked and their sonority is lower than \([j]\), this distinction will not be shown.

6.2 Subjects
The recording was conducted in February 2002 in Tokyo and Okayama in Japan. Thirty seven monolingual Japanese children ranging in age from 2:3 to 3:9 (years: months), average age 2.8, were audio-recorded at their homes or day-care centers in unstructured play sessions, in interaction with both their mother or teacher and one investigator.

6.3 Materials
Target sounds were \([j], [t]\), \([kj]\), \([gj]\) and \([rj]\), which according to Maekawa (1993) and Yoneyama (personal communication, 2002) are the most frequent contracted sounds. Care was taken in selecting the stimuli from children’s vocabulary. When no highly familiar real words could be found in the picture books for children, made-up names for stuffed dolls were used. They are underlined in the table below.
Table 5. Target sounds and stimulus words

<table>
<thead>
<tr>
<th></th>
<th>kya [kja]</th>
<th>kyu [kju]</th>
<th>kyo [kjo]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>kyabetsu (cabbage)</td>
<td>kyuuri (cucumber)</td>
<td>kyorotyan</td>
</tr>
<tr>
<td></td>
<td>kyuukyusha (ambulance)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>rya [rj a]</th>
<th>ryu [rj u]</th>
<th>ryo [rj o]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ryapuu kun</td>
<td>ryyu kun</td>
<td>ryooma kun</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>gya [gia]</th>
<th>gyu [gju]</th>
<th>gyo [gjo]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>gyaasuke</td>
<td>gyunyu (milk)</td>
<td>gyorotyan</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>sya [fa]</th>
<th>syu [fu]</th>
<th>syo [fo]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>densya (train)</td>
<td>tissyu (tissue)</td>
<td>syoohoosya (fire engine)</td>
</tr>
<tr>
<td></td>
<td>syanpuu (shampoo)</td>
<td>syuukuriimu (cream puff)</td>
<td>syooyu (soy sauce)</td>
</tr>
<tr>
<td></td>
<td>syatsu (shirt)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>tya [tfa]</th>
<th>tyu [tfu]</th>
<th>tyo [tfo]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tyairo (brown)</td>
<td>tyuurippu (turip)</td>
<td>tyootyo (butterfly)</td>
</tr>
<tr>
<td></td>
<td>xxtyan (hypocoristic suffix)</td>
<td>pikatyuu</td>
<td>tyokoreeto (chocolate)</td>
</tr>
<tr>
<td></td>
<td>otya (tea)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* There are no Japanese native words that contain the light syllables “kyu”, “gyu” and “ryu”.

Target sounds 1 to 3 were collected for the purpose of examining the error pattern of the “Cj” unit. Target sounds 4 and 5 (palato-alveolar consonants) are discussed separately later in this paper.

6.4 Test format
Toys and picture books designed for two year-olds were used in order to collect the target contracted sounds in the format of an object-naming task. As for make-up names for stuffed dolls, the researcher introduced the dolls and asked the children to repeat their names; “This is XX kun (Mr. XX), Can you say his name?” All sessions were transcribed by the researcher, and errors were counted and error patterns categorized. A second rater tested 10 percent of transcriptions and obtained an inter-rater reliability of 80 percent.

6.5 Findings
6.5.1 Pattern of errors in kjV, rjV and gjV
As some children did not know the target words and made no response to the researcher’s request, the total number of productions for each sound did not always reach 37. Words that contained “kyu” were familiar to many children and were produced more than once per child.
Error pattern was restricted to simplification and hardly any vowel epenthesis errors were observed. In other words, Japanese children did not violate the moraic timing frame even when producing unfamiliar words. This result supports the findings of Demuth (1996), Fee (1992), and Boysson-Bordies et al. (1992) that children have early sensitivity to the prosodic structure of words, regardless of ambient languages.

For both L1 and L2 learners, palatalized flaps caused more difficulty than palatalized velar stops. The L1 learners made an equal number of errors for /kj/ before all three vowels where it can occur. These findings contrast with the result from Experiment 1 that

* Both /CV/ and /jV/ were counted as simplification errors, CV.

**Figure 2. Proportion and Pattern of L1 Learners’ Errors**
(The entire production of each sound was taken as 100%)

**Table 6. Number of L1 learners’ Errors**

<table>
<thead>
<tr>
<th>Correct form</th>
<th>Incorrect CjV</th>
<th>CVCV</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>kya</td>
<td>36</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>kyu</td>
<td>47</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>kyo</td>
<td>25</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>rya</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>ryo</td>
<td>10</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>ryo</td>
<td>10</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>gya</td>
<td>25</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>gyu</td>
<td>21</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>gyro</td>
<td>27</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>206</td>
<td>48</td>
<td>69</td>
</tr>
</tbody>
</table>

Error pattern was restricted to simplification and hardly any vowel epenthesis errors were observed. In other words, Japanese children did not violate the moraic timing frame even when producing unfamiliar words. This result supports the findings of Demuth (1996), Fee (1992), and Boysson-Bordies et al. (1992) that children have early sensitivity to the prosodic structure of words, regardless of ambient languages.

For both L1 and L2 learners, palatalized flaps caused more difficulty than palatalized velar stops. The L1 learners made an equal number of errors for /kj/ before all three vowels where it can occur. These findings contrast with the result from Experiment 1 that
L2 learners did not make any errors in producing /kju/ and /gju/, which have an equivalent sound in English. Incorrectly produced palatalized velar stops [kj] turned out to be all palato-alveolar affricates, whereas the substitution errors of [gj] varied from [j] to [tʃ].

6.5.2 Pattern of simplification
Simplification errors were further divided into CV and jV. The jV type appeared most for the palatalized flap, while CV was the dominant pattern of simplification for “kyV” and “gyV”. This result could be because many of the children have not acquired the flap. Of the children who produced /ju/ for the palatalized flap, 65 percent also produced other sounds (typically [d]) for the plain flap, whereas [k] and [g] were acquired by most of the children. In addition, the correct production of [r] was seen in less than half of CV and most of the flaps were substituted with [d]. The substitution of [d] for [r] is often observed in Japanese (Ueda, 1996) and for the rhotic in English-acquiring children (Ingram, 1989).

If sonority strongly governed the simplification mechanism, “dV” should be a dominant pattern. Thus, the simplification pattern used by Japanese children is explained better by the principle “deletion of the marked element”.

6.5.3 Palato-alveolar consonants
Although palato-alveolars are the analogue to palatalized consonants in the inventory of Yo-on (See Table 1), children did not make analogous errors, replacing palatalized /ʃ/ with “plain” [s] and the like. The graph in Figure 4 below presents the rate of correct production and the pattern of errors produced by the children.
They did replace /tʃ/ with [t] before /a/ and /o/, but this is probably a case of "deaffrication" and not of "fronting", since there were no errors of [ts] for /tʃ/ before /u/. ([ts] is the allophone of /t/ before /u/). In fact, the most common error was the replacement of /ʃ/ with the affricate [tʃ]. Interestingly, the substitution of [tʃ] was also observed for /kj/ and for other consonants in non-target words such as the /t/ in “tukue” (desk). The early acquisition of affricates was also observed in these data as well. It is clear that the palatal affricate is the earliest sound to be acquired by Japanese children among [ʃ], [tʃ] and [ts]. The relative backness of tongue position in articulating [tʃ] compared with other sounds is possibly attributed to the ease of articulation for children whose oral cavity is smaller than adults.

6.5.4 Related errors

Another very frequent error in non-target words was similar to substitution of [tʃ] for /t/ in “tukue (desk)”. This was the substitution of the palato-alveolar [ʃ] for the alveolar fricative /s/ in words such as “sakana” (fish).

Figure 5 indicates the production of “ka-so” transcribed at the time of data collection for Experiment 2.
The substitution for palato-alveolar consonants was frequently observed in all alveolar fricatives. The low error rate in “si” [ʃi] implies that [ʃ] is acquired before [s]. This raised the question of whether the substituted palato-alveolars have some aspect of the alveolar fricative [s]. The palatalization gesture of children was further examined using spectrographic analysis to see whether error [ʃ] and target [ʃ] have the same spectral character.

The substitution errors of [ʃ] for [s] and the production of [ʃ] by the same child were compared in words, such as “kirinsan / penginsan” (Mr. Giraffe / Mr. Penguin) and “densha”(train). Auditorily they both sound like [ʃ].

The midpoint centroid value of fricative noise and F2 locus at the onset of the following vowel were calculated in seven children for whom this comparison could be made. The midpoint centroid value of [s] is expected to be higher than [ʃ], while the F2 locus of [s] is lower than [ʃ].

Figure 5. Production of “ka-so”

---

5 Centroid value = a weighted average of all the frequencies present in the spectrum (Zsiga, 2000)
The graph presents a covert contrast in the productions of palato-alveolar [ʃ]; the sound perceived as [ʃ] for [s] has a spectral characteristic that is close to [s]. The difference between two articulations was statistically significant: F(1,12) = 11.1 > 4.75, p =0.05. That is, some children are trying to produce [s], but have not yet achieved the target sound.

7 Summary and discussion

The productions of contracted sounds by L1 and L2 learners of Japanese presented an interesting contrast in the acquisition of moraic timing. Apart from a few accidental utterances of “CiV”, all L1 learners’ errors of contracted sounds occurred within the time frame of one mora. On the other hand, half of the L2 learners’ errors were moraic timing errors, such as vowel lengthening and insertion of gemination as well as vowel epenthesis even though they repeated the researcher’s pronunciation. These findings suggest that L2 learners need to make conscious efforts to acquire moraic timing frame, while L1 learners are equipped with the prosodic structure of words even for complicated sounds that are expected to be acquired late. The effect of native language phonology was found to be significant in sound acquisition as Vihman (1992) proposed.

The pattern of simplification error was further investigated. “Deletion of the marked element” and “maximal sonority rise” were possible principles to account for the error pattern. It was found that “ryV” was the only contracted sound that had an equal number of occurrences between “CV” and “yV” for L1 learners. The palatalized flap is recognized as a marked sound cross-linguistically (Hall, 2000), and consequently the
overall acquisition rate of [r] was low in Japanese as well as in English, which does not have the palatalized flap. L1 learners seem to have used the strategy of avoidance of the flap [r]. That is, “Deletion of the marked element” is a principle at work in Japanese children’s simplification patterns. The sonority-driven principle that governs the simplification pattern of consonant clusters did not operate on Japanese children’s production. We thus conclude that the /Cj/ unit is phonologically a palatalized consonant in Japanese. In contrast, the simplification patterns used by L2 learners were predominantly in the form of CV and created maximal sonority rise. If the palatal glide /j/ was treated as part of the vowel, the simplification pattern of an unfamiliar sound /CjV/ would be /jV/ and not /CV/ in L2 learners’ production. Apparently, English learners of Japanese regarded the contracted sounds as consonant clusters. Although we require further evidence to claim that the “Cj” unit in English is a consonant cluster, the different simplification patterns certainly derive from the phonological difference of the “Cj” unit in English and Japanese.

Among five palatalized consonants, affricates were acquired earliest by Japanese children aged between two and three years, followed by the palatalized voiceless velar stop [k], the palatalized voiced velar stop [g], palatal fricatives and the palatalized flaps. The strong preference for palato-alveolar consonants in the children’s substitution errors led us to conduct an acoustic analysis of their palato-alveolar consonant [ʃ]. The analysis revealed that the attempt by some children to produce [s] resulted in a sound whose spectral characteristic falls between [s] and [ʃ].

The general view that children are good at producing Yo-on, palatalized consonants, seems to come from the fact that many substitution errors of [t] or [s] are seen in palato-alveolar consonants [tʃ] or [ʃ]. Yet, children are in the process of learning the target sound and try to differentiate it from the palato-alveolar consonant, as we observed clearly in the comparison of the spectral characteristics of these two consonants. One possible explanation is that children generalize the palatalization rule before [i] in Japanese and apply it to a consonant before the other vowels. Further investigation of speech errors by children whose L1 is not Japanese is necessary for a fuller understanding of L1 production of palatalized consonants.

**Acknowledgements**

I would like to thank Mary Beckman at Ohio State University for her advice on the acoustic analysis and the Japanese participants (Megumi-room, Kyodo-Hoikuen, Seikatsu Kenkyuusho in Tokyo arranged by Chiaki Kosaka / Okayama Kohoku-, Ayumi-Hoikuen) for the data collection. I am also grateful to three anonymous Second Language reviewers for their helpful comments, suggestions and corrections. All remaining errors are my own.

---

6 A flap does exist in English, though its distribution is restricted to singleton onsets in unstressed syllables, such as the allophone of [t] in “better”.

---
References


— 46 —


