Sino-Japanese Gemination Revisited: A corpus-based approach

林智凱 / Chihkai Lin

大同大學應用外語學系

Department of Applied Foreign Languages, Tatung University

【摘要】

本研究旨在探討日語漢字音中 T 型(cvTv)及 K 型(cvKv)促音的變異情形。日本漢字音在不同的音韻環境之下，促音的形成有不同的結果。原形為齒齦音的 T 型促音其條件鬆散，後接無聲齒齦塞音(*t*-)、無聲齒齦擦音(*s*-)、無聲軟顎塞音(*k*-)以及喉擦音(*h*-)時，皆會形成促音。但在 K 型促音中，僅後接於無聲軟顎塞音*(k*-)之後，才會形成促音。

本研究依據三本漢字字辭典，建構一語料庫。以對比方式檢視 T 型及 K 型促音的分布及變異情形。語料分布情形顯示無論在組內或是跨組，K 型促音較 T 型促音更具有變異性。約半數之 K 型促音可具有促音形式(-*kk*-)，另一半維持原形(-*vkv*-)，而在 T 型促音中，則是超過九成之語料庫例子為促音(-*tt*-)。

根據語料庫分布情形，本研究認為日本漢字音中促音的形成為固定音韻變化，但仍出現變異。而變異的出現與三原因有關: 使用頻率、音韻限制及構詞。首先，使用頻率高的字，則是較容易不遵守促音規則，其結果可為促音，如數字六(roku)或是維持原形，如數字七(shichi)。使用頻率低的字，則是傾向維持原形。此外，若第一音節無聲化且第二音節促音化，則會使整個字失去前兩個音節的感知顯著性，母音無聲化也可能使得促音化不出現。本研究也提出不同的構詞法，如重疊詞，亦是促音較不易出現的原因。

【關鍵字】

日本漢字音、促音、變異、使用頻率、重疊詞

【**Abstract**】

This paper explores variants of *t*-roots and *k*-roots in Sino-Japanese gemination from a corpus-based approach. Sino-Japanese gemination is subject to the type of a root and the onset of the following syllable. Gemination in *t*-roots occurs when the following consonant is *k*-, *s*-, *t*-, or *h*-; gemination in *k*-roots takes place only when the following consonant is *k*-.

This paper establishes a corpus by collecting data from three Sino-Japanese dictionaries. The results suggest that the *k*-roots show more variants than the *t*-roots do in the corpus. Half of the *k*-roots turn into geminate (-*kk*-), and the other half of the *k*-roots maintain the default form (-*vkv*-), while more than 90% of the *t*-roots have become geminate (-*tt*-).

According to the distribution of the corpus examples, this paper suggests that gemination is a common phonological process in Japanese, but there is still variation in Sino-Japanese gemination, due to frequency, phonology, and morphology. Words that are frequently used like numbers, *roku* 'six' and *shich*i 'seven' tend to deviate from the general gemination process. Words that are rare to be seen tend to resist gemination. In addition, vowel devoicing also results in the resistance of gemination. Finally, the corpus data also reveal that reduplication might block gemination.

【**Keywords**】

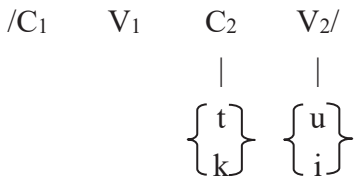
Sino-Japanese, geminate, variation, frequency, reduplication

1. Introduction

The phonological process of gemination (CV ~ /Q/ alternation) in Japanese has occurred in Yamato (or native) Japanese and Sino-Japanese, but the gemination in Yamato Japanese differs from that in Sino-Japanese. The former involves verbal configuration, as in the root tat- 'stand'. Its final form is *tats-u* 'stand-FNL' and the imperfect form is *tat-ta* 'stand-PERF'.[1] Nevertheless, the latter does not specifically deal with verbal configuration but takes place in the medial position of a compound in a root. For example, when *gaku* 'music' and *ki* 'instrument' form a compound, *gaku* + *ki* > *gakki* 'musical instrument', gemination occurs. An output like *gakuki without gemination is illegitimate in Japanese.[2]

The gemination process in contemporary Sino-Japanese has two major categories: *t*-roots and *k*-roots,[3] as discussed in Martin (1952), Vance (1987), Tateishi (1990), Itô and Mester (1996, 2015), Nasu (1996), Numoto (1997), Kurisu (2000, 2011), Huang (2004), Otaka (2009), Labrune (2012), and others.[4] Three phonological restrictions on Sino-Japanese gemination have been proposed in the literature. First, Itô and Mester (1996, 2015) have discussed a segmental composition for *t-roots* and *k-roots* in contemporary Sino-Japanese, as in (1).

(1)



[1] For the gloss, FNL=final, PERF=perfect.

[2] It should be carefully distinguished between doubling and geminate. Doubling in general refers to double consonants, CC, which include nasals like *mm* and non-nasals like *pp* (Akamatsu 1997). Geminate is a term limited to non-nasals, namely, *sokuon* in Japanese phonology.

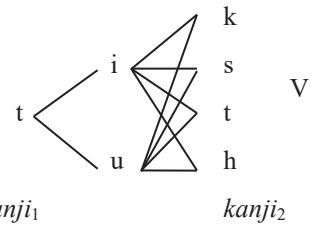
[3] Different terms are used in Itô and Mester (1996) and Itô and Mester (2015). Itô and Mester (1996) used *t*-stems and *k*-stems, but Itô and Mester (2015) changed the terminology to *t*-roots and *k*-roots. This paper follows Itô and Mester (2015) and adopts *t*-roots and *k*-roots.

[4] From a diachronic perspective, there should be *p*-roots, which have undergone sound changes and then turned into long vowels (Numoto 1989). Thus, *p*-roots no longer exist in contemporary Sino-Japanese. More details can be seen in Tateno (2012).

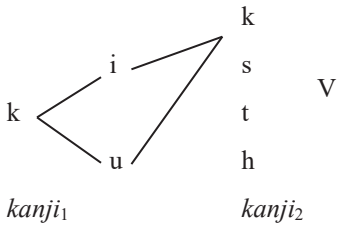
According to Itô and Mester (1996, 2015), the second vowel of *t*-roots and *k*-roots must be a high vowel, /i/ or /u/. Non-high vowels such as /a/ and /o/ are not permitted in this position. In addition to the segmental composition, there is a restriction on the phonological environment. Gemination takes place when the *t*-roots and *k*-roots are in compounds, but the contact of the second syllable in the first *kanji* and the first syllable in the second *kanji* triggers different gemination. Huang (2014: 110-111) presents a process of gemination for *t*-roots and *k*-roots, as seen in (2), where there are four consonants for the syllable, *k*-, *s*-, *t*- or *h*-, in the second *kanji*.

(2)

a. *t*-roots



b. *k*-roots



In (2), the *k*-roots are more restrictive than the *t*-roots are in gemination. The *t*-roots have to undergo gemination when the following root starts with a voiceless velar stop, a voiceless alveolar fricative, a voiceless alveolar stop or a glottal fricative. On the other hand, the *k*-roots only undergo gemination when they are followed by another root starting with a voiceless velar stop. Examples of Sino-Japanese gemination are shown in (3).[5]

(3)

a. *t*-roots + *k*-

e.g. 雪 *setsu* 'snow' + 花 *ka* 'flower' → 雪花 *sekka* 'snow flake'

b. *t*-roots + *s*-

e.g. 切 *setsu* 'bend' + 衝 *shou* 'rush' → 切衝 *sesshou* 'negotiate'

[5] When the onset of the first syllable in the second *kanji* is *h*-, the geminate is a bilabial stop *-pp-*. This paper omits this process.

c. *t*-roots + *t*-

e.g. 圧 *atsu* ‘press’ + 倒 *tou* ‘fall’ → 圧倒 *attou* ‘overwhelm’

d. *t*-roots + *h*-

e.g. 劣 *retsu* ‘inferior’ + 敗 *hai* ‘fail’ → 劣敗 *reppai* ‘failure’

e. *k*-roots + *k*-

e.g. 学 *gaku* ‘learn’ + 校 *kou* ‘school’ → 学校 *gakkou* ‘school’

The outline in (2) and the examples in (3) are associated with another phonological restriction on Sino-Japanese gemination. Itô and Mester (1996: 32) suggest that Sino-Japanese gemination does not occur when the second syllable starts with a voiced consonant, as in the examples 別段 (*betsu* + *dan*) ‘particularly’ and 学芸 (*gaku* + *gei*) ‘art and science’. The former is *betsu-dan* rather than *beddan; the latter is *gaku-gei*, not *gaggei.[6] When the following consonants are nasals *m*- and *n*-, there is no gemination. Examples of nasals are 国民 *kokumin* ‘citizen’ (*kommin), 国内 *kokunai* ‘domestic’ (*konnai). In addition, there is no gemination for liquid *r*-, as in 国立 *kokuritsu* ‘national’ (*korritsu).

Although the three phonological constraints account for most gemination in the *k*-roots and *t*-roots in Sino-Japanese, there are two types of counterexamples. The first type is an overgeneralization of gemination. As discussed in Vance (1987), Huang (2004) and Labrune (2012), *roku* ‘six’, for example, still has gemination when it is followed by classifiers starting with something other than a voiceless velar stop, as illustrated in (4).

[6] The underlying forms of the roots in Sino-Japanese are still debatable. On the one hand, Itô and Mester (1996, 2015) suggest that allomorph listing plays a larger role, and the second vowel of $C_1V_1C_2V_2$ is epenthetic. In their hypotheses, the default of each Sino-Japanese root is a closed syllable, CVC. On the other hand, the underlying form is disyllabic, CVCV, especially in traditional Japanese phonology. Shibatani (1990: 168) suggests that Sino-Japanese gemination is a consequence of deleting the second vowel in the CVCV structure and then being assimilated to the following consonant. Hamada (1950: 101) contends that gemination is a consequence of vowel devoicing and deletion. Kurisu (2000) assumes that the underlying form is either /(C)VC/ or /(C)VCV/, without specifying the default structure.

(4)

a.	*roku* + *satsu*	→	*rokusatsu*	'six-CL (books)'
b.	*roku* + *tou*	→	*rokutou*	'six-CL (large animals)'
c.	*roku* + *ko*	→	*rokko*	'six- CL (things)'
d.	*roku* + *haku*	→	*roppaku*	'six-CL (beats)'
e.	*roku* + *hoN*	→	*roppoN*	'six-CL (long things)'

The exceptions in (4d) and (4e) violate the generalization in (2). Unlike *t*-roots, a glottal fricative in the second root after *k*-roots does not trigger gemination. According to the generalization in (2), the readings of (4d) and (4e) should be *rokuhaku* and *rokuhoN* rather than the actual readings, *roppaku* and *roppoN* with gemination. However, the two exceptions are due to the high frequency of the root *roku* in Japanese.[7] Huang (2004: 111) contends that exceptions like (4d) and (4e) are mainly attested in numbers, such as 'six' and *hyaku* 'hundred', before a glottal fricative, which often turns into a bilabial stop.

The second type is the resistance of gemination. As suggested by the generalization in (2), there should be gemination when *t*-roots are followed by *k*-, *s*-, *t*- or *h*-, and *k*-roots by another *k*-. Nevertheless, counterexamples are observed in the compound of *baku* 'wheat' and *ki* 'air'. Its reading is *bakuki* 'the scent when wheat ripens', instead of *bakki with gemination. This type of counterexamples is also due to frequency, but it is an effect of low frequency that blocks gemination. The postulation that the low frequency blocks gemination has been explored by Otaka (2009). Otaka's (2009) results of production tests have suggested that frequency is a significant factor for irregular gemination. He found that when there were unfamiliar roots (low frequency), native speakers of Japanese preferred preserving the default CV structure to producing gemination in the outputs. In Otaka's (2009: 259) production tests, there were three classes, as shown in (5).

[7] Whether or not gemination occurs might have different interpretation. According to Labrune (2012: 32), the compound of *roku* 'six' and *hou* 'law' is interpreted as 'six kinds of law' when there is no gemination in the reading *rokuhou*. On the other hand, the interpretation of *roku* 'six' and *hou* 'law' as in the reading *roppou* with gemination is 'the Compendium of Laws' in particular.

(5)

Class 1:	The morphemes that always generate a geminate consonant regardless of the onset consonant of the second morpheme, unless it is voiced.
	Examples: *ichi* ‘one’, *zitsu* ‘real’, *zyuu* ‘ten’
Class 2:	The morphemes that generate a geminate consonant when the onset consonant of the second syllable of the first morpheme is identical to the onset consonant of the first syllable of the second morpheme, and they are both voiceless.
	Examples: *soku* ‘instant’, *seki* ‘red’, *haku* ‘white’
Class 3:	The morphemes that do not generate a geminate consonant at all, regardless of the phonological environment.
	Examples: *syuku* ‘inn’, *shichi* ‘seven’, *teki* ‘proper’

In (5), class 1 includes two consonants: *t*- and *h*-, and class 2 only includes *k*-. Examples in class 3 are *t*- or *k*- without gemination. Otaka’s (2009) classes of Sino-Japanese are based on the restrictions on the phonological environment in which gemination occurs. The less restriction the phonological environment has on gemination, the more frequent the morphemes are assumed to be. No gemination means that the frequency is the lowest. In Otaka’s (2009) production study, each class includes two distinct *kanji*, and each *kanji* derives four compound words. Therefore, there were eight compounds in each class. Otaka’s (2009) results have suggested that the rate of compounds in class 1 is the highest, and class 2 shows a high frequency when the morpheme is followed by consonant *k*-. The rate of class 3 is the lowest. Consequently, Otaka (2009) concludes that the three classes of Sino-Japanese morphemes behave differently. In terms of frequency, Otaka’s (2009) proposal suggests that class 1 is more frequent than class 2, and class 3 is the least frequent in Sino-Japanese.

Although Otaka (2009) categorizes three classes for Sino-Japanese gemination, there are some problems in Otaka’s (2009) class 3. By definition, this class lacks gemination, but it is extemporary to list this class, due to the fact that there are some counterexamples with gemination. For example, Otaka listed *teki* ‘proper’ in class 3,

and all of the examples lack gemination. Nevertheless, this morpheme alternates between the forms with and without gemination, as in 適歸 'return', which can be *tekiki* or *tekki*.

The inconsistency in Otaka's (2009) classes requires more data and detailed investigation. To explore Sino-Japanese gemination from another perspective, this paper adopts a corpus-based approach and focuses on *t*-roots and *k*-roots, which correspond to Otaka's (2009) classes 1 and 2, respectively. To understand more about the irregularity of Sino-Japanese gemination in *t*-roots and *k*-roots, this paper addresses two issues: (a) when they appear in unfamiliar words, how do *t*-roots and *k*-roots behave differently in gemination? (b) what is the variation rate in the two roots? To answer the two issues, this paper establishes a corpus by collecting data from three Sino-Japanese dictionaries. This paper is organized as follows. Section 2 introduces the steps of establishing the corpus and the data selection criteria. Section 3 reports the distribution of the corpus examples. Section 4 discusses variants in the corpus and provides possible causes for variation. Section 5 concludes this paper.

2. A corpus for Sino-Japanese *t*-roots and *k*-roots

To account for the irregularity that *t*-roots and *k*-roots resist gemination when they are followed by the other root starting with *k*-, *s*-, *t*- or *h*-, this paper establishes a corpus that contains roots with two *kanji*. The sources are twofold: one wordlist and three Sino-Japanese dictionaries. This paper uses *Kaitei joyo kanjihyo* 改定常用漢字表 'Revised Common *Kanji* List' (Agency of Cultural Affairs 2010) as the basic wordlist for the corpus. There are 344 *kanji* for the corpus: 120 *kanji* of *t*-roots and 224 *kanji* of *k*-roots. This paper extracted all the *t*-roots and *k*-roots from this wordlist and then consulted them with three Sino-Japanese dictionaries: *Kadokawa Shin Jigen* 角川新字源 (Ogawa, Nishida and Akatsuka 2004), *Shin Kangorin* 新漢語林 (Kamata and Yoneyama 2005) and *Zenyaku Kanjikai* 全訳漢辞海 (Sato and Hamaguchi 2006).

The corpus data were sorted according to the following three steps. The corpus individually listed all the roots. Next, the roots that appeared altogether in the three dictionaries were selected. Finally, the selected roots were compared for further analyses in terms of variation rate.

This paper collected all the compounds of the 120 *t*-roots and 224 *k*-roots in the three dictionaries, and the compounds contain two *kanji*. With regard to *t*-roots, there are 1025 compounds from *Kadokawa Shin Jigen*, 979 compounds from *Shin Kangorin*, and 910 compounds from *Zenyaku Kanjikai*. As for *k*-roots, there are 469 compounds from *Kadokawa Shin Jigen*, 418 compounds from *Shin Kangorin*, and 443 compounds from *Zenyaku Kanjikai*.

After data collection, the next step is to select the compounds that all appear in the three dictionaries. In total, 664 compounds of *t*-roots and 286 compounds of *k*-roots with two *kanji* were included in the corpus. According to how they vary in the corpus, the compounds were divided into three groups: (a) within-group variation, (b) between-group variation and (c) gemination. The criteria for classifying the compounds are as follows. Table 1 shows a format of data classification.

Table 1: Format of data classification

Examples		Dictionaries			Classification		
		Kadokawa Shin Jigen	*Shin Kangorin*	*Zenyaku Kanjikai*	a	b	c
閣下	'Your Excellency'	kakka	kakuka kakka	kakka	Yes		
敵国	'a hostile nation'	tekkoku	tekikoku	tekikoku	No	Yes	
越権	'go beyond one's power'	ekken	ekken	ekken	No	No	Yes
説書	'storytelling'	setsusho	setsusho	setsusho	No	No	No

*a: within-group variation;
b: between-group variation;
c: gemination

The within-group variation refers to whether a compound has two readings in the same dictionary. If the dictionary compilers provided a compound with two readings, the compound was marked as *yes* in the group of within-group variation. For example, the compound 閣下 'Your Excellency' in *Shin Kangorin* has two

readings: *kakuka* without gemination and *kakka* with gemination. In the other two dictionaries, this compound has only one reading *kakka*. Thus, the compound 閣下 'Your Excellency' is marked as *yes* in the group of within-group variation.

A compound with only one reading was marked as *no* in the group of within-group variation. Compounds in this group were compared to the group of between-group variation, which classifies the situations when the three dictionaries do not agree with one another on the reading of a compound. In this situation, the compound is marked as *yes* in the group of between-group variation. Take the compound 敵国 'a hostile nation' as an example. In *Shin Kangorin* and *Zenyaku Kanjikai*, it is *tekikoku* without gemination, while in *Kadokawa Shin Jigen*, it is *tekkoku* with gemination. Given that there are inconsistent readings in the three dictionaries, 敵国 'a hostile nation' was marked as *yes* in the group of between-group variation.

The last step is to classify compounds without any variation in the corpus, namely, those marked as *no* in the between-group variation. In this group, compounds with gemination were marked as *yes*, as in 越権 *ekken* 'go beyond one's power'. On the other hand, compounds without gemination were marked as *no*, as in 説書 'storytelling', *setsusho*, which is consistent in the three dictionaries. The results of the distribution are reported in Section 3.

3. Results

Table 2 shows the distribution of t-roots and k-roots in the corpus.

Table 2: Distribution of *t*-roots and *k*-roots in Sino-Japanese

	t-roots				*k*-roots			
Within-group variation	Yes	No			Yes	No		
	21	643			118	168		
Between-group variation		Yes	No			Yes	No	
		11	632			23	145	
Gemination			Yes	No			Yes	No
			612	20			118	27

In the group of within-group variation of *t*-roots, 21 tokens (3%, 21/664) show

inconsistent readings in the same dictionary, and 643 tokens (97%, 643/664) have consistent readings. In the 643 tokens with consistent readings, only 11 tokens (2%, 11/632) are not consistent among the three dictionaries, and 632 tokens (98%, 632/643) show no variants in the corpus. Among the 632 tokens in the gemination group, 612 tokens (97%, 612/632) include gemination, and 20 tokens (3%, 20/632) resist gemination.

In the group of within-group variation of *k*-roots, 118 tokens (41%, 118/286) vary in their readings in the same dictionary, and 168 tokens (59%, 168/286) show consistent readings. Among the 168 tokens in the group of between-group variation, 23 tokens (14%, 23/168) are inconsistent in the readings among the three dictionaries, and 145 tokens (86%, 145/168) are consistent in the readings. Out of the 145 tokens in the gemination group, 118 tokens (81%, 118/145) contain gemination, and 27 tokens (19%, 27/145) lack gemination.

In Table 2, there is a considerable difference between the two roots in the distribution. The two roots vary in how the readings are recognized in the three dictionaries. The variation comes from the sum of the tokens with two readings from the group of within-group variation and those from the group of between-group variation. As suggested by the percentage, variation is low for *t*-roots (5%, (21+11)/664), while approximately half of the *k*-roots have two readings (49%, (118+23)/286). It is clear that *t*-roots have fewer variants than *k*-roots in their readings when one token has two readings in the same dictionary. In other words, *t*-roots are relatively stable in gemination when they are followed by the other consonants *k*-, *s*-, *t*- or *h*-.

Before we move on to the discussion in Section 4, one issue regarding vowels in *t*-roots and *k*-roots needs clarification. Itô and Mester (1996, 2015) point out that the segmental composition of Sino-Japanese must use a high vowel /i/ or /u/ for the second vowel in a root, as in 七 *shichi* ‘seven’ and 敵 *teki* ‘enemy’ for a high front vowel, and 越 *etsu* ‘cross’ and 国 *koku* ‘country’ for a high back vowel. It should be clarified whether the second vowel affects gemination in *t*-roots and *k*-roots. Table 3 presents the distribution of the two high vowels in *t*-roots.

Table 3: Distribution of *t*-roots in Sino-Japanese (focusing on the second vowel in the first *kanji*)

	high back vowel			high front vowel		
Within-group variation	Yes	No		Yes	No	
	19	538		2	105	
Between-group variation		Yes	No		Yes	No
		11	527		0	105
Gemination		Yes	No		Yes	No
		520	7		92	13

557 tokens show that a high back vowel appears as the second vowel in the first *kanji*. In the group of within-group variation, 19 tokens (3%, 19/557) vary in their readings in the same dictionary, and 538 tokens (97%, 538/537) show consistent readings. In the group of between-group variation, 11 tokens (2%, 11/538) show different readings and 527 tokens (98%, 527/538) have only one reading. Among the 527 tokens with consistent readings, 520 tokens (99%, 520/527) show gemination, and only seven tokens (1%, 7/527) lack gemination.

With respect to the high front vowel as the second vowel in the first *kanji*, there are 107 tokens in Table 3. Two tokens (2%, 2/107) have inconsistent readings in the same dictionary, and 105 tokens (98%, 105/107) have only one reading. In the group of between-group variation of high front vowels, no token is marked as *yes*; all the tokens in this group show no variation. 92 tokens have gemination (88%, 92/105) and 13 tokens lack gemination (12%, 13/105).

Table 4 shows the distribution of the two high vowels in *k*-roots.[8]

Table 4: Distribution of *k*-roots in Sino-Japanese (focusing on the second vowel in the first *kanji*)

	high back vowel				high front vowel			
Within-group variation	Yes	No			Yes	No		
	97	145			25	18		
Between-group variation		Yes	No			Yes	No	
		21	124			2	16	
Gemination			Yes	No			Yes	No
			105	19			8	8

There are 242 tokens for high back vowels. In the group of within-group variation, 97 tokens (40%, 97/242) have two readings in the same dictionary, and 145 tokens (60%, 145/242) have only one reading. In the group of between-group variation, 21 tokens (14%, 21/145) show different readings and 124 tokens (86%, 124/145) are consistent in their readings. Among the 124 tokens, 105 tokens (85%, 105/124) have gemination, and 19 tokens (15%, 19/124) lack gemination.

There are 43 tokens in the group of high front vowels in Table 4: 25 tokens (58%, 25/43) with inconsistent readings in the same dictionary and 18 tokens (42%, 18/43) with only one reading. Among the 18 tokens, merely two tokens (11%, 2/18) show differences between groups. Half of the 16 tokens have gemination, and the other half has no gemination.

The above two tables reveal that the high front vowel in the first *kanji* shows slightly more tendency than the high back vowel to resist gemination in Sino-Japanese, but gemination still takes place in the two high vowels in *t*-roots and *k*-roots. The tokens with gemination outnumber those without gemination in high back vowels in the two roots and high front vowels in *t*-roots. As for high front vowels in

[8] The total amount of tokens in Tables 3 and 4 do not amount to 286 tokens. There are only 285 tokens because in the two tables, there is one ambiguous token 力行 'try hard to practice', which could be *ryokkou* (< *ryoku* + *kou*) or *rikkou* (< *riki* + *kou*).

k-roots, the tokens are equal (8 tokens vs. 8 tokens).

4. Discussion

The results in Section 3 show that some corpus instances do not undergo gemination. Section 4.1 discusses the exceptions, and Section 4.2 provides possible factors that influence the process of gemination. As the results in Section 3 also show different variation rates in *t*-roots and *k*-roots, Section 4.3 explores the distribution of the different degrees of variation.

4.1 When gemination does not occur

In Table 2, 20 tokens in *t*-roots and 27 tokens in *k*-roots resist gemination. The tokens of the *t*-roots that violate the generalization in (2) are listed in (6).[9]

(6)

a. *-tuC-*

逸口 *itsukou*	‘make an inappropriate remark’	質的 *sitsuteki*	‘target’
		説書 *setsusho*	‘storytelling’
穴見 *ketsuken*	‘narrow-minded’	蜜酒 *mitsushu*	‘sweet wine’

b. *-tiC-*

七去 *shichikyo*	‘seven conditions for divorce’	七竅 *shichikyou*	‘seven facial orifices’
七教 *shichikyou*	‘seven disciplines’	七経 *shichikei*	‘seven classics’
七賢 *shichiken*	‘seven noble men’	七書 *shichisho*	‘seven classics’
七出 *shichishitsu*	‘seven conditions for divorce’	七生 *shichishou*	‘seven lives’
七政 *shichisei*	‘seven stars’	七星 *shichisei*	‘Big Dipper’
七夕 *shichiseki*	‘seventh day of	七徳	‘seven virtues’

[9] In (6), the capital C in *tuC-* and *tiC-* refers to consonants *k-*, *s-*, *t-* and *h-*.

			the seventh lunar month’	*shichitoku* 七歩 *shichiho*	‘seven steps’
c.	察 *satsusatsu*	察	‘in detail’	卒 *sotsusotsu*	卒 ‘in a hurry’

The irregularities in (6) are divided into three subgroups: *-tuC-* (6a), *-tiC-* (6b) and reduplication (6c). Five corpus instances of *-tuC-* and 13 corpus examples of *-tiC-* are listed in (6a) and (6b). Two corpus instances of reduplication are listed in (6c). First, the five corpus instances in (6a) do not have anything in common. It might be simply due to their low frequency in Sino-Japanese. The corpus examples in (6b) are only attested when the first *kanji* is 七 *shichi* ‘seven’. This fact suggests a biased distribution where 七 *shichi* ‘seven’ resists gemination. Vowel devoicing in Japanese might result in the resistance of gemination in 七 *shichi* ‘seven’. This issue will be discussed in Section 4.2. As for the two instances in (6c), probably reduplication blocks gemination, but a quick conclusion should not be made based on only two examples. More data are needed in the future.[10]

The tokens of the *k*-roots that violate the generalization in (2) are listed in (7).

(7)

a. *-kuk-*

屋下 *okuka*	‘under the roof’	角掎 *kakuki*	‘take the horn and pull the feet’
覚海 *kakukai*	‘Another name for Buddhism’	覚劍 *kakuken*	‘sharpness of understanding’
曲肱 *kyokukou*	‘bend one’s elbow’	谷響 *kakukyou*	‘echo in the valley’
釈教 *shakukyō*	‘Sakyamuni	釈甲	‘take off armor’

[10] Special thanks go to one of the reviewers, suggesting that onomatopoeia can provide more examples for the restriction on gemination. For instance, *katsukatsu* ‘scarce’ and *kotsukotsu* ‘steadily’ do not undergo gemination. It is an interesting issue to explore gemination in onomatopoeia in the future.

		Buddha's teaching'	*shakukou*	
	粛恭 *shukukyou*	'reverent'	粛啓 *shukukei*	'honorific'
	淑景 *shukukei*	'beautiful day'	濯盥 *takukan*	'wash'
	逐客 *chikukaku*	'ask guests to leave'	麦気 *bakuki*	'the scent when the wheat ripens'
	覆考 *hukukou*	'inspect carefully'	約契 *yakukei*	'oath of alliance'
	約儉 *yakuken*	'simple and plain'		
b.	*-kik-*			
	隙駒 *gekiku*	'pass by rapidly'	尺蠖 *sekikaku*	'inchworm'
	色界 *shikikai*	'the rūpa-dhātu'	壁光 *hekikou*	'light from wall hole'
	暦官 *rekikan*	'an official who manages calendars'	歴階 *rekikai*	'ascend the flight of steps'
c.	郭隗 *kakukai*	'a person's name'	郭巨 *kakukyo*	'a person's name'
	易經 *ekikyou*	'Book of Changes'	赤縣 *sekiken*	'place name'

Likewise, the 27 tokens are divided into three subgroups: *-kuk-* (7a), *-kik-* (7b) and proper nouns (7c). There are 17 examples in (7a), six in (7b), and six in (7c). The fact that the 27 tokens do not undergo gemination might be due to their low frequency in Sino-Japanese. As for the six proper nouns without gemination, morphological constraints might be a factor that forbids gemination to take place across the boundary. For example, the morphological boundary in 郭隗 *kakukai* is between

the surname 郭 *kaku* and the given name 隗 *kai*. Gemination like *kakkai that occurs across the morphological boundary for 郭隗 would be illegitimate.

4.2 What blocks/feeds gemination?

Counterexamples in the corpus have been presented in Section 4.1. One type of counterexamples is attested in numbers, *shichi* 'seven' for instance. Although *roku* 'six' and *shichi* 'seven' are numbers, they vary in gemination. *Roku* 'six' is not limited to the restriction in (2) for *k*-roots. As shown in (4), glottal fricative *h*- after *roku* 'six' also undergoes gemination, as in *roppoN* 'six-CL (long things)' from *roku* + *hoN*. On the other hand, *shichi* 'seven' resists gemination, as suggested by the counterexample, 七歩 *shichiho* 'seven steps', in which the glottal fricative does not turn into geminate, *-pp-*.

According to Huang (2004: 111), it is the high frequency of use that frees *roku* 'six' from the phonological constraint in (2). As for *shichi* 'seven', there is another phonological motivation, namely, vowel devoicing of *shi* in *shichi*. The vowel devoicing prevents gemination in *chi* because if vowel devoicing and gemination concurrently take place, the auditory saliency of the first part of the word blurs. For example, in 七教 *shichikyou* 'seven disciplines', the high vowel in *shi* is devoiced. If gemination and vowel devoicing occur simultaneously, *shi* becomes [sḭ] and the word would be reduced to [sḭkkjo].

In addition to the irregularity in numbers, morphology also plays a role in blocking Sino-Japanese gemination, such as reduplication and morphological boundaries. In *t*-roots, reduplication as in 察察 *satsusatsu* 'in detail' blocks gemination, and the output cannot be *sassatsu.[11] In *k*-roots, proper nouns do not undergo gemination. In 郭隗 for example, it is *kakukai*, not *kakkai.[12]

In the corpus, gemination in *t*-roots also shows an intriguing irregularity. When the second *kanji* starts with a voiced consonant, such as *b* or *d*, gemination should be blocked, and the output should lack gemination, as in the examples 別段 (*betsu* +

[11] No example is observed in *k*-roots with reduplication in the corpus.

[12] Gemination is permitted in some proper nouns. For example, the poet's name 白居易 can be *haku-kyoi* or *hakkyoi*. Gemination is allowed in this example because this proper noun is frequently used in Japanese.

dan) 'particularly' and 学芸 (*gaku* + *gei*) 'art and science'. The blocking effect due to voiced consonants has been a constraint in Japanese phonology.[13] Nevertheless, unconventional gemination in voiced consonants is also observed in the corpus. The voiced consonant in the second *kanji* is first devoiced and then gemination is generated. Take 逸材 'a man of talent' as an example. The citation form is *itsu* for 逸 and *zai* for 材. The output should be *itsuzai* without gemination, but there are two readings in the three dictionaries. In *Kadokawa Shin Jigen*, it is read as either *itsuzai* or *issai*. In *Shin Kangorin*, it is *issai*, yet it is only *itsuzai* in *Zenyaku Kanjikai*. Given that this paper does not look into the devoicing process of voiced consonants and gemination in this situation, the irregularities in this type of gemination is left for future research.

4.3 Variation in *k*-roots and *t*-roots in the dictionaries

In Table 2, there is a difference between *t*-roots and *k*-roots in the ratio of variation. In *t*-roots, 32 tokens show variation (5%, 32/664). Approximately half of the *k*-roots have variants (49%, 141/286). The high ratio in the within-group variation in the *k*-roots should be discussed. The distribution of the 118 tokens in the group of within-group variation in *k*-roots reveals more details of the gemination process in *k*-roots, as seen in the distribution in Table 5 below.

There are six conditions in Table 5. The first condition is that the token in the three dictionaries has two variants. In this condition, there are only three corpus instances, as in 責過 'blame', which is marked as *sekika* and *sekka* in the three dictionaries. In the second condition, two dictionaries have variants and the third one has gemination. There are 18 tokens, such as 夕景 'view of sunset', which is marked as *sekkei* or *sekikei* in two dictionaries, and only as *sekkei* in one dictionary. It is also found that two dictionaries both have variants and the third one lacks gemination. The only example is 責課 'collect tax', which is *sekika* or *sekka* in two dictionaries, but it is only *sekika* in the third dictionary.

[13] Exceptions are found in loanwords from English as in *beddo* 'bed', which phonetically is more like *betto*.

Table 5: Distribution of *k*-roots in Sino-Japanese (within-group variation)

Conditions	a.	b.	c.	d.	e.	f.	Total
Tokens	3	18	1	62	9	25	118

* a = variants in all the three dictionaries
b = variants in two dictionaries and the other one with gemination
c = variants in two dictionaries and the other one without gemination
d = variants in one dictionary and the other two with gemination
e = variants in one dictionary and the other two without gemination
f = variants in one dictionary, one with gemination and one without gemination

The majority in Table 5 goes to the fourth condition in which there is a variant in one dictionary, but there is gemination in the other two dictionaries. For instance, 臆見 'subjective view' is *okuken* or *okken* in one dictionary, but in the other two dictionaries, it is only *okken*. The fifth condition is that there is a variant in one dictionary, but there is no gemination in the other two dictionaries, as in 託寄 'entrust', which is marked as *takuki* or *takki* in one dictionary, but as *takuki* in the other two dictionaries. The last condition is that there is a variant in one dictionary, and there are different readings in the other two dictionaries. Take 赤脚 'barefoot' as an example. It can be *sekikyaku* without gemination, *sekkyaku* with gemination or *sekikyaku ~ sekkyaku*, respectively.

The distribution in Table 5 indicates that gemination is still pervasive in the group of within-group variation. If the 62 tokens in Table 5 are added to the 118 tokens marked as gemination in Table 2, gemination becomes prominent in Sino-Japanese *k*-roots, as suggested by 63% of the corpus instances (180/286). If *t*-roots are taken into account, it is noticeably obvious that *k*-roots exceed *t*-roots in terms of variants, given that the ratio for the groups with variation in *t*-roots is low, with only 32 tokens in the corpus.

5. Conclusion

This paper has not only revisited Sino-Japanese gemination in *t*-roots and *k*-roots but also explored irregularities that resist gemination. The results conform to

Otaka's (2009) proposal that the gemination of *t*-roots exceeds that of *k*-roots. However, there are two major differences in this paper as compared to Otaka's (2009) findings. According to Otaka's (2009: 264-265) data, the average rate of *t*-roots with gemination is 89%, and the average rate of *k*-roots with gemination is 78%. Otaka (2009) suggests that both *t*-roots and *k*-roots are high in the rate of gemination. Nevertheless, this corpus-based approach to gemination shows a larger gap between *t*-roots and *k*-roots in Sino-Japanese: *t*-roots considerably higher than *k*-roots in their rates of gemination. Approximately 92% (612/664) of the corpus instances in *t*-roots show gemination, while only 41% (118/286) of the corpus instances in *k*-roots have gemination. The results in this paper prove that there is a significant gap between *t*-roots and *k*-roots in gemination.

This paper has also discussed variants of *t*-roots and *k*-roots in the corpus. The variation rates affect the results of gemination in *t*-roots and *k*-roots. As shown in Table 2, *k*-roots have remarkably more tokens of variation in the corpus, whereas most tokens of *t*-roots have undergone gemination. This difference between *t*-roots and *k*-roots in the variation rates is not observed in Otaka's (2009) results. Thus, this paper proposes that the variation rate in gemination plays a crucial role in distinguishing *t*-roots and *k*-roots.

Finally, this paper has also explained why more *k*-roots show variation. It was expected that *k*-roots would show less variation than *t*-roots in gemination, since the phonological environment for *k*-roots is far more restrictive than that for *t*-roots. The results suggest, however, that *k*-roots vary much more than *t*-roots do. One of the possible factors for the discrepancy would be that *k*-roots are less frequently used than *t*-roots. In the corpus, *t*-roots remarkably outnumber *k*-roots (664 tokens vs. 286 tokens). As the *k*-roots are less frequently used, compilers of the three dictionaries would have more difficulty in determining the readings.

In the future, two issues can be investigated. First, the interaction of voiced consonant devoicing and gemination in Japanese. As the corpus instance 逸材 'a man of talent' suggests, the irregularity has shown that low frequency triggers deviation and clearly gemination leads to an unusual phonological process in Japanese. In the future, some novel words can be used, and native speakers of Japanese will be invited to participate in a production test. The second issue for future

study is to look into sequential gemination, as in the words 北極圈 'arctic circle' and 白血球 'leukocyte'. The former is a compound of 北極 *hokkyoku* 'arctic' and 圈 *ken* 'circle', and the latter is a compound of 白 'white' and 血球 'blood cell'. The reading of 北極圈 'arctic circle' is *hokkyokuken* with one gemination, but theoretically, it can contain two geminations, as in *hokkyokken*. As for 白血球 'leukocyte', its reading is *hakkekkyu*. The two examples vary in internal branching, and there are variants in the readings. More examples are needed to show whether sequential gemination is acceptable in Japanese phonology.

References

Akamatsu, T. (1997). *Japanese Phonetics: theroy and practice*. Newcastle: Lincom Europa.

Hamada, A. (1950). The "Sokuon" and the "Hatsuon". *Studies in the Humanities* (Osaka City University), 1(1), 91-114.

Huang, H. (2004). *Huaren de Reben Yuyinxue* [Japanese Phonetics for Chinese]. Taipei: Jiliang books.

Itô, J., & Mester, A. (1996). Stem and word in Sino-Japanese. In *Phonological Structure and Language Processing: Cross-Linguistic Studies* (Speech Research Series. Vol. 12) eds by Otake, Takashi, and Ann Cutler. 13-44. Berlin: Mouton de Gruyter.

Itô, J., & Mester, A. (2015). Sino-Japanese Phonology. In *Handbook of Japanese Phonetics and Phonology.* ed. by Kubozono, Haruo. 289-312. Berlin: Mouton de Gruyter.

Kurisu, K. (2000). Richness of the base and root fusion in Sino-Japanese. *Journal of East Asian Linguistics* 9: 147-185.

Kurisu, K. (2011). Lexical Indexation in Sino-Japanese Exceptionality. In *Japanese/Korean Linguistics* vol. 18, ed. by McClure, William and Marcel den Dikken. 201-213. Stanford: CSLI publications.

Labrune, L. (2012). *The Phonology of Japanese*. Oxford: Oxford University Press.

Martin, S. (1952). *Morphophonemics of Standard Colloquial Japanese*. Supplement to *Language*. Dissertation No.47. Baltimore, Maryland: The Linguistic Society

of America.

Nasu, A. (1996). Niji-kango ni okeru sokuonka genshoo: Saitekisei-riron niyoru bunseki [Gemination in Sino-Japanese words: An analysis based on Optimality Theory], *Onseigakkai kaihoo [The Bulletin of the Phonetic Society of Japan]* 213, 27–40.

Numoto, K. (1997). *Nihon Kanjion no Rekishiteki Kenkyu—taikei to hyoki wo megutte* [A Study of the History of Sino-Japanese — focusing on system and orthography]. Tokyo: Kyuko shoin.

Otaka, H. (2009). *Phonetics and Phonology of Moras, Feet, and Geminate Consonants in Japanese.* Lanham, MD: University Press of America.

Shibatani, M. (1990). *The Language of Japan*. Cambridge: Cambridge University Press.

Tateishi, K. (1989). Phonology of Sino-Japanese morphemes. *University of Massachusetts Occasional Papers in Linguistics* 13, 209-235. Amherst, Massachusetts: GLSA.

Tateno, Y. (2012). Gendai Nihongo niokeru Shinnai Ryuseionn no Sokuonka [Consonant Gemination of Shinnai Nissho-on in Modern Sino-Japanese]. *Bulletin of the Faculty of Language and Literature* (Bunkyo University), 25(2), 1-21.

Vance, T. (1987). *An Introduction to Japanese Phonology*. New York: State University of New York Press.

Dictionaries

Ogawa, T., Taichiro Nishida and Kiyoshi Akatsuka. (2004). *Kadokawa shin jigen.* Tokyo: Kadokawa shoten.

Kamata, T., and Torataro Yoneyama. (2005). *Shin Kangolin*. Tokyo: Taishukan.

Sato, S. and Fujio Hamaguchi. (2006). *Zenyaku Kanjikai* (second edition). Tokyo: Sanseido.

Online source

Agency of Cultural Affairs. 2010. *Kaitei joyo kanjihyo* [Revised Common Kanji List]. (Retrieved from http://www.bunka.go.jp/seisaku/bunkashingikai/sokai/sokai_10/pdf/kaitei_kanji_toushin.pdf)

本論文於 2019 年 12 月 10 日到稿，2020 年 1 月 20 日通過審查。