Final draft: March 13, 2023
Tanaka, Yu (to appear) Vowels in Sino-Japanese roots: Revisited through the lens of Ural-Altaic vowel harmony. *Proceedings of the Workshop on Altaic Formal Linguistics (WAFL)* 16.

VOWELS IN SINO-JAPANESE ROOTS: REVISITED THROUGH THE LENS OF URAL-ALTAIC VOWEL HARMONY*

YU TANAKA
Doshisha University

1 Introduction

Sino-Japanese words, old loans from Chinese, have certain specific syllable structures (see Ito and Mester, 2015 and references therein). One of the common configurations that a Sino-Japanese root takes is $(C_1)V_1C_2V_2$, where C_2 is either *t* or *k* and V_2 is *i* or *u*, as is illustrated in (1).

(1) A common Sino-Japanese root structure:

$$\begin{array}{cccc} (C_1) & V_1 & C_2 & V_2 \\ & \left\{\begin{array}{c} i \\ e \\ a \\ o \\ u \end{array}\right\} & \left\{\begin{array}{c} t \\ k \end{array}\right\} & \left\{\begin{array}{c} i \\ u \end{array}\right\} \end{array}$$

The quality of V_2, namely *i* or *u*, is largely predictable from its phonological environment (Tateishi, 1990; Ito and Mester, 1996, 2015; Kurisu, 2000; Burness, 2016). As discussed in more detail below, it can be seen as a result of vowel harmony.

This study reexamines the distributional patterns of Sino-Japanese root-final vowels. I show that, although at first glance they may seem the relics of historical, irregular vowel harmony, they can be reanalyzed synchronically, considering data of recent loans and Ural-Altaic languages.

2 Data and issues

2.1 Data description

In the literature, the predictability of V_2 of a $(C_1)V_1C_2V_2$ root has been described as backness harmony triggered by preceding V_1, as shown schematically in (2).

*I thank the audience at WAFL 16, especially Donca Steriade and Deniz Özyıldız, as well as four abstract reviewers for their comments and suggestions. This study was partially supported by JSPS Kakenhi Grant 20K13019.

(2) Backness harmony in Sino-Japanese roots

$$(C_1)\ V_1\ \ C_2\ \ V_2$$

The basic patterns are summarized in (3) with representative examples.[1]

(3) V_2 is predictable from V_1

a. If V_1 is front {*i*, *e*}, V_2 may be *i*
一 *i t **i*** ‘one’ 式 *s i k **i*** ‘style’ 席 *s e k **i*** ‘seat’

b. Elsewhere, V_2 is *u*
圧 *a t **u*** ‘pressure’ 足 *s o k **u*** ‘foot’

Backness harmony does not apply across the board, however, and there are not a few exceptions. Its applicability seems to be affected by the type of intervening C_2 as well as the height of harmony-triggering V_1. If C_2 is *t*, V_2 is mostly *u* regardless of the backness of V_1 (e.g. 室 *situ* ‘room’; 鉄 *tetu* ‘steel’; 凸 *totu* ‘bump’), and only a handful of roots with C_2-*t* actually conform to backness harmony (e.g. 一 *iti* ‘one’; 日 *niti* ‘sun, Japan’). In roots with C_2-*k*, however, V_2 very often harmonizes with preceding front V_1 and is realized as *i*. Harmony is obligatory if V_1 is mid-front *e* (e.g. 席 *seki* ‘seat’; 歴 *reki* ‘history’), and is somewhat variable if it is high-front *i* (e.g. 式 *siki* ‘style’; 力 *riki* ‘power’; cf. 陸 *riku* ‘land’); otherwise, V_2 is always *u* (e.g. 足 *soku* ‘foot’; 国 *koku* ‘country’).

To examine the distributional patterns of V_2 more closely, I extracted all the Sino-Japanese roots of the $(C_1)V_1C_2V_2$ configuration from the list of Standardized Chinese Characters (常用漢字 *Jōyō Kanji*) published by the Japanese government (Bunkachō, 2010). In all, there are 392 such roots. Table 1 shows the number of those roots broken down by V_1 and C_2V_2. Empty cells are shown in dark grey and marginal ones in light gray.

Table 1: Number of common SJ roots

V_1 \ C_2V_2	*t*		*k*	
	i	*u*	*i*	*u*
i	7	28	8	10
e	1	39	**39**	0
a	3	35	0	100
o	0	11	0	83
u	0	11	0	17

Simply eyeballing the raw counts, we observe two kinds of asymmetries. First, as suggested above, the effect of backness harmony looks stronger for roots with C_2-*k* than for those with C_2-*t*. After C_2-*t*, V_2-*i* is always outnumbered by V_2-*u*, and this is true even when V_1 is front *i* (7 vs. 28) or *e* (1 vs. 39), where backness harmony is expected. In the case of C_2-*k*, however, harmonic V_2-*i* is proportionately more common when V_1 is front: *i* (8 vs. 10) and *e* (39 vs. 0).

[1] Throughout the paper, the transcriptions are broad and shown in italics unless detailed phonetic realizations are relevant to discussion. For example, *ti* and *tu* would be [tɕi] and [tsu] respectively in narrower transcriptions.

The other kind of asymmetry is seen in the C_2-*k* columns with respect to the height of V_1. Note the abundance of C*e̲ki̲* (39; boldfaced) as well as the absence of disharmonic C*eku* (0). This contrasts with the relatively small number of C*i̲ki̲* (8) and not a few cases of disharmonic C*iku* (10). These data suggest that mid-front *e* in V_1 is a stronger harmony trigger than high-front *i* in V_1.

2.2 Diachronic facts

The variability and asymmetry in the data are partly attributable to history. The variation between *i* and *u* in V_2 is due to differences in the time of borrowing. CVCV roots, originally CVC in the source Chinese languages, underwent vowel epenthesis (e.g. 德 *|tək| → *toku̲* 'virtue'). It is speculated that epenthesis of *i*, as opposed to *u*, was not uncommon in earlier stages of the Japanese language (see e.g. Poser, 1984; Frellesvig, 2010:283).[2] In fact, all the *i*-final roots except C*eki*, that is, C*iti*, C*eti*, C*ati*, and C*iki*, reflect the pronunciations of old loans (呉音 *go-on*), which were brought into Japanese through the Korean peninsula around or before the 7th century.[3] Later loans (漢音 *kan-on*) generally preferred *u*-epenthesis, replacing *i*-final roots. However, some high-frequency items and religious terms (e.g. 日 *niti̲* 'sun, Japan'; 一 *iti̲* 'one'; 八 *hati̲* 'eight'; 色 *siki̲* 'material') have retained the archaic pronunciation (Hayashi, 1980; Komatsu, 1995; Haraoka, 2019). This has resulted in the coexistence of *i* and *u* as V_2.

The fact that *i*-epenthesis is more common in roots with C_2-*k* than those with C_2-*t* could also be explained in relation to the diachronic fact mentioned above. Historical documents suggest that, unlike CV*k*, some CV*t* roots were pronounced without vowel epenthesis until around the 16th century (see Hashimoto, 1948; Abe, 2017). For instance, the Japanese-Portuguese dictionary compiled by Jesuit missionaries in 1603 transcribes certain Sino-Japanese words with CV*t* roots as if they have coda consonants (e.g. 退屈 *taicut* 'boring'; 発熱 *fotnet* 'fever'; the transcription is the original); the same words are pronounced with epenthetic vowels in present-day Japanese (cf. 退屈 *taikutu̲* 'boring'; 発熱 *hatu̲netu̲* 'fever'). This means that CV*k* underwent vowel epenthesis earlier, when *i* was still commonly used, while CV*t* did so after *u* became the default epenthetic vowel (see above). It is also documented that even in old CV*t* roots that had originally undergone *i*-epenthesis, the vowel was changed to *u* (e.g. 脱 *dati* > *datu* 'strip'; 別 *beti* > *betu* 'other'; see Hayashi, 1980 and discussion in §4). These historical facts have yielded the *t*-*k* asymmetry that we see today.

What about the *i*-*e* asymmetry? As stated above, the use of *u* as an epenthetic vowel became the norm at some point in the history. However, even late borrowings adopted *i* as V_2 when V_1 was mid-front *e* (e.g. 歴 *re̲ki̲* 'history'; 敵 *te̲ki̲* 'enemy'), resulting in the large number of C*eki* forms. This was not the case when V_1 was high-front *i* (e.g. 陸 *riku̲* 'land'), and disharmonic C*iku* forms were created with *u*-epenthesis. As a consequence, it appears as though there is a difference between V_1-*e* and V_1-*i* in their harmony-triggering effects in present-day Japanese.

2.3 Relevance to synchronic phonology

As discussed in the previous section, the distributional patterns of final vowels in Sino-Japanese roots can mostly be explained in terms of diachrony. However, even if a phonological phenomenon exhibits complexity rooted in historical quirks, it should still be treated as a synchronic issue as

[2]It has also been claimed that Old Japanese had some form of vowel harmony (Arisaka, 1934), although it is unclear whether this has played any role in vowel epenthesis in Sino-Japanese words.

[3]See Frellesvig (2010:258–292) for the details of the history of borrowing and adaptation of Chinese words.

long as there is evidence that grammatical principles are somehow involved. In fact, such evidence comes from language-internal and typological facts.

2.3.1 Loanword adaptation

The process of backness harmony is actually not limited to Sino-Japanese roots. Recent loanwords occasionally show *i*-epenthesis, as opposed to standard *u*-epenthesis, after a syllable containing a front vowel and coda *k* in the source language (Shoji and Shoji, 2014). Examples are shown in (4).

(4) *i*-epenthesis in loanwords

a. In C*ek*_

|tɛksəs|	→	*t<u>e</u>k<u>i</u>sasu*	テキサス	'Texas'
|mɛksɪkoʊ|	→	*m<u>e</u>k<u>i</u>siko*	メキシコ	'Mexico'
|lɛksɪkɑn|	→	*r<u>e</u>k<u>i</u>sikon*	レキシコン	'lexicon'
|tɛkst|	→	*t<u>e</u>k<u>i</u>suto* (∼ *tek<u>u</u>suto*)	テキスト	'text'

b. In C*e:k*_

|steɪk|	→	*sut<u>e:</u>k<u>i</u>*	ステーキ	'steak'
|keɪk|	→	*k<u>e:</u>k<u>i</u>*	ケーキ	'cake'
|bɹeɪk|	→	*bur<u>e:</u>k<u>i</u>* (∼ *bure:k<u>u</u>*)	ブレーキ	'break'

As can be seen, *i*-epenthesis occurs in the contexts of C*ek*_ (adaptation of English |Cɛk|) and C*e:k*_ with long *e:* (English |Ceɪk|). It should be noted that this loanword phenomenon is quite variable. There are words that rather epenthesize *u* in the same contexts (e.g. |nɛkst| → *nek<u>u</u>suto* 'next'; |meɪk| → *me:k<u>u</u>* 'make, makeup'), or those that show within-item variation (see the words for 'text' and 'break' in (4) above).[4] In addition, words with *i*-epenthesis tend to sound somewhat old-fashioned, suggesting that this adaptation strategy may now be dying out (Shoji and Shoji, 2014). Nevertheless, the data still indicate that harmonic *i*-epenthesis is not merely a historical vestige found in old Sino-Japanese roots, but is governed by the phonology of present-day Japanese.

Furthermore, we can even find asymmetries of the same kinds as those in Sino-Japanese roots. *i*-epenthesis may also occur in C*ik*_ and C*i:k*_ with high-front *i(:)* (e.g. |mɪksɚ| → *m<u>i</u>k<u>i</u>sa:* 'mixer'; |lik| → *r<u>i:</u>k<u>i</u>* 'leek'); however, these cases are much rarer than C*e(:)ki* forms exemplified in (4). This indicates that the *i-e* asymmetry is also present in loanword adaptation; that is, mid-front *e* triggers backness harmony more than *i*. Additionally, *i*-epenthesis never happens in C*i(:)t*_ or C*e(:)t*_ with C_2-*t* (e.g. |skeɪt| → *suk<u>e:</u>t<u>o</u>*, **suke:t<u>i</u>* 'skate'), in line with the *t-k* asymmetry noted above.[5]

2.3.2 Typology

The *i-e* asymmetry in vowel harmony is also not unique to Japanese, but is found in other languages. Specifically, similar height-dependent harmony patterns are attested in a group of

[4] In the case of the word for 'break' in (4b), the variable forms are also associated with different meanings. *bre:ki* with *i* usually refers to 'break (of vehicle)', while *bre:ku*, a newer form, is mainly used to mean 'break' as in 'getting a big break in the entertainment industry'.

[5] That said, the blocking of *i*-epenthesis in this case is confounded with other factors. Epenthesizing *i* or *u* after *t* would cause affrication, as in *ti* [tɕi] or *tu* [tsu]. Most probably due to a strong faithfulness requirement to the consonant (preventing affrication), *o*-epenthesis is usually employed for coda *t* regardless of the preceding vowel (see e.g. Irwin, 2011:110; Shoji and Shoji, 2014; Kubozono, 2015).

languages commonly labeled "Ural-Altaic" languages. In Old Mongolian, of the two reconstructed front unrounded vowels */i/ and */e/, high front */i/ in word-initial syllables was neutral to backness harmony and could be followed by either front or back vowels, while mid front */e/ acted as a real front vowel in the harmony system, and was always followed by front vowels (see Svantesson et al., 2005:113-116,190-194). In sound-symbolic words in present-day Korean, high unrounded /i/ and /ɨ/ in non-initial syllables behave as "neutral" vowels, and they allow following vowels to be disharmonic (Kim-Renaud, 1978; but see Larsen and Heinz, 2012 for complications). These descriptions suggest that higher vowels tend to be weaker harmony triggers than lower ones.

The patterns of vowel harmony in Hungarian even more closely resemble those in Japanese. Hungarian has backness harmony, which is notably seen in suffixation (see Vago, 1980; Hayes et al., 2009 and references therein for details). The vowels in suffixes (e.g. dative suffix /nɛk∼nɔk/) agree with the rightmost vowels of stems in backness, as is shown in (5) with the examples taken from Hayes et al. (2009).

(5) Backness harmony in suffixation in Hungarian (Hayes et al., 2009)

a. If the stem-final vowel is front, front suffix [nɛk]

[y̲ʃt-nɛ̲k]	'cauldron-DAT'
[ʃofø̲ːr-nɛ̲k]	'chauffeur-DAT'

b. If the stem-final vowel is back, back suffix [nɔk]

[ɔblɔ̲k-nɔ̲k]	'window-DAT'
[glykoː̲z-nɔ̲k]	'glucose-DAT'

The four front unrounded vowels /i, iː, eː, ɛ/ are considered "neutral" in that they do not strictly enforce backness harmony when they appear stem-finally. For example, if a stem contains a back vowel followed by a front unrounded vowel, the suffix is variably front or back (e.g. [honveː̲d-nɛ̲k] ∼ [honveː̲d-nɔ̲k] 'Hungarian soldier-DAT'). If a stem only contains neutral vowels, the suffix is usually front, agreeing with the phonetic frontness of the stem vowels; even so, there is also a modest number of stems that take back suffixes. Examples are given in (6).

(6) Variability with neutral vowels (Hayes et al., 2009)

a. All-neutral stems: usually front suffix [nɛk]

[tsiː̲m-nɛ̲k]	'garden-DAT'
[kɛ̲rt-nɛ̲k]	'address-DAT'
[rɛ̲pɛ̲s-nɛ̲k]	'splinter-DAT'

b. All-neutral stems: sometimes back suffix [nɔk]

[hiː̲d-nɔ̲k]	'bridge-DAT'
[dɛ̲reː̲k-nɔ̲k]	'waist-DAT'

The variability is in fact dependent on the height of the neutral vowels. Hayes et al. (2009) conducted a corpus study and a nonce-word experiment. The results show that, both in real and nonce words, stems ending in high [i] or [iː] take more back suffixes (i.e. phonetically disharmonic [nɔk]) than those ending in high-mid [eː], which in turn take more back suffixes than those ending in low-mid [ɛ]. Put differently, among the neutral vowels, the lower the vowel, the more often it takes the suffix form that is phonetically more harmonic. Hayes et al. (2009) call this a "height effect," referring to a possible functional explanation (see Kaun, 1995, 2004). Recall now that we see similar patterns in Japanese; mid *e* requires the following vowel to be harmonic *i* more than high *i* does.

2.4 Interim summary

To summarize the discussion thus far, the final vowel of a (C)VCV Sino-Japanese root, which is historically an epenthetic vowel, takes the form of either *i* or *u*. Even though the vowel quality is not completely predictable, the trends of data can be characterized as being driven by backness harmony with height and place effects. Similar patterns are found in recent loanwords and in the vowel harmony systems of several Ural-Altaic languages.[6] This suggests that the phenomenon is productive and also governed by general phonological principles that are cross-linguistically active. I thus argue that the patterns of Sino-Japanese root-final vowels should not be simply dismissed as fossilized, historical vestiges; they should rather be attributed to the synchronic grammar of Japanese. In the following section, I propose an analysis of the patterns within the framework of Maximum Entropy Harmonic Grammar.

3 Analysis: MaxEnt modeling

3.1 MaxEnt HG

Maximum Entropy Harmonic Grammar (MaxEnt HG) is a probabilistic model of Harmonic Grammar (Legendre et al., 1990), which employs Optimality-Theoretic constraints (Prince and Smolensky, 1993/2004) with numerical weights rather than strict rankings. (For the details of MaxEnt HG and its early application in the field, see, e.g., Goldwater and Johnson, 2003; Jäger, 2007; Hayes and Wilson, 2008; Hayes et al., 2009.)

A MaxEnt HG model can generate a probability distribution over a set of output forms of a given input based on their violations of weighted constraints. For instance, the probability of some output y mapped from input x, or $P(y|x)$, is calculated using the formula in (7).

(7) Formula for calculating the probability of an output candidate

$$P(y|x) = \frac{\exp(\sum_{k=1}^{n} w_{\mathrm{k}} C_{\mathrm{k}}(x,y))}{Z} \qquad \text{, where}$$

$$Z = \sum_{y \in Y(x)} \exp(\sum_{k=1}^{n} w_{\mathrm{k}} C_{\mathrm{k}}(x,y))$$

There are n constraints (C_1, C_2, ..., C_{n}), and constraint C_{k} has some weight w_{k}. $C_{\mathrm{k}}(x, y)$ denotes the number of violations (counted in negative numbers, as in –1) assigned by constraint C_{k} to output y given its input x. The penalty score of an output candidate is calculated by multiplying the number of violations of each constraint by the associated weight and summing the results. Then,

[6] Here, I am not committed to the controversial view that these languages form a single language family. Instead, these data should be understood as indicating that vowel harmony systems in languages of the world that are not genetically related, or only distantly related if any, show similar patterns.

the natural exponential function of the sum is divided by Z, which is the sum of the exponentiated penalty scores of all possible output candidates for that input x (i.e. all y in the set $Y(x)$).

MaxEnt HG is also associated with a mathematically defined learning algorithm (Berger et al., 1996). When it is provided with training data along with a set of constraints, the model converges on the optimal constraint weights that maximize the probability of the observed data. The weights learned from the data can further be used to make predictions about how novel words will be pronounced.

Given its basic architecture, MaxEnt HG is suited for capturing linguistic data with variation and how such data are learned and reproduced by speakers. I thus adopt MaxEnt HG in order to model the synchronic phonological grammar of Japanese that generates the variable vowel harmony patterns in Sino-Japanese roots.

3.2 UR and constraints

The underlying representation of a CVCV Sino-Japanese root in synchronic analysis has been a subject of debate. One might assume that it is /CVC/ and [i] or [u] is epenthesized in final position, on the ground that the quality of the vowel is largely predictable. One may argue that it should rather be /CVCi/ or /CVCu/, as the final vowel is not completely predictable and some kind of lexical specification is necessary. There have been other proposals, including ones that mingle the two positions. (See e.g. Tateishi, 1990; Ito and Mester, 1996, 2015; Kurisu, 2000; Labrune, 2012:30-34; Burness, 2016; Poppe, 2022 for discussion). In this study, I posit /CVC/ and consider the final vowel on the surface to be epenthetic. This analytic choice is for the sake of simplicity, however, and is not crucial to the account proposed below. For example, a form like /CVCi∼CVCu/ would also be possible.[7]

On the assumption that the underlying form is /CVC/, I propose two constraints in (8) to capture the occurrence of vowel epenthesis.

(8) Constraints on vowel epenthesis

a.	NOCODA	Coda consonants are disallowed.	(e.g. *[kok])
b.	DEP-V	Vowel epenthesis is disallowed.	(e.g. */kok/ → [koku̲])

NOCODA disallows consonants in coda position, while DEP-V bans vowel epenthesis. In the data of CVCV Sino-Japanese roots in isolation, the last vowel is always present. It is thus expected that NOCODA has a higher weight than DEP-V, or the former outranks the latter in terms of constraint rankings.[8]

To replicate the height effect (i.e. the *i*-*e* asymmetry) in backness harmony, I employ two versions of harmony-trigger constraints in (9) following Hayes et al.'s (2009) analysis of Hungarian.

[7] A reviewer asks how a learner could posit /CVC/ if the form never surfaces as such in the language. It is worth noting here that a CVCV root actually quite often shows up in the form of CVC without the last vowel in root-root compounds (e.g. 学 *gaku* 'study' + 会 *kai* 'meeting' → 学会 *gak-kai* 'academic society'; see the references above for more details). Thus, it is not the case that /CVC/ is some abstract form that is too far removed from surface forms.

[8] As stated just above, NOCODA may be satisfied in compounds. The use of a constraint like CODACOND (Ito and Mester, 1994) would be more accurate given that Japanese actually allows certain types of codas. Here, I use NOCODA as it suffices for the purpose of this study.

(9) Two harmony-trigger constraints (Hayes et al. 2009)

a.	* [–back] [+back]	Vowels agree in backness with a preceding front vowel. (e.g. *{i, e}-u)
b.	$*\begin{bmatrix}-\text{back}\\-\text{high}\end{bmatrix}$[+back]	Vowels agree in backness with a preceding non-high front vowel. (e.g. *e-u)

(9a) is a constraint that is meant to capture the general backness harmony pattern. As non-high front *e* acts as a strong harmony trigger, it is expected that (9b), a specific version of (9a), will also gain some weight. It is important to note that the two constraints have been proposed to account for cross-linguistic tendencies of vowel harmony, and not specifically for the data of Sino-Japanese roots, and that they are argued to have functional motivations (see Kaun, 1995, 2004; Hayes et al., 2009).

For the place effect (the *t*-*k* asymmetry), I use IDENT[dorsal]. Additionally, two specific versions of DEP constraints are included in order to capture the differences between *i* and *u* as epenthetic vowels.

(10) Constraints on the choice of the epenthetic vowel

a.	IDENT[dor]	A change in [dorsal] is disallowed.	(e.g. */set/ → [setɕi])
b.	DEP-i	Epenthesis of [i] is disallowed.	(e.g. */kok/ → [koki])
c.	DEP-u	Epenthesis of [u] is disallowed.	(e.g. */kok/ → [koku])

The articulation of palatal consonants is characterized as involving tongue dorsum movement, and they are specified for both [+coronal] and [+dorsal] in some feature systems (see Hayes, 2009; cf. Recasens, 1990 for discussion of the actual articulatory characteristics of palatals and further complications). Epenthesis of *i* in the CV*t* context yields the sequence *ti*, which is phonetically realized as [tɕi]. The alveolo-palatal [tɕ] is a [+dorsal] segment, and thus a /t/ → [tɕ] alternation incurs a violation of IDENT[dor]. A high weight of IDENT[dor] would prevent *i*-epenthesis in the case of CV*t* but not CV*k*, capturing the the *t*-*k* asymmetry.[9] Also, in present-day Japanese, *u* is more commonly used as an epenthetic vowel than *i*. A higher weight of DEP-i than DEP-u would capture this preference.

3.3 Procedure and results

Using MaxEnt Grammar Tool (Wilson et al., 2006), I constructed a MaxEnt HG model with the seven constraints proposed above, and fitted the model to the data of 392 Sino-Japanese roots in Table 1. Then, I had it predict the output forms of novel inputs with the learned constraint weights, on the assumption that the same number of Sino-Japanese roots as in the training data were newly coined and provided to the grammar. For example, the model was given 18 C*ik* forms (the sum of 8 C*iki* and 10 C*iku* in the real data) as novel inputs, which would possibly be produced as C*ik*, C*iki* or C*iku* as output forms.

I first show in Table 2 the constraint weights learned by the model. Note that some constraints gained zero weights. Although these constraints do not play any role in accounting for the data in analysis, they are still shown here for comparison with their theoretically opposing constraints.

[9] One could also use IDENT([distributed]) to achieve the same effect. What feature is used is not germane to the issue here. Further note that epenthesis of a high vowel after *t* also causes affrication (*ti* [tɕi]; *tu* [tsu]), violating IDENT[delayed release]. As the constraint is not relevant to the discussion, it is not included in the model.

Table 2: Constraint weights

Constraint	Weight
NoCoda	17.712
Dep-V	0
* [–back] [+back]	4.856
$*\begin{bmatrix}-\text{back}\\-\text{high}\end{bmatrix}$[+back]	0.898
Ident[dor]	2.970
Dep-i	4.200
Dep-u	0

The weights generally follow the expectations stated above. NoCoda has a high weight, since CVC with a coda never occurs on the surface (e.g. 国 *kok<u>u</u>* 'country'; **kok*), as far as the data of roots produced in isolation are concerned. In contrast, Dep-V's weight is zero as it is always violated (but see Footnote 7 for the data of compounds where the constraint is satisfied). As for the harmony-trigger constraints, not only [–back][+back] (*{i, e}-u) but also [–back,–high][+back] (*e-u) have gained some weight. This indicates that the model detected the general trends of backness harmony as well as non-high *e*'s strong role as a trigger, which would produce the height effect. A relatively high weight of Ident[dor] dampens *i*-epenthesis specifically after CV*t*. Dep-i also has a higher weight than Dep-u, reflecting the status of *u* as the default epenthetic vowel.

Table 3 shows the number of output forms of novel (hypothetical) CVCV Sino-Japanese roots predicted by the model with the constraint weights above. The table is made in the same format as for Table 1. The original counts are also shown in parentheses. Note that the predicted numbers are not integers, as they are computed stochastically.

Table 3: Number of hypothetical SJ roots predicted by MaxEnt HG

V_1 \ C_2V_2	*t*		*k*	
	i	*u*	*i*	*u*
i	3.1 (7)	31.9 (28)	6.1 (8)	11.9 (10)
e	7.8 (1)	32.2 (39)	**32.2** (39)	6.8 (0)
a	0.0 (3)	38.0 (35)	1.5 (0)	98.5 (100)
o	0.0 (0)	11.0 (11)	1.2 (0)	81.8 (83)
u	0.0 (0)	11.0 (11)	0.3 (0)	16.7 (17)

By and large, the model reproduced the basic patterns in the training data. For example, the *i*-*e* asymmetry is seen from the abundance of harmonic C*eki* (32.2; boldfaced) with V_1-*e* as a trigger in comparison to the modest number of harmonic C*iki* (6.1) or the relatively large number of disharmonic C*iku* (11.9) with V_1-*i*. The model has also learned the *t*-*k* asymmetry in that its predictions show general preferences for *u*-epenthesis in CV*t* roots. Notice that *u* always outnumbers *i* after C_2-*t*, just as in the training data.

That said, the model also seems to have failed in several respects. It did not replicate the total absence of C*eku*, predicting a certain amount of such forms (6.8; cf. 0 in the training data). One other notable problem is that it produced relatively many C*eti* forms despite their scarcity in the actual data of Sino-Japanese roots (7.8; cf. 1). Lastly, C*ati* is predicted to never occur, although there are a few instances in real data (0; cf. 3; e.g. 八 *hati̲* 'eight'). In the next section, I discuss the cause of these issues and potential solutions, their implications, and a general assessment of the proposed grammar.

4 Discussion

The constructed MaxEnt HG model has successfully generated the basic patterns of vowel harmony in Sino-Japanese roots, but has also shown its limitations. First, the grammar produces some C*eku* forms, which is incongruous with real data. This is mainly due to the strong preference for *u* as an epenthetic vowel in the overall data. Note that *u*-epenthesis is dominant when roots with a back vowel as V_1 are also considered (see the large number of C*aku* and C*oku*). Conversely, *i*-epenthesis is greatly disfavored, and DEP-i necessarily receives a high weight. Thus, the model cannot produce enough C*eki* even with the strong triggering effect of V_1-*e*; hence the appearance of C*eku*.

It should be noted, however, that this actually seems closer to what is seen in the data of recent loans. Again, *u*-epenthesis is the default choice in present-day Japanese, and *i*-epenthesis occurs only occasionally and even variably in the C*e(:)k* context in loanwords (e.g. |nɛkst| → *neku̲suto* 'next'; |tɛkst| → *teki̲suto*∼*teku̲suto* 'text'; see §2.3.1). This suggests that, in the course of loanword adaptation, Japanese speakers with the knowledge of the patterns of Sino-Japanese roots have also applied backness harmony probabilistically to newly borrowed words. In other words, the model may not have completely failed in predicting novel forms.

The second problem is the overgeneration of C*eti*. This is mostly because the model learned the *i*-*e* asymmetry in CV*k* roots, giving a high weight to [–back,–high][+back] (*e-u), and the effect was carried over to the domain of CV*t*. However, the *i*-*e* asymmetry is not found with CV*t* in the data of real Sino-Japanese roots. In fact, the number of C*iti* (7) is even slightly higher than that of C*eti* (1). That is to say, the opposite patterns are found between CV*k* roots and CV*t* roots; V_1-*e* is a stronger harmony trigger in CV*k*, but V_1-*i* is stronger in CV*t*. It may be the case that the height effect actually depends on the place of intervening C_2 as well. To capture this, one could possibly split the harmony-trigger constraints into two types, one for CV*k* and the other for CV*t*. Such a move, however, should be justified with solid evidence for the motivation of the constraints. It is beyond this study's scope to examine whether any such evidence can be found from functional or typological standpoints.

At least some diachronic data point to an interesting fact. As stated in §2.2, Hayashi (1980) documents that some CV*t* roots had originally undergone *i*-epenthesis but later changed the epenthesized vowel to *u* (e.g. 脱 *dati* > *datu̲* 'strip'; 別 *beti* > *betu̲* 'other'). By investigating old Chinese-Japanese dictionaries from different periods, he further shows that the *i* > *u* change happened gradually, and the timing of occurrence was different depending on the context. Specifically, the change happened in the order of C*ut_* ≫ C*ot_* ≫ C*at_* ≫ C*et_* ≫ C*it_*. It makes sense that *i* first became *u* in the back vowel contexts (i.e. C*ut_*, C*ot_*, and C*at_*), if it was motivated by harmony. The order C*et_* ≫ C*it_* would then suggest that *i* resisted the sound change more in the context of C*it_* than in C*et_*. In other words, the harmony effect was stronger with V_1-*i* than with V_1-*e* after C_2-*t*, just as discussed in the previous paragraph. Further investigations on the history of Japanese as well as cross-linguistic vowel harmony patterns are needed to determine whether these should be incorporated into synchronic grammar as constraints.

The problem of C*ati* is a little more challenging. If *i*-epenthesis is motivated by backness harmony, there is no reason for selecting *i* in C*at_*. In fact, this may have to be explained in terms of lexical idiosyncrasy, rather than principled phonological factors. The three rare Sino-Japanese roots of the C*ati* shape are shown in (11).

(11) C*ati* roots

八 *hati* 'eight' 鉢 *hati* 'pot, bowl' 罰 *bati* 'divine punishment'

All these roots are very old forms. The first example is one of the Sino-Japanese numerals, which show peculiar patterns in many respects. The use of *i* here may also be due to some sort of paradigm leveling effect, since *i* often appears in other numerals as well (e.g. 一 *it<u>i</u>* 'one'; 七 *sit<u>i</u>* 'seven'). The second example is etymologically Sino-Japanese, but it is one of the morphemes that sound "nativized" to the ears of native Japanese speakers. It very often undergoes compound voicing known as rendaku (e.g. *hi* + *hati* → *hi-**b**ati* 'brazier'; *ueki* + *hati* → *ueki-**b**ati* 'flowerpot'), which a typical Sino-Japanese morpheme would usually resist. The last one actually has an alternative form with *u* (罰 *batu* 'punishment'), which is more generally used. The form with *i*, which is older, is used specifically to refer to 'punishment by supernatural beings'. All these cases may simply be lexicalized, which would not entail a grammatical treatment.

Despite all these issues, however, it should be highlighted once again that the basic patterns of backness harmony found in the data are well captured by the model. Specifically, the height and place effects are replicated with standard faithfulness constraints as well as markedness constraints proposed for another language with active vowel harmony (*[–back][+back] and *[–back,–high][+back]; Hayes et al., 2009). Sino-Japanese root-final vowels have long been studied in the literature, and the variability and skew in the data have also been noted (see e.g. Tateishi, 1990; Ito and Mester, 1996, 2015; Kurisu, 2000; Burness, 2016). None of these previous studies, however, have analyzed them in the context of cross-linguistic tendencies of vowel harmony. They are now reconsidered and brought back into the field of theoretical phonology. One of the next steps of this research would be to examine the true productivity of the patterns using nonce words. I leave this for future work.

5 Conclusion

This study has revisited the distribution of root-final vowels in Sino-Japanese roots. Although the variable and complex patterns are partly due to language-specific historical quirks, I have discussed them from a broader perspective. Pointing out their similarities to data of recent loanwords and vowel harmony systems of Ural-Altaic languages, I have argued that they should be accounted for by a synchronic phonological grammar. I have given an analysis within the framework of MaxEnt HG and shown that the basic patterns can be captured by standard constraints, some of which are also used for vowel harmony in Hungarian. The results and findings hold theoretical implications and also raise new issues, opening up possibilities for future research.

References

Abe, Seiya. 2017. Onin-shi [The history of phonology]. In *Nihon-go no oto [The sounds of Japanese]*, ed. Takuya Okimori and Hajime Kimura, 76–98. Tokyo: Asakura Shoten.

Arisaka, Hideyo. 1934. Kodai Nihongo ni okeru onsetsu ketsugō no hōsoku. *Kokugo to Kokubungaku* 11.

Berger, Adam L., Stephen A. Della Pietra, and Vincent J. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics* 22:39–71.

Bunkachō. 2010. Jōyō-kanji hyō [List of Chinese characters in regular use]. Published by Agency for Cultural Affairs. Available at:
https://www.bunka.go.jp/kokugo_nihongo/sisaku/joho/joho/kijun/naikaku/kanji/.

Burness, Phillip. 2016. Unproductive alternations and allomorph storage: The case of Sino-Japanese. *Journal of Japanese Linguistics* 32:79–99.

Frellesvig, Bjarke. 2010. *A history of the Japanese language*. Cambridge: Cambridge University Press.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, 111–120. Stockholm: Stockholm University.

Haraoka, Kaoru. 2019. Nihongo ni okeru kanjion no juyō to tsūjiteki hensen ni kansuru kenkyū [A study of the adoptation of Sino-Japanese sounds and their historical development in Japanese]. *Studies in Language and Culture (Tokyo Women's Christian University)* 57–76.

Hashimoto, Shinkichi. 1948. Kokugo no onsetsu kōzō no tokushitsu ni tsuite [The characteristics of syllable structures of Japanese]. *Kokugogaku* 1:24–43.

Hayashi, Chikafumi. 1980. Go-on-kei jion ni okeru zetsunai nisshōon no kana hyōki ni tsuite [On the kana-notation of Ancient Chinese final /-t/ in Japanese go-on materials]. *Kokugogaku* 122:55–69.

Hayes, Bruce. 2009. *Introductory phonology*. Malden: Wiley-Blackwell.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379–440.

Hayes, Bruce, Kie Zuraw, Péter Siptár, and Zsuzsa Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85:822–863.

Irwin, Mark. 2011. *Loanwords in Japanese*. Studies in Language Companion. Amsterdam and Philadelphia: John Benjamins.

Ito, Junko, and Armin Mester. 1994. Reflections on CodaCond and Alignment. In *Phonology at Santa Cruz*, Vol. 3, 27–46. Linguistics Research Center, University of California, Santa Cruz.

Ito, Junko, and Armin Mester. 1996. Stem and word in Sino-Japanese. In *Phonological structure and language processing: Cross-linguistic studies*, ed. Takashi Otake and Anne Cutler, 13–44. Berlin: Mouton de Gruyter.

Ito, Junko, and Armin Mester. 2015. Sino-Japanese phonology. In *The handbook of Japanese language and linguistics: Phonetics and phonology*, ed. Haruo Kubozono, 289–312. Berlin: De Gruyter Mouton.

Jäger, Gerhard. 2007. Maximum entropy models and stochastic optimality theory. In *Architectures, rules, and preferences: Variations on themes by Joan Bresnan*, ed. Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling, and Chris Manning, 467–479. Stanford: CSLI Publications.

Kaun, Abigail. 1995. The typology of rounding harmony: An optimality theoretic approach. Doctoral Dissertation, University of California, Los Angeles.

Kaun, Abigail. 2004. The phonetic foundations of the rounding harmony typology. In *Phonetically-based phonology*, ed. Bruce Hayes, Robert Kirchner, and Donca Steriade, 87–116. Cambridge: Cambridge University Press.

Kim-Renaud, Young-Key. 1978. Semantic features in phonology: Evidence from vowel harmony in korean. *Korean Linguistics* 1:1–18.

Komatsu, Hideo. 1995. Nihon jion no shotaikei: Dokujuon seibi no mokuteki o chūshin ni [The systems in Sino-Japanese phonology: Systematization for the purpose of sutra recitation]. In *Nihon kanjion shi ronshū [Papers on the history of Sino-Japanese phonology]*, ed. Hiroshi Tsukishima, 13–37. Kyuko Shoin.

Kubozono, Haruo. 2015. Loanword phonology. In *The handbook of Japanese phonetics and phonology*, ed. Haruo Kubozono, Handbooks of Japanese language and linguistics 2, 313–361. Berlin and New York: De Gruyter Mouton.

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

Labrune, Laurence. 2012. *The phonology of Japanese*. The Phonology of the World's Languages. Oxford: Oxford University Press.

Larsen, Darrell, and Jeffrey Heinz. 2012. Neutral vowels in sound-symbolic vowel harmony in Korean. *Phonology* 29:433–464.

Legendre, Géraldine, Yoshiro Miyata, and Paul Smolensky. 1990. Harmonic grammar — A formal multi-level connectionist theory of linguistic well-formedness: An application. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, 884–891. Mahwah, NJ: Lawrence Erlbaum Associates.

Poppe, Clemens. 2022. Sino-Japanese C∼CV alternations in Relational Morphology. Presentation at Phonological Association in Kansai. On-line, December 2022.

Poser, William. 1984. Hypocoristic formation in Japanese. In *Proceedings of the West Coast Conference on Formal Linguistics 3*, ed. Mark Cobler, Susannah MacKaye, and Michael Wescoa, 218–229. Stanford: Stanford Linguistics Association.

Prince, Alan, and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Cambridge, MA: Blackwell Publishers.

Recasens, Daniel. 1990. The articulatory characteristics of palatal consonants. *Journal of Phonetics* 18:267–280.

Shoji, Shinichi, and Kazuko Shoji. 2014. Vowel epenthesis and consonant deletion in Japanese loanwords from English. In *Supplemental Proceedings of the 2013 Annual Meeting on Phonology*, ed. John Kingston, Claire Moore-Cantwell, Joe Pater, and Robert Staubs.

Svantesson, Jan-Olof, Anna Tsendina, Anastasia Karlsson, and Vivan Franzén. 2005. *The phonology of Mongolian*. The Phonology of the World's Languages. Oxford: Oxford University Press.

Tateishi, Koichi. 1990. Phonology of Sino-Japanese morphemes. *University of Massachusetts Occasional Papers in Linguistics* 13:209–235.

Vago, Robert. 1980. *The sound pattern of Hungarian*. Washington, DC: Georgetown University Press.

Wilson, Colin, Ben George, and Bruce Hayes. 2006. UCLA Maxent Grammar Tool. Available at: http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/.