Phonology (2025), **Vol. 42**:e3 1–41
doi:10.1017/S0952675725000016

PHONOLOGY

ARTICLE

How tight is the link between alternations and phonotactics?

Jongho Jun, Hanyoung Byun, Seon Park and Yoona Yee

Department of Linguistics, Seoul National University, Seoul, Republic of Korea.

Corresponding author: Jongho Jun; Email: jongho@snu.ac.kr

Received: 4 January 2022; **Revised:** 9 August 2023; **Accepted:** 20 November 2023

Keywords: phonotactics; derived environment effect; alternation

Abstract

This study tests the hypothesis that alternation patterns with strong lexical support are more robust than those with no, or weak, lexical support. Focusing on three alternation patterns in Korean with varying productivity and generality, we measured lexical support in two ways. First, we conducted an acceptability-rating experiment investigating Korean speakers' judgements on non-words with and without violations of the phonotactic constraints motivating the alternations. In addition, we performed a simulation of learning a maximum entropy (MaxEnt) Harmonic Grammar from a dictionary corpus. The results of the experiment and computational modelling confirmed the hypothesis by showing that if an alternation is robust, its associated phonotactic constraint is learned with a high weight from the MaxEnt simulation, and it affects the participants' well-formedness ratings for non-words. Consequently, the results of this research support the claim of a tight link between alternations and phonotactics.

Contents

1. Introduction

This study addresses the relationship between alternations and phonotactics. In standard Optimality Theory (Prince & Smolensky, [1993] 2004), the same phonological constraints are employed to account for both alternations and their corresponding phonotactic restrictions, and thus a very tight link between alternations and phonotactics is assumed by default. This assumption is in line with previous studies (Hayes 2004; Prince & Tesar 2004; Jarosz 2006, 2011; Pater & Tessier 2006; Chong 2017, 2021) showing that alternations are easier to learn if they enforce lexically supported phonotactic patterns than if they do not. One well-known alternation pattern with mismatching phonotactics is derived environment, or non-derived environment blocking, effects (Kiparsky 1973 and others). Derived environment effect phenomena, in particular morphologically derived environment effects, have been described in the literature as phonological processes applying across morpheme boundaries in languages where their potential target sequences are attested within morphemes. In languages with derived environment effect patterns, the same sequences would be well-formed in lexical phonotactics but ill-formed across morphemes. Chong's (2017; 2019) recent studies on two well-known cases of derived environment effects, namely Korean palatalisation and Turkish velar deletion, suggest that derived environment effect phenomena do not necessarily pose a problem for the claim of a link between phonotactics and alternations. The results of Chong's corpus studies show that the constraint for Korean palatalisation is general and active both across stem–suffix boundaries and within morphemes whereas the constraint for Turkish velar deletion is active only in a small number of morphological contexts, and inactive within morphemes, indicating that phonotactics and alternations do not completely mismatch. In addition, the results of Chong's (2017; 2021) artificial language learning experiment show, in line with the previous studies mentioned above, that alternations with no phonotactic support are harder to learn than those with phonotactic support. Accordingly, alternations with no phonotactic support are predicted to be rare or unlikely.

However, not all previous phonological studies argue or assume that there is a tight link between alternations and phonotactics. Note that in early generative phonology (Chomsky & Halle 1968), in which lexical phonotactics is expressed by morpheme structure constraints whereas alternations result through the application of phonological rules, phonotactics may in principle be independent of alternations. Some recent studies (e.g., Paster 2013; Do & Yeung 2021) claim that alternations with matching phonotactics are accidental and language-specific, and thus there is no universal link between them.

Given the two opposing views on the connection between phonotactics and alternations, we test the following hypothesis for a link between phonotactics and alternations.

(1) *Hypothesis for a link between phonotactics and alternations*
Alternation patterns with strong lexical support are more robust (i.e., more productive and more general) than those with no, or weak, lexical support.

We focus on three alternation patterns in Korean: palatalisation, vowel harmony and laryngeal obligatory contour principle (OCP), which are known to have derived environment effects, and which vary in productivity and generality. As will be presented in §2, in Korean, palatalisation is highly productive and general whereas vowel harmony and laryngeal OCP are limited in both productivity and generality.

The test is carried out in two ways. First, we measure lexical support for each of the three alternation patterns, using Korean speakers' judgement ratings on non-words, under the assumption that speakers' judgements are a good proxy for lexical phonotactics. The hypothesis in (1) predicts that the more robust an alternation pattern is, the more its associated phonotactic restriction will affect participants' non-word ratings. Second, we measure lexical support for the same alternation patterns by conducting a simulation of learning a maximum entropy Harmonic Grammar (MaxEnt; e.g., Goldwater & Johnson 2003; Hayes & Wilson 2008). Specifically, we trained the UCLA Phonotactic Learner (Hayes & Wilson 2008) on a Korean monomorpheme lexicon, on the assumption that the learned weights of constraints

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

reflect over- or under-representation of a pattern in the lexicon. The hypothesis in (1) predicts that the more robust an alternation pattern is, the more likely its associated phonotactic constraint is to be learned and the greater its weight will be.

In order to measure lexical support for the alternation patterns, we conduct an acceptability judgement experiment on Korean non-words in addition to MaxEnt modelling on a Korean corpus for the following reasons. First, the corpus adopted in the present study might be flawed in part, and therefore, it may not be a good representative of the linguistic environment of an average Korean speaker. More importantly, speakers' phonotactic knowledge might not directly reflect statistical patterns in the raw lexicon if the speakers are biased towards under- or over-learning certain lexical patterns (Hayes *et al.* 2009; Becker *et al.* 2011; Hayes & White 2013, and others). Finally, by conducting both MaxEnt modelling on lexical data and an acceptability judgement test, we are able to find out to what extent speakers' phonotactic grammar is derived from the lexical data.

The results of the experiment and computational modelling, which are highly correlated with each other, are overall consistent with the predictions made from the hypothesis in (1), although some results are unclear. Specifically, the participants' ratings were lower for non-words violating the palatalisation constraint, and a high-weighted general palatalisation constraint was learned in the MaxEnt simulation. Further, in contrast, the participants' ratings were not lower for non-words violating vowel harmony constraints than for those satisfying them, and no vowel harmony constraints were learned in the simulation. Finally, the results of the acceptability-rating experiment showed no indication of laryngeal OCP, which might be due to a floor effect, and laryngeal OCP constraints prohibiting only some subsets of the potential target sequences were learned with low weights in the simulation. Based on these results, we conclude that there is a tight link between alternations and phonotactics, which we believe should be captured in the grammar and/or in its learning algorithm.

The rest of this article is organised as follows. In §2, we present background information of the selected alternation phenomena in Korean, discussing their productivity and generality. In §§3 and 4, we conduct an experiment checking Korean speakers' phonotactic judgements and perform computational modelling of Korean phonotactic grammar using a dictionary corpus. We report the procedures and results of the experiment and computational modelling, and compare their results to see if both methods of inquiry yield convergent results or not. In §5, we discuss some alternative analyses and interpretations of the results of the experiment and computational modelling. The final section concludes the article.

2. Some alternation patterns in Korean

This section presents background information about the three alternation patterns in Korean that we will focus on to test the hypothesis in (1). Our main concern will be to examine how productive and general each of these patterns is. As illustrated below, two of the three cases (palatalisation and vowel harmony) concern the constraint that drives the alternation itself, while the remaining case (laryngeal OCP in compound tensing) involves a constraint that blocks the alternation. Consequently, in our discussion, we will concentrate on the effect of application of the process in question in the former cases and the effect of blocking of the process in question in the latter case.

2.1. Palatalisation

In Korean, coronal stops before high front vocoids (henceforth TI) become alveo-palatal affricates, as shown in (2a) and (2b). This palatalisation does not apply within morphemes, as shown in (2c), and thus it has been considered an instance of a derived environment effect.

(2) *Korean palatalisation*

a. *Applies before inflectional suffixes*

i.	/pat^{h}-i/	[pac^{h}i]	'field-NOM'
ii.	/pat^{h}-ilaŋ/	[pac^{h}ilaŋ]	'field-COM'
iii.	/pat^{h}-ita/	[pac^{h}ita]	'field-COP'

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

b. *Applies before derivational suffixes*

i.	/mat-i/	[maci]	'eldest son-NMLZ'
ii.	/kot-i/	[koci]	'straight-ADV'
iii.	/put^{h}-i/	[puc^{h}i]	'adhere-CAUS'

c. *Blocked within a morpheme*

i.	/mati/	[mati]	'joint'
ii.	/ʌti/	[ʌti]	'where'
iii.	/t^{h}i/	[t^{h}i]	'blemish'

Korean palatalisation applies obligatorily at stem–suffix boundaries, regardless of whether the suffixes are inflectional, as in (2a), or derivational, as in (2b) (Cho 2009: 466). This indicates that Korean palatalisation in morphologically derived environments is not only productive but also general.

In order to have a better understanding of the productivity and generality of palatalisation, we need to consider two complications involving other processes which may interact with palatalisation. First, the palatalised forms occurring before inflectional suffixes (2a) are not the only attested surface forms. In Korean, noun stem-final coronal obstruents are in variation mainly with [s] before vowel-initial suffixes. Thus, when noun stems ending in coronal stops are combined with /i/-initial suffixes, [s]-final variants may occur in addition to the palatalised forms, as shown in (3a). This stem-final obstruent variation is not limited to the palatalisation context, as shown in (3b) (see Jun 2010 for more details of Korean noun stem-final obstruent variation).

(3) *Korean noun stem-final obstruent variation*

a. *Before /i/-initial suffixes*

i.	/pat^{h}-i/	[pac^{h}i] ~ [pasi]	'field-NOM'
ii.	/pat^{h}-ilaŋ/	[pac^{h}ilaŋ] ~ [pasilaŋ]	'field-COM'

b. *Before other vowel-initial suffixes*

i.	/pat^{h}-ɨl/	[pat^{h}ɨl] ~ [pasɨl]	'field-ACC'
ii.	/pat^{h}-esʌ/	[pat^{h}esʌ] ~ [pasesʌ]	'field-LOC'

Given that either palatalised forms or [s]-final variants, not [t^{h}]-final forms, are attested in a palatalising context, as can be seen in (3a), we may consider palatalisation to apply obligatorily, and the resulting palatalised forms such as [pac^{h}i] may further be subject to free variation targeting stem-final obstruents, optionally yielding [s]-final forms such as [pasi]. Note that although innovative [s]-final forms may occur as free variants of the palatalised forms, none of the attested surface forms in the palatalisation context violate the palatalisation constraint. Thus, it is true that the palatalisation constraint, that is, *TI, is never violated across stem–suffix boundaries. This palatalisation, optionally followed by noun stem-final obstruent variation, is quite robust, and is extended to loanwords. When English words ending in stops are borrowed into Korean, either vowel epenthesis or final coda unreleasing applies: for example, *Matt* [mɛt^{h}ɨ] ~ [mɛt̚] (personal name).[1] Such loanword forms take different allomorphs of the nominative and accusative case markers, following a general allomorphic rule of Korean: vowel- and consonant-initial suffix allomorphs are used after consonant- and vowel-final stems, respectively. Specifically, when loanwords for English words ending in a coronal stop are inflected with nominative case markers, stem forms ending in an epenthetic vowel are suffixed with *-ka* (4a). In contrast, stem forms ending in an unreleased stop are suffixed with *-i* (4b). Here, stem forms

[1]Syllable-final coda obstruents in Korean are always unreleased, though we do not mark this explicitly in phonetic transcriptions elsewhere in this article.

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

before *-i* can end in either [cʰ] or [s], and thus sequences violating the palatalisation constraint, such as [tʰ-i], are unlikely to occur.[2]

(4) *Loanwords ending in a coronal stop, inflected with nominative case markers*

a. *Stem forms ending in an epenthetic vowel*

i. [mɛtʰi-ka] 'Matt-NOM'

ii. [kʰʌtʰi-ka] 'cut-NOM'

b. *Stem forms ending in a consonant*

i. [mɛcʰ-i] ~ [mɛs-i] 'Matt-NOM'

ii. [kʰʌcʰ-i] ~ [kʰʌs-i] 'cut-NOM'

Note that because loanword stem forms with vowel epenthesis end in [tʰ], /tʰ/ can be considered as the underlying form of the loanword stem-final consonant in (4). Then, it is plausible to assume that [cʰ]-final forms occurring before *-i* are the result of applying palatalisation, and [s]-final forms occur due to the stem-final obstruent variation. This indicates that palatalisation, optionally followed by stem-final obstruent variation, applies to loanwords, confirming its high productivity.

The second complication is that TI sequences occurring in compounding and prefix–stem combinations usually undergo n-insertion, in which /n/ is inserted at the juncture of two morphemes when the first morpheme ends with a consonant and the following morpheme begins in a high front vocoid /i/ or /j/. This n-insertion is often optional, and thus TI sequences, which violate the palatalisation constraint, may surface, as shown in (5).[3] However, it seems that such non-palatalised forms are minor variants in the n-insertion context.[4]

(5) *Korean n-insertion (optional)*

a. *Compound*

i. /pak'atʰ-jaŋpan/ [pak'annjaŋpan] (~ [pak'atjaŋpan])
outside-man
'husband'

ii. /nos-jokaŋ/ [nonnjokaŋ] (~ [notjokaŋ])
brass-chamber.pot
'brass chamber-pot'

b. *Prefix–stem (derivational)*

i. /sus-jaŋ/ [sunnjaŋ] (~ [sutjaŋ])
male-sheep
'ram'

[2] To investigate the stem-final variation patterns of Korean nouns and loanwords, Jun & Lee (2007) performed production and acceptability ratings experiments on ten speakers of Kyungsang Korean (a variety spoken in the southeastern province of South Korea). The results of the production experiment for loanwords ending in a coronal stop occurring before a nominative case marker *-i* show that the participants produced mostly [s]-final (and, less frequently, [cʰ]-final) forms, and [tʰ]- or [t]-final forms were rarely produced: [s] (54), [cʰ] (5), [tʰ] (2), and [t] (1) (where the frequency of each attested stem-final consonant is shown in parentheses). The results of the acceptability-ratings experiment show that [cʰ]- and [s]-final forms received higher mean ratings than [tʰ]- and [t]-final forms: [mɛcʰ-i] (3.9) > [mɛs-i] (3.3) > [mɛtʰ-i] (2.9), [mɛt-i] (2.2); [kʰʌs-i] (4.5) > [kʰʌcʰ-i] (3.5) > [kʰʌtʰ-i] (2.7), [kʰʌt-i] (1.7) (where mean ratings shown in parentheses are based on a Likert scale of 1 (impossible) to 5 (perfect)). We, as native speakers of Korean, think that such a dominant use of palatalised and [s]-final forms for loanwords ending in a coronal stop is generally true in other dialects of Korean, including Seoul Korean.

[3] The phonetic transcriptions in (5) reflect two automatic phonological processes in Korean: obstruent nasalisation (i.e., an obstruent becomes nasal before a nasal) and coda neutralisation (i.e., coda obstruents are neutralised into homorganic unreleased lenis stops). Coda neutralisation followed by resyllabification applies across prefix–stem and stem–stem boundaries.

[4] According to the results of Jun's (2021) survey on Seoul Korean n-insertion, only a minority of participants chose forms with TI sequences, as opposed to n-inserted forms, as their preferred pronunciations: 2, 3, 7 and 4 out of 22 Seoul Korean participants chose [pak'atjaŋpan], [notjokaŋ], [sutjaŋ] and [cʰʌtjʌlim], respectively.

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

ii. /cʰʌs-jʌlɨm/ [cʰʌnnjʌlɨm] (~ [cʰʌtjʌlɨm])
first-summer
'early summer'

Consequently, it seems that there are no morphologically derived environments in Korean where the palatalisation constraint is violated frequently.

In summary, Korean palatalisation is highly productive at stem–suffix boundaries, and can also be considered general, given that it applies in all morphologically derived environments, except where n-insertion may apply.

2.2. Vowel harmony

To aid the understanding of Korean vowel harmony, let us first examine the vowel phoneme inventory of a dialect of Korean spoken in the Seoul-Gyeonggi area (henceforth, Seoul Korean), shown in (6).

(6) *Seoul Korean vowel phoneme inventory*

	Front		Back	
	Unround	Round	Unround	Round
High	i	y	ɨ	u
Mid	e	ø		o
	ɛ		ʌ	
Low		a		

This 10-vowel inventory can be considered a conservative one. The front rounded vowels /y/ and /ø/ are often diphthongised as [wi] and [we], respectively, surfacing faithfully only in some limited environments. In addition, it is controversial whether /e/ and /ɛ/ are merged completely or not. It seems true that the distinction between /e/ and /ɛ/ is completely lost at least among many younger speakers of Seoul Korean (see Lee 1996 and Cho 2016 for relevant discussion).

Korean has two types of vowel harmony, which hold in different domains: the inflection of verbs and adjectives (henceforth verbal inflection) and ideophones. The harmonic groups of vowels, traditionally called 'light' and 'dark', are different between the two vowel harmonies. The light vowel group consists only of /a o/ in the vowel harmony system of verbal inflection, as shown in (7a). In contrast, the light vowel group for ideophones is a larger set, as shown in (7b).

(7) *Vowel harmony: harmonic groups of vowels*

a. *Verbal inflection*
'light': /a o/
'dark': /i e ɛ y ø ɨ ʌ u/

b. *Ideophones*
'light': /a o ɛ ø/
'dark': /i e y ɨ ʌ u/

Since alternations involving vowel harmony occur only in verbal inflection, we will focus on the type of vowel harmony in verbal inflection in the remainder of this article.

In Korean, verb stems never occur in isolation; they are always suffixed. Verbal inflectional suffixes may be classified into the following four types, based on Choi (1985) and Kang (2006).

(8) *Four types of verbal suffix in Korean*

a. Vowel-initial suffixes: for example, *-a/ʌ* (DECL.IND and many other meanings)
b. Consonant-initial suffixes: for example, *-ca* (HORT)
c. (ɨ)C-initial suffixes: for example, *-(ɨ)mjʌn* ('if')
(/ɨ/ is present only when the stem ends in a consonant other than /l/.)
d. (Cɨ)C-initial suffixes: for example, *-(sɨ)p* (addressee honorific)
(Cɨ is present only when the stem ends in a consonant other than /l/.)

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

In Korean verbal paradigm, all vowel-initial inflectional suffixes have two basic allomorphs, [a]- and [ʌ]-initial, as illustrated in (9).[5]

(9) *Allomorphs of some vowel-initial inflectional suffixes in Korean verbal paradigm*

a. a ~ ʌ DECL.IND and other meanings
b. as' ~ ʌs' PAST
c. asʌ ~ ʌsʌ 'and then'
d. ala ~ ʌla IMP

The choice between the two allomorphs depends on which harmonic group the last stem vowel belongs to. The specific selection mechanism is illustrated in (10).

(10) *Vowel harmony in Korean verbal inflection: (last) stem vowel* vs. *suffix vowel*

Stem V	Suffix V	Stem	Suffixed form	
ʌ	ʌ	/mʌk/	mʌk-ʌ	'eat'
e	ʌ	/se/	se-ʌ	'count'
ɛ	ʌ	/mɛc/	mɛc-ʌ	'bear'
i	ʌ	/mil/	mil-ʌ	'push'
ɨ	ʌ	/tɨl/	tɨl-ʌ	'lift'
u	ʌ	/sum/	sum-ʌ	'hide'
ø	ʌ	/tø/	tø-ʌ	'become'
y	ʌ	/hy/	hy-ʌ	'bend'
o	a	/c'oc^{h}/	c'oc^{h}-a	'chase'
a	a	/mac/	mac-a	'correct, be hit'

[a]-initial forms are chosen when the last stem vowel is either /a/ or /o/, whereas [ʌ]-initial forms are chosen after any other stem vowel.

As discussed by Kang (2012: 7–12), in Middle Korean, vowel harmony was widely attested. Not only verbal but also nominal inflections showed harmonic alternations. Harmony also held within native Korean roots. Sino-Korean roots did not obey vowel harmony, but they constituted only a minority of the Middle Korean vocabulary (Kim 2007). Thus, it is generally accepted that vowel harmony was pervasive and productive in Middle Korean. Since then, a variety of historical changes, such as vowel shift, vowel loss, monophthongisation, and the expansion of Sino-Korean vocabulary, have disrupted the vowel harmony system, resulting in its limited application in Contemporary Korean. Specifically, vowel harmony alternation takes place only in verbal inflection, not in nominal inflection.[6] In addition, there exist disharmonic root morphemes in the Korean lexicon, as shown in (11), and thus Korean vowel harmony can be considered an instance of derived environment effect.

(11) *Disharmonic native Korean roots*

a. *Noun*
   i. /tasʌs/ 'five'
   ii. /nolu/ 'roe deer'

b. *Verb and adjective*
   i. /kajʌp/ 'feel pity'
   ii. /tat^{h}u/ 'quarrel'

Since Korean vowel harmony applies only in verbal inflection, it cannot be considered a general process. Moreover, even within verbal inflection, the domain of application of vowel harmony is

[5]Kang (2012: 3) reports that sixteen such alternating suffixes are attested in the National Institute of the Korean Language (NIKL) frequency data.

[6]There exist nominal suffixes that begin with /ʌ/ or /a/, but they do not show vowel harmony alternations. For instance, a vocative case marker /-a/ is always realised as [a], regardless of whether the last stem vowel is dark or light: for instance, /kihoŋ-a/ [kihoŋa] '(name)-VOC' and /kihun-a/ [kihuna] '(name)-VOC'.

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

subject to several further limitations. First, only vowel-initial, not consonant-initial, suffixes participate in vowel harmony: for example, /mʌk-A/ [mʌkʌ] *[mʌka] 'eat-DECL.IND', but /mʌk-ca/ [mʌkc'a] *[mʌkc'ʌ] 'eat-HORT' (where /A/ represents the underlying form of the alternating suffix-initial vowel).[7] Second, only the initial vowels of vowel-initial suffixes alternate due to vowel harmony, and non-initial suffix vowels have no harmonic variants, as can be seen in (9c) and (9d). Third, only the last vowel of the stem can trigger vowel harmony, and stem vowels other than the last one do not affect vowel harmony: for example, /k'apul-A/ [k'apulʌ], *[k'apula] 'behave badly-DECL'. Consequently, vowel harmony holds only between the last vowel of the verbal stem and the initial vowel of the following vowel-initial suffixes. Given that vowel harmony applies only in such a restricted domain, the generality of Korean vowel harmony is quite limited.

In order to ascertain how productive vowel harmony in Korean verbal inflection is, let us begin by considering which vowel, /a/ or /ʌ/, is the underlying form of the alternating suffix-initial vowel. Recall that [ʌ]-initial forms are used for a larger set of trigger stem vowels, that is, all vowels other than /a o/. This is not the only reason why /ʌ/ can be considered as the underlying vowel. The surface forms of the alternating suffixes not adjacent to the preceding stems provide more definite evidence for the underlying form of the suffix-initial vowel. As discussed above, vowel harmony applies only when the alternating suffixes are adjacent to stems. For instance, as shown in (12a), when the DECL.IND suffix is immediately preceded by a stem with a light vowel /a o/, it is realised as [a]. In contrast, as shown in (12b), when the same suffix follows another suffix, it is always realised as [ʌ]. Note that the preceding suffix is realised as a light alternant, [as'], and thus the [ʌ]-initial form of the DECL.IND suffix in (12b) cannot be due to vowel harmony. This clearly suggests that the underlying vowel of the DECL.IND suffix is /ʌ/.

(12) *V-initial suffixes in initial and non-initial positions*

a.	/nok-A/	[nok-a]	'melt-DECL.IND'
b.	/nok-As'-A/	[nok-as'-ʌ]	'melt-PST-DECL.IND'

Given that it is possible that all attested [ʌ]-initial forms of the alternating suffixes are simply the underlying forms, it is only in [a]-initial suffix forms that we can be sure vowel harmony has applied.

Since vowel harmony shows exceptions and variations, [a]-initial suffix forms are not as frequent as expected when the last stem vowel is a light vowel /a o/, as experimentally shown in some previous studies on novel Korean words (Kang 2012, 2016; Jang 2017) and as recently discussed and emphasised by Jo (2023). The most prominent exceptional pattern is that when the last stem vowel is /a/, disharmonic forms are attested often: for example, /mac-A/ [mac-a] ~ [mac-ʌ] 'correct-DECL'. Even more frequent exceptions occur when the stems are multi-syllabic p-irregular (in which stem-final /p/ appears as [w] before vowel-initial suffixes), regardless of whether the last stem vowel is /a/ or /o/: for example, /kølop-A/ [kølow-ʌ] (~[kølow-a]) 'painful-DECL' and /alimtap-A/ [alimtaw-ʌ] (~[alimtaw-a]) 'beautiful-DECL'. Accordingly, the only case where vowel harmony is known to apply obligatorily is with regular verb and adjective stems with /o/ as the last stem vowel: for example, /nok-A/ [nok-a], *[nok-ʌ] 'melt-DECL.IND'. However, the results of some recent experimental studies on Korean vowel harmony (Kang 2012; Jang 2017) show that exceptional disharmonic forms occur even after novel stems with /o/ as the last stem vowel, and that the rate of vowel harmony in novel Korean words does not differ depending on the quality of the last stem vowel, whether /a/ or /o/. Thus, the exceptionless harmony pattern triggered by /o/ is not extended to novel Korean words.

Some quantitative information about how often Korean speakers apply vowel harmony in speech can be obtained from the results of Kang's (2012) production experiment.[8] As shown in (13), in Kang's experiment, exceptional disharmonic forms occurred much more often for novel Korean words than for real Korean words.

[7] The phonetic transcriptions presented in this sentence reflect post-obstruent tensing (i.e., a lenis obstruent becomes tense after an obstruent).

[8] Kang's experiment included 30 native Korean speakers (19 Seoul, 4 Chungcheong, 6 Kyungsang and 1 Jeolla) as participants.

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

(13) *The rate of vowel harmony in Korean real and novel stems with a light vowel* (Kang 2012: 42)

a. Real stems: 81.3%

b. Novel stems: 32%

Furthermore, at least nine out of 30 Korean participants produced only [ʌ]-initial disharmonic forms for novel Korean stems with a light vowel, and additional six participants produced ʌ-initial forms for over 80% of the same novel stems (Kang 2012: 51, Figure 20).[9] A similar pattern indicating complete loss of vowel harmony was observed in the results of Jang's (2017) production experiment in which three out of 13 Seoul Korean participants produced only [ʌ]-initial forms for 12 verb stems with light vowels (four real and eight novel stems). Consequently, the results of the previous experimental studies using novel Korean words suggest that the true productivity of Korean vowel harmony is quite low, or even zero for some speakers.

In summary, Korean vowel harmony is neither general nor productive. The generality of vowel harmony is clearly limited in that it applies only between stem-last vowels and initial vowels of vowel-initial suffixes in verbal inflection. The productivity of vowel harmony is also limited since exceptions are frequent in existing words, and even more frequent in novel words.

2.3. Laryngeal OCP in compound tensing

For an effective illustration of phonation alternations in Korean compound tensing, let us start by examining the Seoul Korean consonant phoneme inventory, which is shown in (14).

(14) *Seoul Korean consonant phoneme inventory (C^h = aspirated, C' = tense or glottalised)*

	Labial	Coronal	Velar	Glottal
Stop	p p^h p'	t t^h t'	k k^h k'	
Affricate		c c^h c'		
Fricative		s s'		h
Nasal	m	n	ŋ	
Liquid		l		
Glide	w	j		

Note that obstruents in Seoul Korean and many other dialects of Korean show three-way phonation contrasts between lenis, aspirated and tense. Among them, the laryngeally marked consonants (aspirated and tense) will be the focus of the current discussion.

In Korean, when a compound consists of two nouns, and the second noun stem (N_2) begins with a lenis obstruent, tensing can occur as shown in (15).

(15) *Examples of Korean compound tensing*

a. /hɛ-pic^h/ [hɛp'it] 'sunlight' ('sun'+'light')

b. /kaɨl-pi/ [kaɨlp'i] 'autumn rain' ('autumn'+'rain')

c. /pom-pi/ [pomp'i] 'spring rain' ('spring'+'rain')

d. /mal-tɛk'u/ [malt'ɛk'u] 'retort' ('language'+'reply')

e. /mal-tat^{h}um/ [malt'at^{h}um] 'dispute' ('language'+'fight')

[9] In a scatter plot in Kang's Figure 20 in which a point represents the percentage of disharmonic forms for each participant, there are only 29 points for 30 participants. Our supposition is that two participants' points completely overlap, but if so, we do not know which two points those were. This is why we say that 'at least' nine, rather than exactly nine, participants produced only disharmonic forms for novel Korean stems with a light vowel /a o/ as the last stem vowel.

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

This tensing is often optional, and it shows lexical variation. One interesting tendency found in previous studies on Seoul Korean compound tensing (Zuraw 2011; Kim 2016) is the avoidance of tensing in words with laryngeally marked segments. Specifically, tensing is less likely when N_2 has tense or aspirated consonants, as in (15d) and (15e). This trend can plausibly be interpreted as the effect of laryngeal co-occurrence restrictions or laryngeal OCP. According to Ito's (2014) study on compound tensing in Yanbian Korean (a variety spoken in China's Yanbian Korean Autonomous Prefecture), this laryngeal OCP effect is more extensive in Yanbian Korean than in Seoul Korean, in that tensing is less likely even when the first component noun stem (N_1) has a laryngeally marked segment. Moreover, the laryngeal OCP effect that was found significant in existing compound nouns was mirrored in novel compounds. The laryngeal OCP effect has also been observed in Kang & Oh's (2019) acceptability rating experiments for existing and novel Korean compounds, which employed speakers of Jeolla Korean (a variety spoken in the southwestern province of South Korea). Thus, it seems that the laryngeal OCP effect is active and productive in compounds in various dialects of Korean.

Note that there exist monomorphemes with multiple laryngeally marked consonants, as shown in (16), and thus the laryngeal OCP effect can be considered an instance of a derived environment effect.

(16) *Examples of native Korean monomorphemes with multiple laryngeally marked consonants*

a. /t'uk'ʌŋ/ 'cover'
b. /t^{h}ok'i/ 'rabbit'
c. /k'at^{h}al/ 'hindrance'
d. /c^{h}alk^{h}ak/ 'with a snap'

Let us now consider how general and how productive the laryngeal OCP is in Korean. We are not aware of any morphological domains other than noun–noun compounding in which laryngeal OCP is active. In Korean, there are several tensing processes other than compound tensing which may create forms with multiple laryngeally marked onset consonants. First, post-obstruent tensing, in which a lenis obstruent becomes tense after an obstruent, is known to be an automatic process, and thus it is generally assumed to apply obligatorily. To the best of our knowledge, no previous studies report the blocking of post-obstruent tensing by the laryngeal OCP. In addition, the initial lenis consonants of adjective, noun and verb stems in Korean become tense optionally, as shown in (17).

(17) *Word-initial tensing (trivially modified from Kang & Oh 2016: 8,* SE *= sentence ender)*

a. *Adjective*
/cak-ta/ [cakt'a] ~ [c'akt'a] 'small-SE'
b. *Noun*
/sonaki/ [sonaki] ~ [s'onaki] 'squall'
c. *Verb*
/kam-ta/ [kamt'a] ~ [k'amt'a] 'to shampoo-SE'

The lexical meaning of adjectives and some nouns, but not verbs, can be intensified with this word-initial tensing. Kang & Oh (2016) collected (Jeolla) Korean speakers' acceptability ratings on word-initial tensing by employing a set of real Korean words which do not undergo word-initial tensing. Based on their statistical analysis of the experiment results, Kang & Oh argue that word-initial tensing is preferred in words with tense consonants. This is the opposite of the laryngeal OCP effect.[10] Thus, the results of Kang & Oh's experiment show no laryngeal OCP effect in word-initial tensing. Consequently, to the best of our knowledge, compound tensing is the only clear case, in which laryngeal OCP is effective in Seoul Korean and other dialects of Korean, and thus laryngeal OCP cannot be considered general in Korean.

[10] We are not convinced that the observed difference is statistically significant, since it seems that no random effects for item or subject were included in the statistical analysis.

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

Let us now consider how productive the laryngeal OCP is in Korean compound tensing. Note that tensing in compounding is variable, and laryngeal OCP decreases the tensing rate when stems have laryngeally marked consonants, as can be seen in (18a).[11] Using the tensing rate data from the previous studies in (18a), we have calculated the rates of blocking by the laryngeal OCP, as shown in (18b).

(18) *Tensing and blocking rates (%) in real and novel Korean compounds by the type of the onset consonant included in the second member (N_2) of a compound* (Seoul Korean data from Kim 2016: 20; Yanbian Korean data from Ito 2014: 367, 374)

		Seoul Korean	Yanbian Korean	
		Existing	Existing	Novel
a.	Tensing in words with N_2-medial…			
	i. lenis (or sonorant) onset	60.7	57.7	69.4
	ii. laryngeally marked onset	30.2	43.5	47.1
b.	Blocking by laryngeal OCP (= $100\% - \frac{(a.ii)}{(a.i)}$)	50.2	24.6	32.1

The resulting rates of blocking in (18b) indicate that the blocking by the laryngeal OCP is far from obligatory in both Seoul and Yanbian Korean. This partial blocking effect is extended to novel compounds, as reported by the experimental studies on Yanbian Korean (Ito 2014) and Jeolla Korean (Kang & Oh 2019). Kang & Oh (2016) discuss their experimental results by providing mosaic plots and statistical tests, not detailed rates of compound tensing, and so we cannot calculate the rate of blocking of compound tensing using their published results. However, Kang & Oh clearly show that the presence of aspirated or tense onsets in N_2 do not completely block compound tensing, regardless of whether N_2 is real or novel, and that the blocking effect is higher in real compounds than in novel compounds. In summary, the results of the previous experimental studies on Korean compound tensing suggest that the laryngeal OCP blocks compound tensing, even when N_2 is a novel stem, but the blocking effect is only partial, not complete.

In this section, we have discussed how general and how productive the laryngeal OCP is in Korean. We may conclude that both generality and productivity of the laryngeal OCP effect are limited, since the effect is attested only in noun–noun compounding, and it is only partial, not even close to complete.

2.4. A summary

Based on the discussions presented in the previous sections, we may determine how the three alternation patterns in Korean differ in robustness which we consider to be a function of productivity (namely, rate of application) and generality (namely, breadth of application domain). It seems clear that Korean palatalisation is highly robust, given that both its productivity and generality are high. In contrast, Korean vowel harmony and laryngeal OCP are low in robustness, since they are limited in both productivity and generality, as summarised in (19).

(19) *Robustness, productivity and generality of three alternation patterns in Korean*

	Productivity	Generality	Robustness
Palatalisation	High	High	High
Vowel harmony	Limited	Limited	Low
Laryngeal OCP	Limited	Limited	Low

[11] Tensing rates presented in (18a.i) are for lenis onsets in the columns for existing Seoul Korean and novel Yanbian Korean compounds, whereas the rates are for lenis and sonorant onsets for existing Yanbian Korean compounds. Ito (2014: 367, Table XII) does not provide separate values for lenis and sonorant onsets.

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

The robustness summarised in (19) may lead us to make the following prediction based on the hypothesis in (1): in Korean, palatalisation will have more lexical support compared to vowel harmony and laryngeal OCP. To test this prediction, we conduct an acceptability rating experiment for non-words in §3 and a computational modelling of learning a phonotactic grammar in §4, modelled loosely on Daland *et al.* (2011).

3. Experiment

In this section, to measure lexical support for each of the three alternation patterns described in the previous section, we perform an acceptability-rating experiment investigating Seoul Korean speakers' judgements on non-words. From the hypothesis in (1), it is predicted that the more robust an alternation is, the more the corresponding phonotactic pattern will influence participants' non-word ratings. More specifically, participants' ratings for target non-words violating a phonotactic constraint will be lower than those for control non-words satisfying it only when the associated alternation is robust. This prediction will be tested against the results of the current experiment.

3.1. Method

This section provides information about the stimuli, participants, and procedure utilised in the experiment.

3.1.1. Stimuli

In the experiment, we created 201 Korean non-words, of which eight were practice items and the rest were test items. (A list of test non-words is provided in the Supplementary Material.) The practice items were presented to participants before the main experiment. They were chosen to expose experimental participants to a wide range of acceptability, from high to intermediate to low.

The test stimuli consisted of three sets, each of which was intended to investigate phonotactics related to one of the three effects (palatalisation, vowel harmony and laryngeal OCP). We did our best to ensure several conditions. First, the words do not sound like any particular real Korean words or phrases. They also do not violate any major phonotactic constraints of Korean other than those we were examining in this study.[12] Further, all stimuli consist of two open syllables, of which the second may include a glide.[13] Thus, the stimuli can be schematised as $C_1V_1C_2(G)V_2$, where C = consonant, V = vowel and G = glide. For V_1 and V_2, we use six simple vowels: /i e a o ʌ u/. Specific details about each set are provided in the following paragraphs (see the Supplementary Material for more details about them).

Sixty-six Korean non-words were used to explore the palatalisation effect. In this set, the lenis and aspirated coronal stops, /t tʰ/, and their non-coronal counterparts, /p k pʰ kʰ/, occur in C_2 position. The high front vocoids, /i j/, and the other glide, /w/, occur in V_2 and G positions. As shown in (20), 20 target non-words with TI sequences and 20 control non-words with non-coronal stops followed by high front vocoids such as /pi/ (henceforth PI) were created. In addition, control non-words with coronal stops followed by vocoids other than high front vocoids such as /ta/ (henceforth TA) were created. Specifically, 10 control non-words with vowels other than /i/ and four control non-words with /w/ were added to the list.

[12] See Park (2020) for a recent review of Korean phonotactic restrictions found in previous studies.

[13] Note that disyllabic words are the shortest ones with which we can examine vowel harmony and laryngeal OCP. It seems that disyllabic words are short enough for Korean participants to consider them as monomorphemic, given that disyllabic words in our Seoul Korean corpus (which consist of native and Sino-Korean words) are predominantly monomorphemes (91%, i.e., 33,954 out of 37,364).

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

(20) *Number of target and control non-words in the palatalisation set by consonant–vocoid sequence type. (The number of non-words of the corresponding type is shown in parentheses)*

	Target with TI		C control with PI		V control with TA	
	ti	(5)	pi, ki	(5)	ta, to, tu	(5)
	tʰi	(5)	pʰi, kʰi	(5)	tʰa, tʰo, tʰu	(5)
	tj	(5)	pj, kj	(5)	tw	(2)
	tʰj	(5)	pʰj, kʰj	(5)	tʰw	(2)
Total		20		20		14

Six non-words with non-coronal stops followed by a labial-velar glide such as /pw/ were adopted to reduce the imbalance of the stimuli set in which a palatal glide /j/ was used predominantly over a labial-velar glide /w/. In addition, six non-words which were originally created to investigate laryngeal OCP and/or vowel harmony turned out to be useful to investigate palatalisation, and thus were also considered in the analysis as part of the current set of non-words. As a result, among 66 non-words in this set, 20 were target items with TI, 17 control items with TA, 23 control items with PI and 6 fillers with non-coronal stops followed by a labial-velar glide.

Eighty Korean non-words were used to explore the vowel harmony effect. The V_1–V_2 sequences used exhausted all possible sequence permutations of the six vowels, /a e i o u ʌ/, and for each V_1–V_2 sequence, two non-words were created, thus yielding 72 non-words in total (6 vowels × 6 vowels × 2 items), as shown in (21).

(21) *Vowel sequences of target and control non-words in the vowel harmony set (number of non-words for each sequence = 2)*

	Disharmonic target	Harmonic control
	ae ai au aʌ	aa ao
	ea eo	ee ei eu eʌ
	ia io	ie ii iu iʌ
	oe oi ou oʌ	oa oo
	ua uo	ue ui uu uʌ
	ʌa ʌo	ʌe ʌi ʌu ʌʌ
Total	16 sequences (n = 32)	20 sequences (n = 40)

Eight non-words which were originally created to investigate palatalisation turned out to be useful to investigate vowel harmony, and thus were also considered in the analysis as part of the relevant set of non-words. Among 80 non-words in this set, 33 were disharmonic target items and 47 were harmonic control items.

Seventy-five Korean non-words were used to explore the laryngeal OCP effect in Korean phonotactics. The C_1–C_2 onset sequences used exhausted all possible sequence permutations of three consonant phonation types, that is, lenis, aspirated and tense. For each C_1–C_2 phonation type sequence except for lenis–lenis, non-words with eight different V_1–V_2 sequences were created. Eight lenis–lenis non-words had already been created for the other two conditions, thus making up 72 non-words (3 phonations × 3 phonations × 8 vowel sequences), as shown in (22).

(22) *Number of target and control non-words in the laryngeal OCP set by onset sequence. (The number of non-words of the corresponding type is shown in parentheses; asp = aspirated.)*

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

	Target		Control			
	Double-lar		No-lar		Single-lar	
	asp–asp	(8)	lenis–lenis	(8)	lenis–asp	(8)
	asp–tense	(8)			lenis–tense	(8)
	tense–asp	(8)			asp–lenis	(8)
	tense–tense	(8)			tense–lenis	(8)
Total		32		8		32

For C_1 and C_2, we used four lenis obstruents /p t c k/ as well as their aspirated and tense counterparts. Among the non-words created to investigate vowel harmony and/or palatalisation, 11 of them, including eight items with lenis–lenis sequences, met the relevant conditions required for this set of non-words. These items were thus also considered in the analysis as part of this set of non-words. The resulting set consists of 32 target non-word items with two laryngeally marked consonants (henceforth 'double-lar') and 43 control items which can be further divided into eight non-words with no laryngeally marked consonants (henceforth 'no-lar') and 35 non-words with a single laryngeally marked consonant (henceforth 'single-lar').

All the Korean non-words from the three sets were recorded as stimuli in a soundproof booth with a high-quality microphone. One female speaker of Seoul Korean produced all of the stimuli in isolation with declarative intonation. The recorded stimuli were RMS equalised in Praat (Boersma & Weenink 2020) and checked to ensure that they could be heard as intended. Five Seoul Korean speakers were employed to transcribe all recorded audio stimuli in the standard Korean orthography. Non-words which were incorrectly transcribed were re-recorded until all five of the transcribers identified them correctly. The stimuli were pseudo-randomised in order, such that stimuli of the same set never followed each other and were always separated by stimuli of different sets.[14] We controlled the presentation-order effect by using four different pseudo-randomised orders.

3.1.2. Participants

Participants were recruited from the community at Seoul National University through online advertising. They were restricted to native speakers of Seoul Korean with ages of at least 18 years. Well-formedness ratings to the test stimuli and demographic information of participants were collected through an online Google Forms survey. All participants (n = 57) were compensated with online gift cards for their participation.

Data from participants who reported themselves as being non-native speakers of Seoul Korean or who did not agree to the IRB statement were excluded from the analysis. In addition, data from participants who chose the same response option for two-thirds or more of the stimuli were excluded from the analysis. The final set of data included responses from 49 participants (38 female and 11 male; mean age = 25.37).

3.1.3. Procedure

The participants were instructed to rate the stimuli, on a 1–6 Likert scale, according to how acceptable they sounded as standard Seoul Korean words, not loanwords.[15] A rating of 1 meant 'unlikely as a Korean word', whereas 6 meant 'likely as a Korean word'. The stimuli were provided in the form of a sound file, which was hosted on YouTube and linked to each question. The instruction did not include any statement about morphological composition of the stimuli, under the expectation that the

[14] This did not include stimuli designed for other purposes but considered in the analysis as part of the same set. Therefore, we did not perfectly avoid the consecutive presentation of stimuli belonging to the same set.

[15] Loanwords show some prominent phonotactic differences from native and Sino-Korean words – for example, word-initial liquids are allowed only in loanwords.

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

participants would view the stimuli spoken in isolation as monomorphemic, based on their disyllabic length, as mentioned in footnote 13. The participants proceeded through the experiment at their own pace. The experiment lasted approximately 25 minutes.

3.2. Results

The mean of all the received responses was 3.05 (see the Supplementary Material for the mean response score, shown under 'human rating', for each test non-word). As anticipated, the mean well-formedness ratings of the eight practice items covered a relatively wide range of the scale, as shown in Figure 1.

Three practice items, [neho] (4.0), [kʌta] (3.55) and [pahe] (3.41), which we, as native speakers of Korean, consider perfectly well-formed, received clearly higher scores than three practice items, [kociɨ] (2.45), [les'ʌ] (2.39) and [pot'ja] (1.63), which we consider ill-formed.[16] Excluding the practice items, we analysed participant ratings on the rest of the stimuli. The total number of responses was 9,457 (49 participants × 193 test non-words). Recall that the test non-words were created in three sets to investigate the three effects – palatalisation, vowel harmony and laryngeal OCP – in the lexicon. In the following subsections, we will analyse and discuss the results separately for each set of test non-words.

3.2.1. Palatalisation

This section concerns the analysis of the results pertaining to palatalisation. The participants gave different ratings for non-words with different consonant–vocoid (CV) sequence types, as shown in Figure 2.[17]

Note in Figure 2 that the target non-words with TI (2.45) were rated lower than the control non-words with TA or PI (3.13), as would be expected if the palatalisation constraint affects the participants' phonotactic judgements. In order to provide a statistical analysis of the observed differences, we fitted a linear mixed effects model in a Bayesian framework, using the `brms` package (Bürkner 2018) in R (R Core Team 2022). We chose a Bayesian implementation of the model for a variety of reasons: it

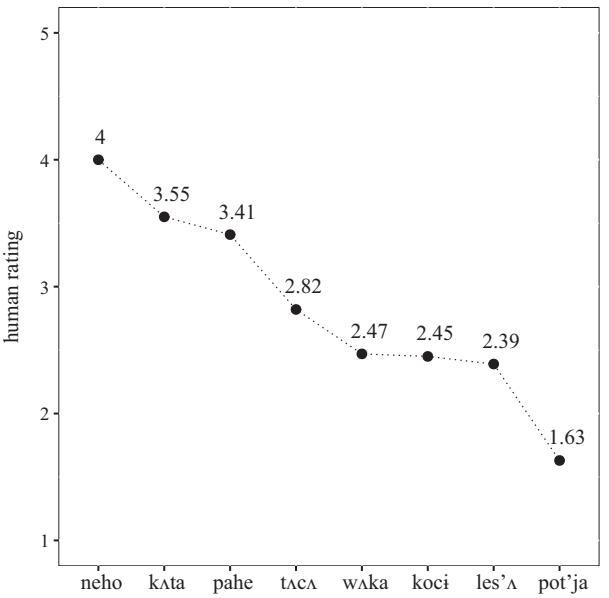

Figure 1. *A plot of ratings of practice items.*

[16]The word-final [ci], the word-initial [l] and the sequence [t'j] do not occur in (at least the surface forms of) native and Sino-Korean words.

[17]Responses to six fillers were excluded from the analysis reported in this section.

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

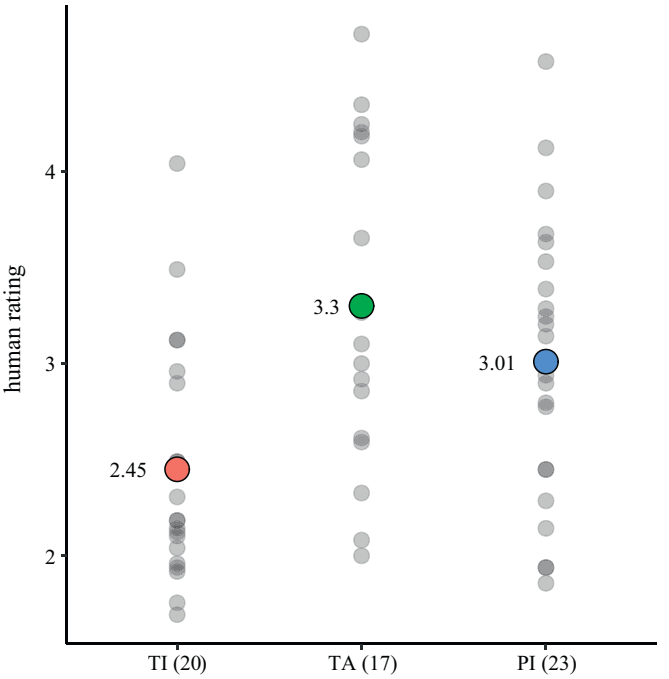

Figure 2. *Palatalisation: Mean Likert rating in the survey of Korean non-words, by CV sequence types. Smaller semitransparent points represent individual non-words, averaged across participants, and larger points represent overall averages. The number of non-words of the corresponding type is shown in parentheses.*

allows us to retrieve quantitative information about a hypothesis of interest, to make direct comparisons between levels of an independent variable, and to fit complex random effects structures, not worrying about convergence issues (Nicenboim & Vasishth 2016; Franke & Roettger 2019). For a hypothesis that a parameter is larger, or smaller, than zero, we will conclude that there is compelling evidence for this hypothesis if the 95% credible interval (CI) for the parameter does not include zero, and/or if the posterior probability of the hypothesis is close to one (Franke & Roettger 2019). The model used the `brms` default priors, with four sampling chains of 2,000 iterations each and a warm-up period of 1,000 iterations. (The same Bayesian setting was used in the statistical model analyses conducted in the rest of the article.) Response ratings and the CV sequence type were included as dependent and independent variables, respectively. The CV sequence type with three levels, TA, PI and TI (baseline), was dummy coded. Following Barr *et al.* (2013), we included the maximal random effect structure justified by the data. Specifically, random intercepts were included for participant and non-word stimulus, and a random slope was included for participant. The results of the Bayesian analysis, presented in (23), show that it is likely true that the target non-words with TI are rated lower than control items with TA and PI, given that 95% CIs for the coefficients TA and PI do not include zero.[18]

(23) *Palatalisation: a Bayesian analysis of a mixed effects linear regression model (fixed effects)*

	Estimate	Est. error	l-95% CI	u-95% CI
(Intercept = TI)	2.45	0.19	2.07	2.83
TA	0.86	0.26	0.35	1.38
PI	0.55	0.24	0.05	1.02

[18]The abbreviations used are as follows: Est. error = estimation error, and l/u-95% CI = lower/upper bound of the 95% CI.

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

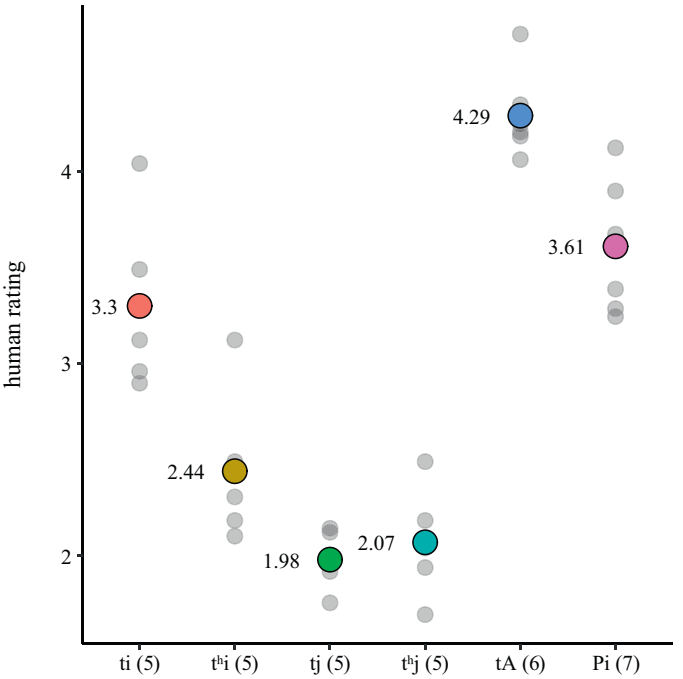

Figure 3. *Palatalisation: Mean Likert rating of target non-words with TI sequences and control non-words with tA and Pi.*

In order to retrieve more quantitative information, we accessed the samples of the fitted model. The sample distributions show that the posterior probabilities of the propositions that TI < TA and that TI < PI are 99.98% and 98.55%, respectively, suggesting that there is compelling evidence for the differences.

In order to have a more accurate understanding of the current ratings pattern, let us consider whether the participants' ratings differ depending on the specific target TI sequence. Four TI sequences, /ti/, /tʰi/, /tj/ and /tʰj/, were used in the target non-words. As can be seen in Figure 3, the mean ratings for three of them, that is, /tʰi/, /tj/ and /tʰj/, are clearly low, below 2.5, and we can consider them to be ill-formed, given that the three ill-formed practice items received scores below 2.5, as shown in Figure 1.

The non-words with /ti/ received relatively higher mean ratings, that is, 3.3, than the rest of the target non-words, but they received relatively lower ratings than their control non-words with tA and Pi. To provide a statistical analysis of the observed differences, we fitted a Bayesian mixed effects linear regression model to the data under consideration. Response ratings and consonant–vocoid sequence type – /ti/ (reference), /tʰi/, /tj/, /tʰj/, /tA/ and /Pi/ – were included as the dependent and independent variables, respectively. The consonant–vocoid sequence type was dummy coded. Random intercepts were included for participant and non-word stimulus, and a random slope was included for participant. The results of the Bayesian analysis, presented in (24), show that it is likely true that the target non-words with /ti/ are rated higher than the rest of target non-words with /tʰi tj tʰj/, given that 95% CIs for the corresponding coefficients do not include zero. The sample distributions also show that the posterior probabilities of the propositions that non-words with /ti/ are higher in ratings than those with /tʰi tj tʰj/ are 100%, suggesting that there is compelling evidence for the differences. These differences can be understood if we consider that aspirated consonants and complex onsets are more marked than lenis consonants and simplex onsets, respectively.

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

(24) *Palatalisation (non-words with TI, tA and Pi sequences): a Bayesian analysis of a mixed-effects linear regression (fixed effects)*

	Estimate	Est. error	l-95% CI	u-95% CI
(Intercept = ti)	3.3	0.18	2.94	3.66
tʰi	−0.86	0.23	−1.32	−0.4
tj	−1.33	0.24	−1.79	−0.85
tʰj	−1.24	0.23	−1.7	−0.77
tA	0.99	0.22	0.56	1.43
Pi	0.3	0.21	−0.11	0.71

Let us now compare /ti/ and its controls, /tA/ and /Pi/. As can be seen in (24), it is likely true that the target non-words with /ti/ are rated lower than those with /tA/, given that the 95% CI for the relevant coefficient, that is, tA, does not include zero. This conclusion is confirmed by the sampling distributions showing that the posterior probability of the proposition that non-words with /ti/ are lower in ratings than those with /tA/ is 100%. On the other hand, it is not clear that the target non-words with /ti/ are rated lower than those with /Pi/, given that the 95% CI for the coefficient Pi includes zero. Also, the sampling distributions show that the posterior probability of the proposition that non-words with /ti/ are lower in ratings than those with /Pi/ is 92.2%. These results suggest that there is some evidence that target items with /ti/ are lower than controls with /Pi/, but the difference is not large enough to be compelling. Thus, it seems that non-words with /ti/ were dispreferred by Korean participants, as suggested by the comparison with controls with /tA/, and that this pre-/i/ dispreference might not be confined to coronal consonants, as suggested by the comparison between targets with /ti/ and controls with /Pi/.

In summary, most of the target TI sequences received clearly low mean ratings, signalling ill-formedness. /ti/, which received the highest ratings among the TI sequences, can still be considered ill-formed, mainly based on its comparison with /tA/ and other clearly well-formed sequences which will be presented in the sections below. These results indicate that the robust alternation patterns of Korean palatalisation are phonotactically supported.

3.2.2. Vowel harmony

If vowel harmony affected the participants' judgements, the disharmonic target non-words would be rated lower than harmonic control ones. This expectation was not met, as can be seen in Figure 4.

Note that the mean response scores were higher for non-words with disharmonic vowel sequences than for those with harmonic vowel sequences. Thus, there is no indication that disharmonic words were less acceptable than harmonic ones. This is confirmed by the results of a Bayesian linear mixed effects model analysis of relevant ratings data. Response ratings and the vowel sequence type were included as dependent and independent variables, respectively. The vowel sequence type with two levels, harmonic and disharmonic (baseline), was dummy coded. Random intercepts were included for participant and non-word stimulus, and a random slope was included for participant. The results of the Bayesian analysis, presented in (25), suggest that there is no vowel harmony effect, given that the coefficient *harmonic* is negative. This is confirmed by the sample distributions showing that the posterior probability that harmonic sequences are higher in ratings than disharmonic ones is only 6.4%.

(25) *Vowel harmony: a Bayesian analysis of mixed-effects linear regression (fixed effects)*

	Estimate	Est. error	l-95% CI	u-95% CI
(Intercept = disharmonic)	4	0.14	3.74	4.27
Harmonic	−0.19	0.13	−0.44	0.05

In summary, disharmonic non-words were not lower in the participants' ratings than harmonic ones. This leads us to conclude that Korean vowel harmony, which is limited in both generality and productivity, is not supported by phonotactics.

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

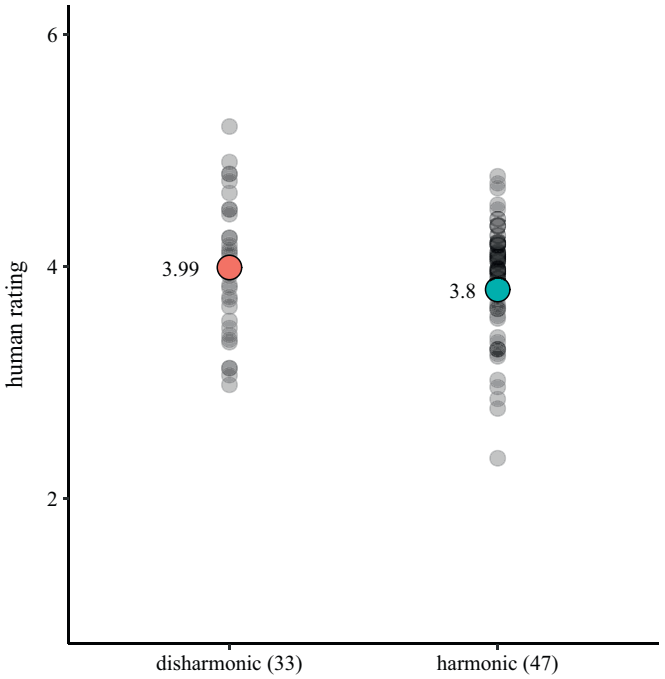

Figure 4. *Vowel harmony: mean Likert rating in the survey of Korean non-words, by vowel sequence type (disharmonic and harmonic).*

3.2.3. Laryngeal OCP

This section discusses the results pertaining to laryngeal OCP. The participants gave different ratings for non-words with different number of laryngeally marked consonants, as shown in Figure 5.

Note in Figure 5 that as the number of laryngeally marked consonants increases, the mean ratings decrease, and that the double-lar non-words were rated the lowest. Most of double-lar items (28/32) received scores below 2.5. It seems clear that they were judged to be ill-formed, given that, as shown in Figure 1, the three ill-formed practice items received scores below 2.5. In contrast, all eight no-lar items received scores above 3.5. In addition, single-lar items received lower mean ratings, which span a wide range of the scale, compared to no-lar items. This suggests that single-lar items may be somewhat ill-formed, possibly violating a constraint that prohibits a (single) laryngeally marked consonant. The key question is whether double-lar items with two laryngeally marked consonants exhibit a superadditive effect, one that exceeds the cumulative effect of two segmental markedness constraints or two instances of the same segmental constraint violation. Such superadditivity is expected under the influence of the Laryngeal OCP which disallows the co-occurrence of two laryngeally marked consonants. (For an in-depth discussion of superadditive effects in phonological data, see Shih 2017.)

In order to determine if there is an effect of laryngeal OCP on participants' judgements, we conducted a Bayesian linear mixed-effects model analysis using relevant ratings data. The dependent variable was the response ratings, while the independent variables included the phonation type (lenis *vs.* laryngeal) of C_1 and C_2, as well as their interaction. It should be noted that the interaction term, which is adopted to capture superadditivity, would be statistically significant if the laryngeal OCP is truly effective. We dummy coded the phonation type with two levels, lenis (baseline) and laryngeal. We included random intercepts for participant and non-word stimulus. All fixed factors were included as by-participant random slopes.

The results of the Bayesian analysis, presented in (26), show that the coefficients for the two fixed factors, the C_1 and C_2 phonation types, are negative, and their 95% CIs do not include zero. This suggests

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

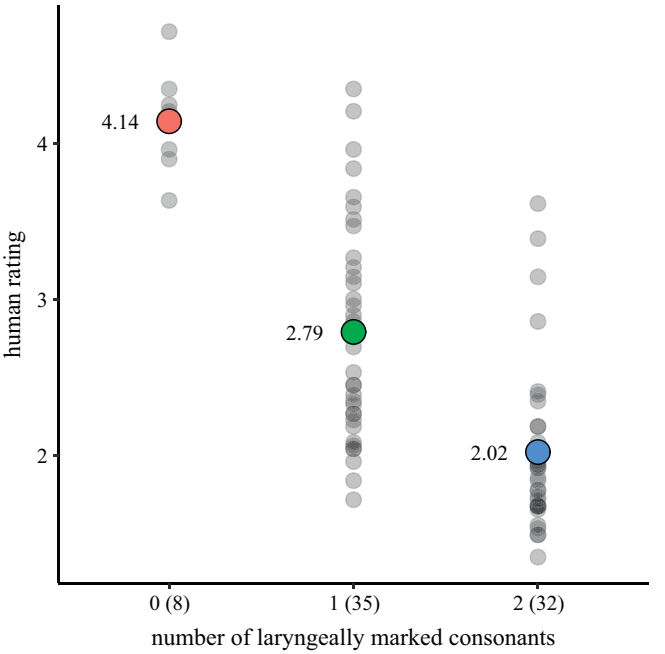

Figure 5. *Laryngeal OCP: Likert rating in the survey of Korean non-words, by number of laryngeally marked consonants.*

that it is likely true that the target non-words are rated lower when either C_1 or C_2 is laryngeally marked compared to when neither is lenis. Thus, the effects of (positional) markedness constraints prohibiting (single) laryngeally marked consonants are confirmed. However, considering that the coefficient for the interaction term is positive – which is the opposite of the prediction under the influence of the Laryngeal OCP – it seems unlikely that a laryngeal OCP effect exists in the ratings data. This conclusion is supported by the sample distributions, which show that the posterior probability of the interaction being below 0 is a mere 3.2%.

(26) *Laryngeal OCP: a Bayesian analysis of Mixed effects linear regression (fixed effects)*

	Estimate	Est. error	l-95% CI	u-95% CI
(Intercept = lenis)	4.15	0.24	3.69	4.62
C_1 laryngeal	−1.49	0.28	−2.05	−0.95
C_2 laryngeal	−1.25	0.28	−1.8	−0.72
C_1 laryngeal : C_2 laryngeal	0.62	0.34	−0.03	1.28

For a more accurate understanding of the current ratings data, let us consider whether the participants' ratings differ depending on the specific phonation type of an onset consonant. As shown in Figure 6, all three different types of double-lar items with (i) tense, (ii) aspirated and (iii) tense and aspirated (henceforth tense//asp) onset sequences were lower in ratings than single-lar non-words with an aspirated onset (henceforth single-asp) which were in turn lower than no-lar items. This is consistent with the response pattern shown in Figure 5. However, single-lar items with a tense onset (henceforth single-tense, mean rating = 2.29) were almost as bad as double-lar items, especially items with two aspirated onsets (henceforth double-asp, mean rating = 2.22).

In summary, the participants' ratings patterns in Figure 6 indicate the following hierarchy: no-lar > single-asp > single-tense and double-lar.

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

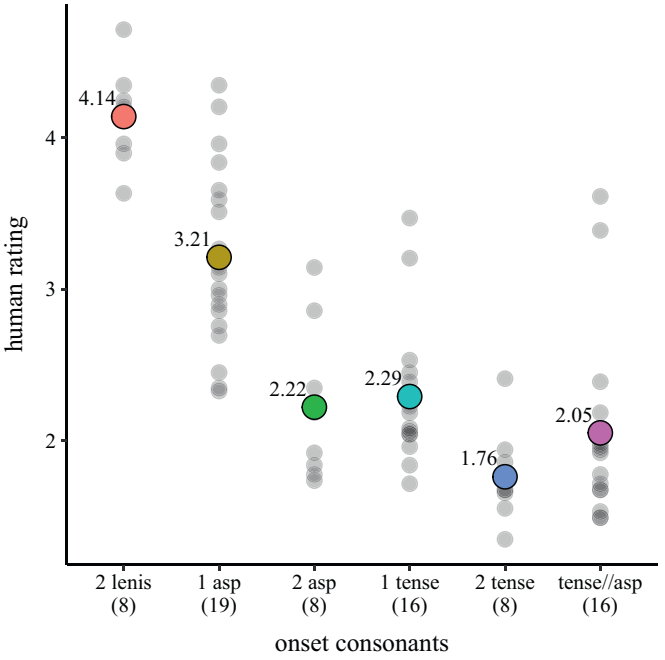

Figure 6. *Laryngeal OCP: Likert rating in the survey of Korean non-words, by onset type.*

To check whether there is a laryngeal OCP effect, we focus on the response results for double-lar items with at least one tense onset, that is, double-tense and tense//asp, which are parallel in onset combination with compounds where blocking of tensing by the laryngeal OCP may occur. If double-tense items are subject to laryngeal OCP, the reduction in ratings from single-tense to double-tense items would be larger than the reduction from no-lar to single-tense items. This predicted superadditive effect did not occur, as can be seen in (27).

(27) *Double-tense* vs. *single-tense*

	Mean Likert rating	Reduced by…
No-lar	4.14	
Single-tense	2.29	1.85
Double-tense	1.76	0.53

If tense//asp items are subject to laryngeal OCP, the reduction from no-lar items to tense//asp items would be larger than the sum of the reductions from no-lar to single-tense items and to single-asp items. As above, the predicted superadditivity effect did not emerge, as can be seen in (28).

(28) *Tense//asp* vs. *single-tense* + *single-asp*

	Mean Likert rating	Reduced by… (relative to no-lar)
No-lar	4.14	
Single-tense, single-asp	2.29, 3.21	1.85, 0.93 (sum = 2.78)
Tense//asp	2.05	2.09

All the above comparisons show subadditivity, not superadditivity, which is not consistent with the prediction under the effect of laryngeal OCP. This, along with the results of the Bayesian mixed effects

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

linear regression analysis shown in (26), indicates that the laryngeal OCP did not affect the participants' phonotactic judgements.

However, before we conclude that there is no laryngeal OCP effect in Korean speakers' phonotactic judgements, note in Figure 6 that the ratings assigned to single-tense items are already quite low, and the ratings of double-lar items can hardly be lowered further. Floor effects of this type might prevent us from finding the superadditive effect. Consequently, based on the results presented in this section, we do not know for sure whether the laryngeal OCP in Korean is phonotactically supported or not.

3.2.4. Summary

Mean participant ratings of the target and control items of each of the three sets of non-words are provided in (29).

(29) *Mean Likert rating in the survey of Korean non-words, by the type of effect and the stimulus group. (The number of non-words of corresponding type is shown in parentheses)*

	Target	Control	Phonotactic effect?
Palatalisation	2.45 (20)	3.13 (40)	Yes
Vowel harmony	3.99 (33)	3.80 (47)	No
Laryngeal OCP	2.02 (32)	3.04 (43)	Unclear

In the set for palatalisation, mean ratings are lower for target items than for control items. Patterns vary depending on the type of TI sequences. Specifically, non-words with /ti/ received higher ratings than the rest of target items, but Korean participants did not judge them to be well-formed. Target non-words with /ti/ (3.3) were clearly lower in mean ratings than their control items with /tA/ (4.29), as well as no-lar items in the laryngeal OCP set (4.14) and non-words in the vowel harmony set, whether disharmonic (3.99) or harmonic (3.8), all of which can be plausibly assumed to be well-formed. Thus, it seems that Korean participants considered all different types of target non-words in the palatalisation set to be ill-formed. This leads us to conclude that Korean cross-morpheme palatalisation with high generality and productivity has strong phonotactic support.

In contrast, in the set for vowel harmony, target items with disharmonic sequences received slightly higher ratings than control items with harmonic sequences, which is the opposite of the prediction under the hypothesis that vowel harmony holds in Korean speakers' phonotactic judgements. The two sets, disharmonic (3.99) and harmonic (3.8), were close in mean ratings to no-lar items of the laryngeal OCP set (4.14), but much higher than single-lar (2.79) and double-lar (2.02) items. It seems that the set of non-words created to test Korean vowel harmony, whether disharmonic or harmonic, were judged to be well-formed. Thus, we may conclude that Korean vowel harmony in verbal inflection with limited generality and productivity has no phonotactic support.

Finally, in the set for laryngeal OCP, the superadditive effect, expected under the effect of laryngeal OCP, did not arise. However, even single-lar control items with a tense onset were very low in the adopted acceptability scale, and thus it was almost impossible to capture the superadditive effect. Consequently, the current response data are unclear about whether the laryngeal OCP active in compound tensing with limited generality and moderate productivity is phonotactically supported or not.

4. Computational modelling

In this section, to test our research hypothesis presented in (1), we measure the lexical support for the three alternation patterns in Korean by conducting a simulation of learning a MaxEnt grammar (Goldwater & Johnson 2003; Hayes & Wilson 2008). Specifically, we train the UCLA Phonotactic Learner (Hayes & Wilson 2008) on a Korean lexicon. Given that observed/expected (O/E) values, used in some previous studies (Pierrehumbert 1993; Frisch *et al.* 2004 and others) for examining lexical

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

statistics, are mathematically flawed (Wilson & Obdeyn 2009), it seems that the weights of constraints in the learned MaxEnt grammar are the only reliable indicator of over- or underrepresentation of a lexical pattern currently available. It is predicted from the hypothesis in (1) that the more robust an alternation pattern is, the more likely its associated phonotactic constraint is to be learned and the higher its weight will be.

The learned grammar with weighted constraints will assign well-formedness ratings to the non-words adopted in the acceptability-rating experiment. The resulting model ratings will be used in testing a prediction from our research hypothesis. More specifically, model ratings for target non-words violating a phonotactic constraint will be lower than those for control non-words satisfying it only when the associated alternation is robust.

In addition, the model ratings will be compared to the participants' ratings (which will be referred to as 'human ratings') presented in the previous section to figure out whether and how much the participants' phonotactic judgements are based on over- or underrepresentation of a pattern in the lexicon. It is often assumed in the phonological literature that human speakers make phonotactic judgements on the basis of their phonotactic grammar, which in turn reflects lexical frequency (Hayes & Wilson 2008; Albright 2009; Daland *et al.* 2011; Chong 2017, and others). If it is true that Korean participants responded using their phonotactic grammars that are constructed based on lexical representation, the model ratings should correlate well with the participants' ratings.

In the remainder of this section, the method of the modelling is first described, and then the results of the modelling are reported and analysed to test our main research hypothesis in (1) and its predictions.

4.1. Method

In order to construct a Seoul Korean lexicon, we first selected words listed as standard Korean in the Standard Korean dictionary (henceforth the SKD, National Institute of the Korean Language 1999; downloaded from https://stdict.korean.go.kr/mypage/download/downloadList.do on 17 September 2020). Among them, we further selected words with a frequency of at least one occurrence in the Sejong corpus. We extracted the citation-form pronunciations of the selected words from the SKD. In addition, we determined their morphological composition (simplex/complex) and etymological origin (native/Sino-Korean/loan) mainly based on the SKD. We prepared three monomorphemic word lexicons, not weighted by token frequency: (i) one with native Korean words ($n = 5,335$), (ii) one with Sino-Korean words ($n = 32,702$) and (iii) one with both native and Sino-Korean words ($n = 38,037$).[19] Each of these was used as training data for modelling phonotactic grammars.

We conducted a simulation of learning a MaxEnt grammar by using UCLA phonotactic learner (Hayes & Wilson 2008). The feature system fed to the learner was constructed mainly based on Cho (2012). Aspirated and tense consonants had different values for the feature [aspirated], but the two were specified as [+tense], thus forming a natural class. The feature [+spread glottis] was assigned only to /h/. Departing from Cho, we distinguished between onset and coda consonants by using the feature [rhyme], as in Daland *et al.* (2011) and Park (2020). For the vowels, we adopted a nine-vowel system, /i e a ʌ o u ɨ y ø/, not distinguishing between /e/ and /ɛ/, which are known in the literature to be no longer contrastive, at least in younger Korean speakers' speech. Although the front rounded vowels, /y ø/, are known to be diphthongised often, it is still possible that they are pronounced as monophthongs at least in some words, and thus we decided to adopt a conservative choice by distinguishing them in the vowel inventory. As recently discussed by Jun (2018), it is quite controversial what the feature involved in Korean vowel harmony is. A variety of features, such as backness, lowness, ATR and RTR, have been proposed. Given this indeterminacy, we specified /a o/ as [+RTR] and the rest of the Korean vowel phonemes as [−RTR], but make no theoretical commitment to this identification of the harmonising feature.

[19] Note that previous modelling studies on phonotactics provided more reliable predictions based on type than token frequencies (Hayes & Wilson 2008: 395; Albright 2009: 20; and the references therein).

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

The learner settings were as follows. The length of constraint (i.e., the maximum gram size), the maximum number of constraints to discover and the highest O/E threshold were set at 2, 300 and 0.3, respectively.[20] Constraints including complement natural classes (designated by the symbol ˆ) were allowed to be found. Autosegmental tiers, called 'projections', were created and fed to the learner so that constraints for nonlocal processes like vowel harmony and laryngeal OCP could be found. Specifically, constraints can be defined on three different projections: (i) default segmental projection, that is, full representation; (ii) vowel projection, that is, the substring consisting of all and only vowels; and (iii) laryngeal projection, that is, the substring of all and only obstruent onsets. Palatalisation, vowel harmony and laryngeal OCP constraints were expected to be found from default, vowel and laryngeal projections, respectively. Finally, no prespecified constraints were fed to the learner.

4.2. Simulation results

The UCLA Phonotactic Learner creates phonotactic constraints through an inductive learning algorithm and assigns numerical weights to them according to the principle of maximum entropy (see Hayes & Wilson 2008 for relevant details). If a form is underrepresented in the training data, a constraint prohibiting it will likely be learned. The more underrepresented a form is, the higher the weight of the constraint prohibiting it. To illustrate, some high-weighted constraints learned from the lexicon of native and Sino-Korean monomorphemes are shown in (30).

(30) *Two high-weighted constraints learned from the lexicon of native and Sino-Korean monomorphemes (son = sonorant, rh = rhyme, cons = consonantal)*

	Constraint	Weight	Comment
a.	*[+son, +dorsal, −rh]	6.605	No [ŋ] in the onset
b.	*[−rh][+cons]	7.682	No onset consonant clusters

The two constraints in (30) reflect well-known phonological restrictions in Korean. In Korean, the velar nasal and true consonant clusters are not allowed at all in the onset.

The learner gave penalty scores to each test non-word depending on the weight of constraints violated by it and the number of violations, as defined by Hayes & Wilson (2008: 383, (4)). Note that such penalty scores (often called harmony scores) are known to match with human well-formedness intuitions (Breiss & Hayes 2020: 362). For model predictions about acceptability of a non-word which will be compared to its human ratings, we used negative values of the penalty scores assigned to it which increase as acceptability increases like human ratings. Henceforth, the negative penalty scores will be referred to as 'model ratings'. Since we trained the learner on three different lexicons, we were provided with three sets of model ratings.

Which lexicon and which set of model ratings should we rely on? We opt for the lexicon consisting of native and Sino-Korean monomorphemes, based on a plausible assumption, adopted by Chong (2019), that ordinary Korean speakers cannot reliably distinguish between native and Sino-Korean words. Still, since we were curious about whether and how the three lexicons differ in explaining the human ratings, we calculated the correlations (Spearman) between human ratings and each of the three sets of model ratings.[21] As can be seen in (31), the ratings given by the MaxEnt trained on both native and Sino-Korean words show the highest correlation with the human ratings although there is only a small difference between the lexicons consisting of only Sino-Korean words and both native and Sino-Korean words, which is not surprising if we consider the fact that the latter lexicon consists mostly of Sino-Korean words (85.97%, i.e., 32,702 out of 38,037).

[20] The O/E is the ratio of the number of observed violations of a constraint, to the expected number of violations. Low O/E values indicate powerful, effective constraints. For details, see Capodieci *et al.* (2009, §5.4) and Hayes & Wilson (2008, §4.2).

[21] We used nonparametric Spearman correlations because many of the data points were concentrated at the ends of the scale.

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

(31) *Correlations (Spearman) between human ratings and model ratings (provided by MaxEnt models trained on three Korean monomorpheme lexicons)*

Native	Sino-Korean	Native and Sino-Korean
0.53	0.71	0.73

In the remainder of this article, the set of native and Sino-Korean monomorphemes in our Seoul Korean corpus will represent the Seoul Korean lexicon, and the ratings given by a MaxEnt model trained on it will be referred to as 'model ratings'.

As shown under 'native and Sino-Korean' in (31), the correlation between human and model ratings (0.73) is high. This high correlation between human and model ratings indicates that the Korean participants' well-formedness judgements were heavily based on the frequency distribution of the Korean lexicon, confirming the assumption that speakers' judgements are a good proxy for lexical phonotactics. In Figure 7, the human ratings for the entire test set of non-words, are plotted against the model ratings. (Human and model ratings for all test non-words are provided in the Supplementary Material.)

The scatter plot in Figure 7 shows an overall trend that the human ratings increase as the model ratings increase, as is consistent with the high positive correlation between the human and model ratings, mentioned above.

Let us now consider how the correlations between human and model ratings differ depending on the stimuli. For each set of test stimuli, the correlation (Spearman) between observed human ratings, presented in the previous section, and the corresponding model ratings were calculated as shown in (32).

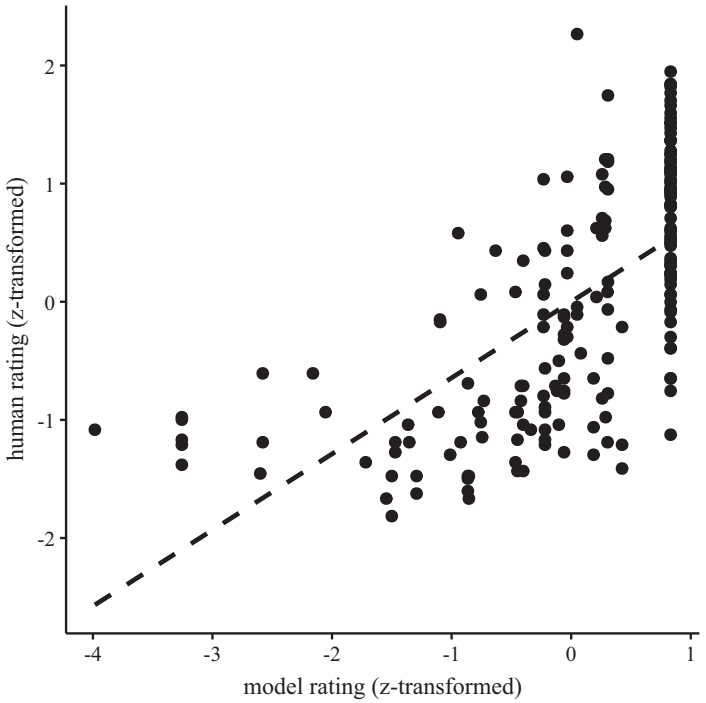

Figure 7. *All test non-words: Correlation between human and model ratings (z-transformed; n* = 193*; dashed line = regression line).*

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

(32) *Correlations (Spearman) between human and model ratings for each of the three sets of test non-words*

Palatalisation (n = 66)	Vowel harmony (n = 80)	Laryngeal OCP (n = 75)
0.58	0.20	0.74

The correlation between human and model ratings is high for the non-words in the laryngeal OCP set, and moderate for the non-words in the palatalisation set. However, the correlation for non-words in the set of vowel harmony (0.20) is low. As will be shown below, the main reason for this relatively low correlation is that many of the non-words in the vowel harmony set were predicted to be rated as almost perfect, and thus the model rating values did not vary sufficiently, leaving little room for correlations with human ratings. In the following subsections, we discuss model ratings for each set of test non-words, testing the predictions from our research hypothesis in (1), and comparing them with the corresponding human ratings.

4.2.1. Palatalisation

In this section, we report and discuss the modelling results pertaining to palatalisation to test the specific predictions, shown in (33), which can be made from the hypothesis that Korean palatalisation, which is general and highly productive across morphemes, has strong lexical support.

(33) *Predictions about the results of the MaxEnt simulation, under the hypothesis that Korean palatalisation has strong lexical support.*

a. The palatalisation constraint *TI is learned in the MaxEnt simulation and its weight is high.

b. Target non-words violating *TI are lower in model ratings than control non-words satisfying *TI.

As shown in (34), the first prediction in (33a) seems to be confirmed since a general palatalisation constraint prohibiting TI sequences was learned in the current simulation, and its weight (3.286) is relatively high.

(34) *Palatalisation constraint learned from the MaxEnt simulation*

Constraint	Weight	Comment
*[−strident][+high,−back]	3.286	No alveolar stops before high front vocoids

The weight 3.286 is not as high as the weights of constraints reflecting exceptionless generalisations in Korean, for example, no velar nasal in the onset (6.605) and no true onset consonant clusters (7.682), as shown in (30). However, as will be shown below, it is much higher than the weights of any constraints with the effect of vowel harmony or laryngeal OCP.

The second prediction in (33b) is also confirmed by the model ratings in the violin plot in Figure 8.

Note that target non-words with TI sequences are clearly lower in mean ratings than control non-words with TA and PI sequences. Recall that the observed lower ratings of non-words with TI sequences were also true for human ratings. The specific word-by-word correspondence can be seen in Figure 9, in which the human ratings for the palatalisation set are plotted against the model ratings.

Note that about half of the non-words with TI sequences, that is, target items (marked with filled circle) are located in the lower left quadrant of the plot, indicating that they are both predicted and judged to be ill-formed. Most of the control items with TA or PI sequences (marked with empty triangle and circle, respectively) are both predicted and judged to be better formed than the target items.

For a more accurate understanding of the current model ratings pattern, let us consider whether the model ratings differ depending on the specific target TI sequence. Recall that four TI sequences, /ti/, /t^{h}i/, /tj/ and /t^{h}j/, were used in the target non-words. As can be seen in Figure 10, non-words with no glide (/ti t^{h}i/) received higher ratings than those with a glide (/tj t^{h}j/). Among target non-words with no

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

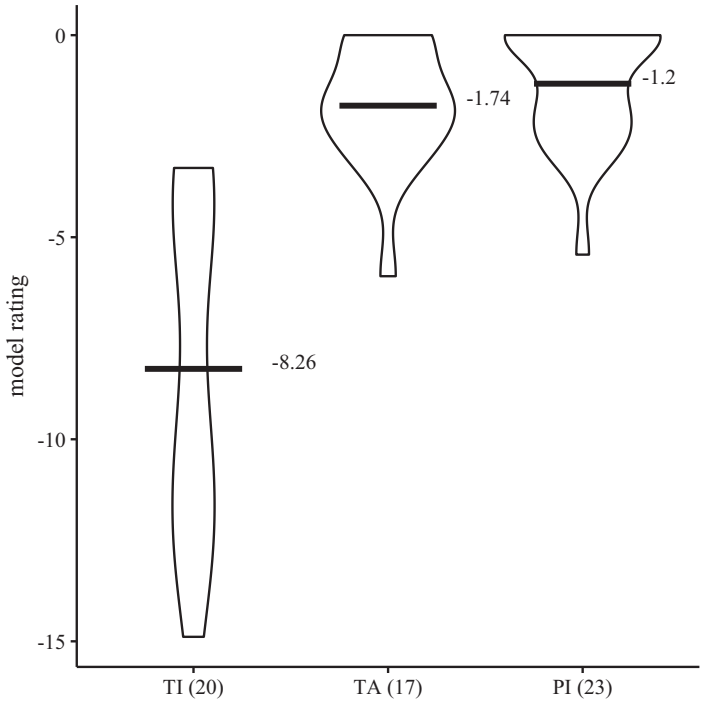

Figure 8. *Palatalisation: Model ratings of non-words, by CV sequence type. Crossbars represent mean ratings.*

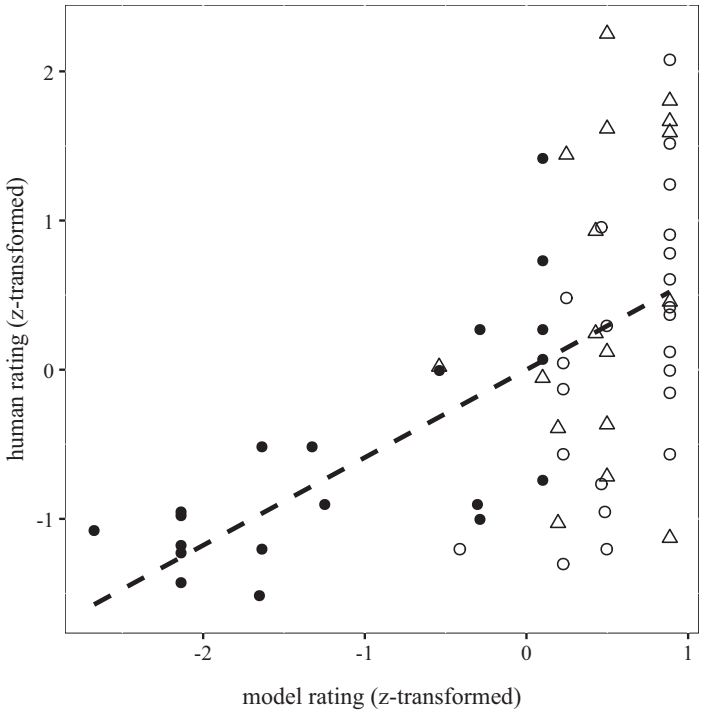

Figure 9. *Palatalisation: Correlation between human and model ratings (n = 66; dashed line = regression line; filled circle = target with TI, empty triangle = control with TA, empty circle = control with PI).*

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

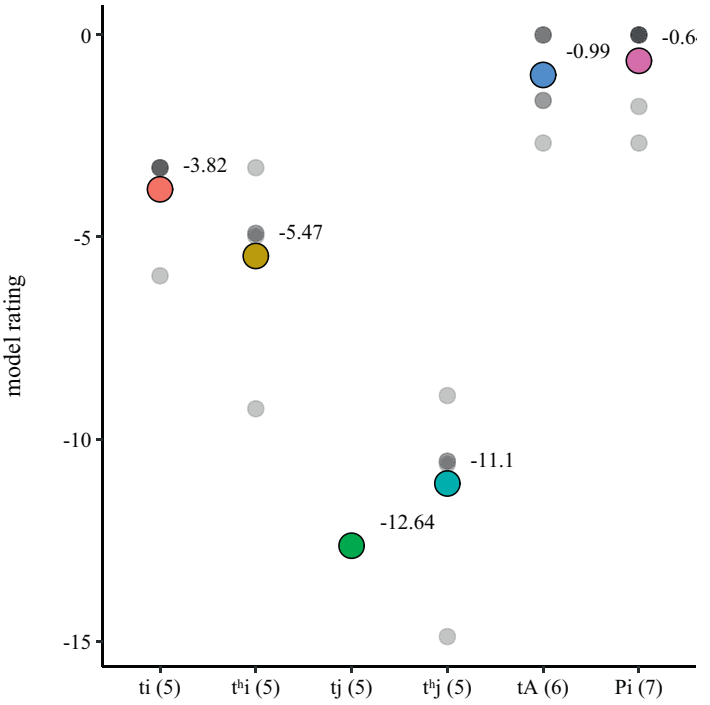

Figure 10. *Palatalisation: Mean model rating of target non-words with TI sequences and control non-words with tA and Pi. Smaller semitransparent points represent individual non-words, and larger points represent overall averages.*

glide, those with a lenis onset (/ti/) received higher ratings than those with an aspirated onset (/tʰi/). Non-words with /ti/ were lower in mean ratings than control non-words with /tA/ and /Pi/.

To provide a statistical analysis of the observed differences, we fitted a Bayesian linear regression model to the current model ratings data. Model ratings and consonant–vocoid sequence type – /ti/ (reference), /tʰi/, /tj/, /tʰj/, /tA/ and /Pi/ – were included as the dependent and independent variables, respectively. The consonant–vocoid sequence type was dummy coded. The results of the Bayesian analysis, presented in (35), suggest that it is likely true that the target non-words with /ti/ are rated higher than target non-words with /tj tʰj/, and that the target non-words with /ti/ were rated lower than control non-words with /tA Pi/, given that 95% CIs for the corresponding coefficients do not include zero. The sample distributions confirm all these conclusions, and, in addition, show that the posterior probability of the proposition that non-words with /ti/ are higher than those with /tʰi/ is 95.38%, suggesting that we can be 95.38% certain that items with /ti/ sequences are more well-formed than those with /tʰi/.

(35) *Palatalisation (non-words with TI, tA and Pi sequences): a Bayesian linear regression model*

	Estimate	Est. error	l-95% CI	u-95% CI
(Intercept = ti)	−3.79	0.69	−5.21	−2.45
tʰi	−1.67	0.98	−3.59	0.3
tj	−8.84	0.97	−10.7	−6.99
tʰj	−7.3	0.99	−9.18	−5.37
tA	2.81	0.96	0.91	4.71
Pi	3.17	0.9	1.39	5

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

The most important observation here is that model ratings of target non-words with /ti/, as well as other TI, sequences are meaningfully lower, compared to control non-words with /tA Pi/, confirming the prediction in (33b). This is in part consistent with the corresponding results of an acceptability-rating experiment, presented in §3.2.1. Recall that there is compelling evidence that the participants gave lower ratings to the non-words with /ti/ relative to /tA/, not /Pi/. Thus, it seems that the lexical support for palatalisation is even stronger in the MaxEnt simulation than in speakers' phonotactic judgements. An additional discrepancy between model and human ratings for non-words in the palatalisation set can be seen; the difference between non-words with /tʰi/ *vs.* /tj tʰj/ is larger in model ratings than in human ratings. This might be in part due to a floor effect in which the human ratings were already low for non-words with /tʰi/ (2.44), and so the participants could not give much lower ratings for non-words with /tj tʰj/.

In conclusion, the two predictions in (33) are confirmed by the results of the current simulation, indicating that lexical items with TI sequences in the Korean lexicon are underrepresented, and thus that general and productive cross-morphemic palatalisation in Korean has strong lexical support.

4.2.2. Vowel harmony

In this section, we report and discuss the modelling results pertaining to vowel harmony to test the specific predictions presented in (36) which can be made from the hypothesis that Korean vowel harmony with limited generality and productivity has no, or weak, lexical support.

(36) *Predictions about the results of the MaxEnt simulation, under the hypothesis that Korean vowel harmony has no, or weak, lexical support*

a. Vowel harmony constraints are not learned at all, or are learned with low weights.

b. Target non-words with disharmonic vowel sequences are *not* lower in model ratings than control non-words with harmonic vowel sequences.

To test the prediction in (36a), we investigated the vowel projection constraints learned in the current simulation. All vowel projection constraints learned either were trivial (such as *## 'no word without a vowel') or else involved front rounded vowels [y ø] or the high back unrounded vowel [ɨ], all of which have highly limited distributions in Korean. Note that none of these vowels were used in the current set of non-words. What is important here is that no vowel harmony constraints prohibiting disharmonic vowel sequences were learned, confirming the prediction in (36a). This indicates that disharmonic vowel sequences are not underrepresented in the Korean lexicon.

To test the second prediction in (36b), we investigated the distribution of model ratings of the non-words in the vowel harmony set. As can be seen in Figure 11, the mean model ratings are slightly higher for disharmonic non-words than for harmonic ones, thus confirming the prediction in (36b).[22]

Recall from §3.2.2 that higher scores for disharmonic words were also found in the human ratings. The specific word-by-word correspondence can be seen in Figure 12, in which the human ratings for the vowel harmony set are plotted against the model ratings.

Note in Figure 12 that many items, either disharmonic (filled circle) or harmonic (empty circle), were located near the high end of the model rating scale. Specifically, as shown in (37), the majority of the test items in the vowel harmony set (68 out of 80 items) were predicted to be perfect (i.e., 0) or near-perfect (i.e., −1.62 and −1.69). The limited variation due to the dominance of such (almost) perfectly well-formed non-words can explain why the correlation between human and model ratings was relatively low (0.20) for the non-words in the vowel harmony set, presented in (32).

[22] We performed a Bayesian linear regression analysis for the relevant model ratings data. The sample distributions show that the posterior probability of the proposition that disharmonic non-words are lower in model ratings than harmonic non-words is only 1.75%.

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

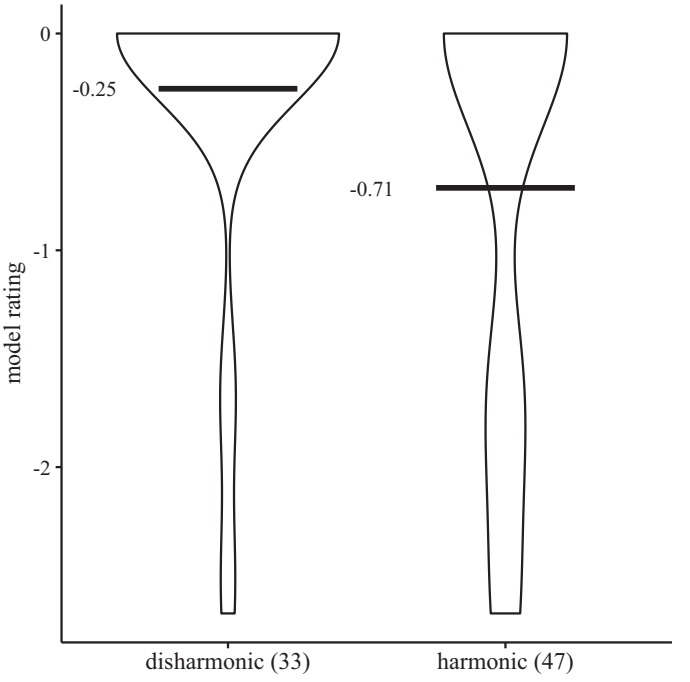

Figure 11. *Vowel harmony: Model ratings of Korean non-words, by the type of vowel sequence (disharmonic vs. harmonic). Crossbars represent mean ratings.*

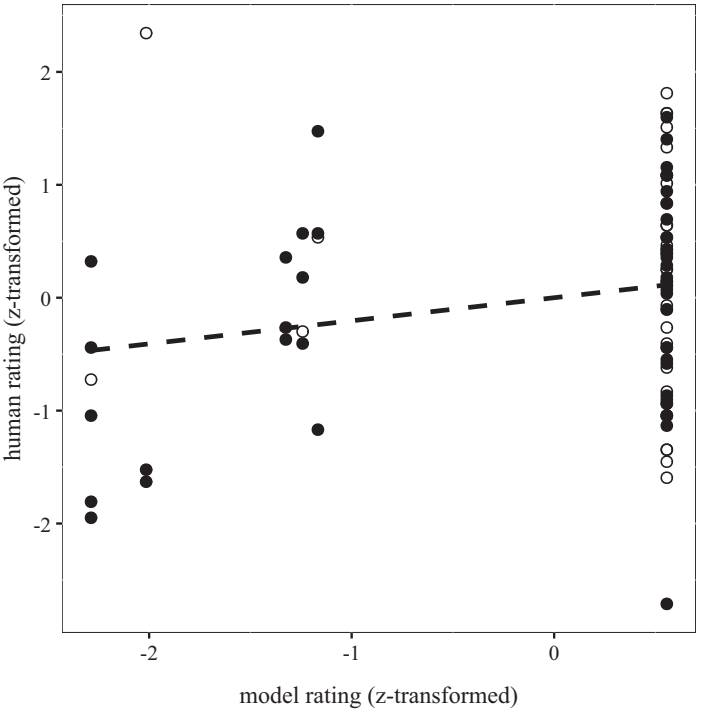

Figure 12. *Vowel harmony: Correlation between human and model ratings (n = 80; dashed line = regression line; filled circle = disharmonic, empty circle = harmonic).*

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

(37) *Number of non-words in the vowel harmony set by model rating*

Model rating	Total	(Disharmonic, Harmonic)
0	60	(29, 31)
−1.62	4	(1, 3)
−1.69	4	(1, 3)
−1.77	3	(0, 3)
−2.42	3	(1, 2)
−2.68	6	(1, 5)
	80	(33, 47)

In conclusion, the two predictions in (36) are confirmed by the results of the current simulation, indicating that lexical items with disharmonic vowel sequences in the Korean lexicon are not underrepresented, and thus that Korean vowel harmony with limited generality and productivity has no lexical support.

4.2.3. Laryngeal OCP

In this section, we report and discuss the modelling results pertaining to laryngeal OCP to test the specific predictions presented in (38) which can be made from the hypothesis that laryngeal OCP in Korean with limited generality and productivity has no, or weak, lexical support.

(38) *Predictions about the results of the MaxEnt simulation, under the hypothesis that laryngeal OCP in Korean has no, or weak, lexical support.*

a. The laryngeal OCP constraints are not learned at all, or are learned with low weights.
b. Target non-words violating laryngeal OCP are *not* lower in model ratings than control non-words satisfying it.

To test the first prediction in (38a), we investigated the laryngeal projection constraints learned in the current simulation. No general laryngeal OCP constraint (i.e., *[+tense][+tense]) was learned. However, two of the three constraints learned on the laryngeal projection, (39a) and (39b), express more specific laryngeal OCP generalisations. In (39a) *[+tense][+aspirated], with a weight of 1.256, penalises aspirated–aspirated and tense–aspirated onset sequences, and in (39b) *[+aspirated][−aspirated,+tense], with a weight of 0.564, penalises aspirated–tense onset sequences. No constraint prohibiting tense–tense sequences was learned on the laryngeal projection. In summary, relatively low-weighted laryngeal OCP constraints were learned only for some potential target sequences, which we may consider to be consistent, at least in part, with the prediction in (38a). The set of learned laryngeal OCP constraints and their weights suggest that only some of onset sequences violating the laryngeal OCP are weakly underrepresented and the rest are not underrepresented in the Korean lexicon. This may indicate that laryngeal OCP in Korean compounding has weak lexical support.

(39) *Laryngeal projection constraints learned from the MaxEnt simulation (asp = aspirated)*

	Constraint	Weight	Comment
a.	*[+tense][+asp]	1.256	No laryngeally marked onsets before aspirated onsets
b.	*[+asp][−asp,+tense]	0.564	No aspirated onsets before tense onsets
c.	*[−asp,+tense][segment]	1.993	No tense onsets in non-final syllable

To test the second prediction in (38b), we investigated the distribution of model ratings of the non-words in the laryngeal OCP set. The model ratings differ for non-words with different numbers of laryngeally marked consonants, as shown in Figure 13.

Note in Figure 13 that as the number of laryngeally marked consonants increases, the model ratings decrease, with the double-lar non-words being predicted to have the lowest ratings. Recall from §3.2.3

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

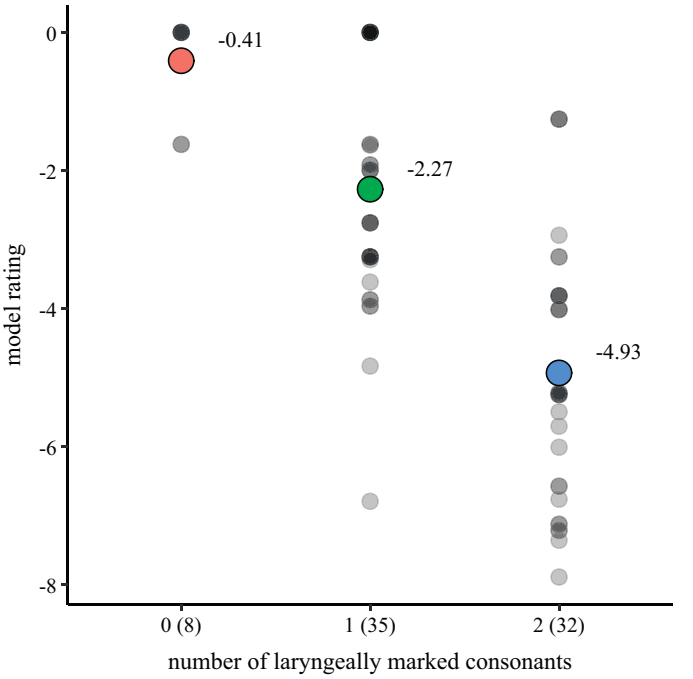

Figure 13. *Laryngeal OCP: Model rating in the MaxEnt simulation by number of laryngeally marked consonants.*

that the observed hierarchy, no-lar > single-lar > double-lar, was also true for human ratings. The more specific word-by-word correspondence can be seen in Figure 14, in which the human ratings for the laryngeal OCP set are plotted against the model ratings.

Note that many of the double-lar items, that is, target items (marked with black circles), are located in the lower left quadrant of the plot, indicating that they are both predicted and judged to be ill-formed. Among control items, all eight no-lar items (marked with empty circles) are located at the upper right quadrant of the plot, indicating that they are both predicted and judged to be well-formed, whereas single-lar items (marked with grey circles) show an intermediate pattern between double and no-lar items, covering a relatively wide range.

In order to have a more accurate and better understanding of the ratings pattern under consideration, let us consider how the patterns differ depending on the specific phonation type of an onset consonant. From Figure 15, in which model ratings of test non-words are plotted by the onset type, we can make some observations. First, single-lar items are not as good as no-lar items, as we have seen in Figure 13. Second, single-asp items are higher in model ratings than double-asp (and all other double-lar) items. Third, single-tense items are higher in model ratings than double-tense (and tense//asp) items, but lower than double-asp items. Thus, having one tense onset is already quite bad in the simulated grammar, as is also the case in the human judgements presented in §3.2.3. As we can see by comparing Figure 6 and Figure 15, one notable difference between human and model ratings is that the differences between single-tense *vs.* double-tense and tense//asp items are smaller in human ratings than in model ratings, which can be attributed to a floor effect.

As we have done in §3.2.3, let us now check whether there is a superadditive effect, focusing on the modelling results for double-lar items with at least one tense onset, that is, double-tense and tense//asp. If double-tense items are subject to laryngeal OCP, the reduction in ratings from single-tense to double-tense items would be larger than the reduction from no-lar to single-tense items. This predicted superadditivity effect did not occur, as can be seen in (40).

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

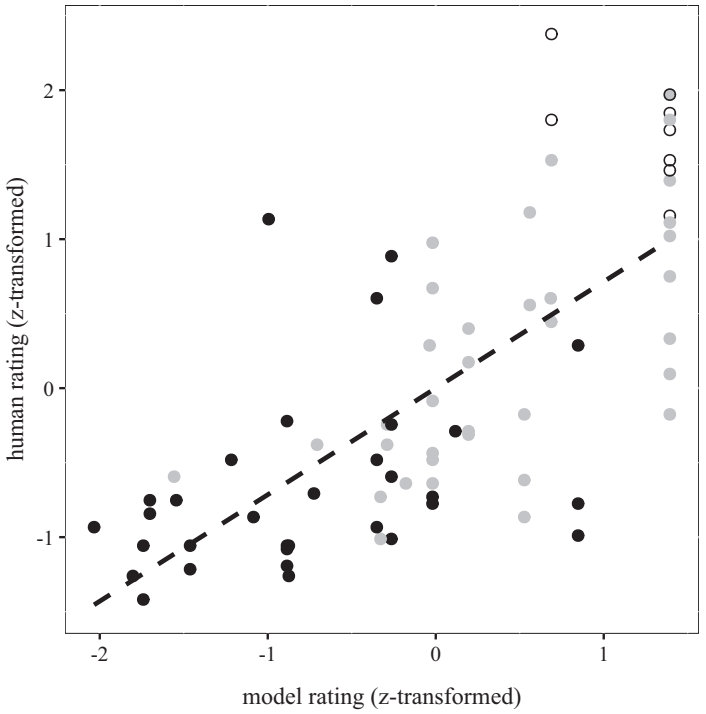

Figure 14. *Laryngeal OCP: Correlation between human and model ratings (n = 75; dashed line = regression line; empty circle = no-lar (n = 8); grey circle = single-lar (n = 35); black circle = double-lar (n = 32)).*

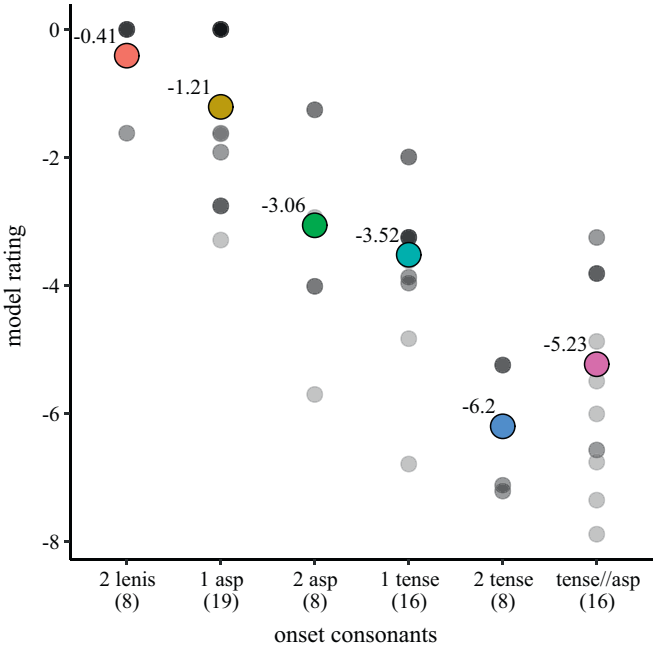

Figure 15. *Laryngeal OCP: Model rating in the MaxEnt simulation, by the onset type.*

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

(40) *Double-tense* vs. *single-tense*

	Mean model rating	Reduced by…
No-lar	−0.41	
Single-tense	−3.52	3.11
Double-tense	−6.2	2.68

If tense//asp items are subject to laryngeal OCP, the reduction in ratings from no-lar items to tense//asp items would be larger than the sum of the reductions from no-lar to single-tense items and from no-lar to single-asp items. This predicted superadditivity emerged, as can be seen in (41).

(41) *Tense//asp* vs. *single-tense* + *single-asp*

	Mean model rating	Reduced by… (relative to no-lar)
No-lar	−0.41	
Single-tense, single-asp	−3.52, −1.21	3.11, 0.8 (sum = 3.91)
Tense//asp	−5.23	4.82

Thus, only non-words with tense//asp, not double-tense, sequences show a superadditive effect in mean model ratings, which is consistent with the laryngeal OCP constraints in (39a) and (39b), learned in the MaxEnt simulation. However, the evidence for a superadditive effect in (41) is not compelling, according to the results of a Bayesian linear regression model analysis (not shown), in which model ratings and the onset consonant type were included as the dependent and independent variables, respectively. The sample distributions show that the posterior probability of the proposition that Δ(no-lar, tense//asp) > Δ(no-lar, single-tense) + Δ(no-lar, single-asp) is only 90% (where Δ(A, B) represents the difference between A and B). Consequently, model ratings data indicate that the laryngeal OCP is not active in the Korean lexicon.

In summary, the results of the MaxEnt simulation indicate that only some subsets of potential target sequences of laryngeal OCP are possibly slightly underrepresented in the Korean lexicon, since co-occurrence constraints were learned only for those subsets (and with relatively low weights), not for the rest of the target sequences, and model ratings for double-lar non-words show no compelling evidence for superadditivity, which is predicted to occur under the hypothesis that laryngeal OCP is active in the Korean lexicon. These results lead us to conclude that the laryngeal OCP in Korean has no, or possibly weak, lexical support.

4.2.4. Summary

In this section, we have reported and discussed the results of the MaxEnt simulation, investigating whether target sequences are underrepresented in the Korean lexicon. The investigation was performed in two parts.

First, we investigated the learned constraints and their weights, summarised in (42).

(42) *Constraints and weights learned in the MaxEnt simulation*

	Constraint	Weight
Palatalisation	General	3.286
Vowel harmony	None	–
Laryngeal OCP	Subset	0.56 ~ 1.26

Note that a general constraint was learned only for palatalisation, and its weight is clearly higher than the weights of the constraints learned for other effects. For vowel harmony, no constraint, general or specific, was learned. For laryngeal OCP, constraints were learned, but they prohibit only some subsets of the target double-lar sequences. These modelling results indicate that TI sequences are

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

underrepresented; disharmonic vowel sequences are not underrepresented; and some, but not all, double-lar sequences are underrepresented, and those only weakly so.

Second, we investigated the patterns of well-formedness ratings assigned to the test non-words by the learned MaxEnt model. To test the effects of palatalisation and vowel harmony, we compared the model ratings between target and control items. In contrast, to investigate the effect of laryngeal OCP, we focussed on whether there is a superadditive effect in the model ratings of double-lar items. The relevant model ratings along with the corresponding human ratings are summarised in (43).

(43) *Model and human ratings, by the type of effect and the stimulus group.*

		Target	Control	Phonotactic effect?
Palatalisation	Model	−8.26	−1.43	Yes
	Human	2.4	3.13	Yes
Vowel harmony	Model	−0.25	−0.71	No
	Human	3.99	3.80	No
Laryngeal OCP	Model	'no superadditivity'		No
	Human	'no superadditivity' or 'floor effect'		Unclear

In the set for palatalisation, mean model ratings are clearly lower for target items than for control items. Model ratings vary depending on the type of TI sequences. Non-words with /ti/ are higher in model ratings than the rest of target items, as in human ratings, but they are still lower than the corresponding control items. Thus, all different types of target non-words in the palatalisation set received low model ratings, indicating that palatalisation affected the model's ratings for non-words. In contrast, in the set for vowel harmony, target items with disharmonic sequences (−0.25) are slightly higher in model ratings than control items with harmonic sequences (−0.71), as in human ratings. This indicates that vowel harmony did not affect the model's ratings. Finally, in the set for laryngeal OCP, we found that there is no compelling evidence for a superadditive effect, indicating that the laryngeal OCP did not affect the model's ratings.

As summarised in (44), the above two sets of investigation together indicate that target sequences of palatalisation are underrepresented in the Korean lexicon, but those of vowel harmony are not. In addition, only some subsets of target sequences of laryngeal OCP may be underrepresented. These indications lead us to conclude that Korean palatalisation has strong lexical support whereas vowel harmony and laryngeal OCP has no, or weak, lexical support. Recall in §3.2.4 that the same conclusions were made from the results of the experiment, except for unclear lexical support for laryngeal OCP.

(44) *Underrepresentation of target sequences and lexical support for the three alternation patterns*

	Modelling		Experiment
	Underrepresentation	Lexical support	Lexical support
Palatalisation	Yes	Strong	Strong
Vowel harmony	No	No	No
Laryngeal OCP	No (possibly weak for some subsets)	No (or weak)	Unclear

Given that Korean palatalisation applying across morphemes is high in robustness whereas vowel harmony in verbal inflection and laryngeal OCP in compound tensing are low in robustness, the results of the current research summarised in (44) generally support the hypothesis in (1).

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

5. Discussion

The results of the current experiment and computational modelling, presented in the previous section, show that the robustness of an alternation correlates well with the strength of the corresponding phonotactics, supporting the claim that there is a tight link between alternations and phonotactics. We believe that human speakers' mental grammar is equipped with a mechanism responsible for this link. These results and indications are consistent with the following previous studies and theories that assume a universal link between alternations and phonotactics: (i) previous studies (Hayes 2004; Prince & Tesar 2004; Jarosz 2006, 2011; Pater & Tessier 2006; Chong 2017, 2021) showing that alternations with phonotactic support are easier to learn than those with no phonotactic support; (ii) studies on derived environment effects (Chong 2017) arguing that true derived environment patterns with a perfect mismatch between phonotactics and alternations are elusive; (iii) phonological theories assuming a unified mechanism for alternations and phonotactics, such as Optimality Theory (Prince & Smolensky [1993] 2004); and (iv) modelling studies showing that a generalisation typically leaks from a tautomorphemic domain into a heteromorphemic domain (Martin 2011). On the other hand, the current research results and findings are inconsistent with previous studies and theories that argue for no universal link between alternation and phonotactics, including early generative phonology (Chomsky & Halle 1968) and some more recent studies (e.g., Paster 2013; Do & Yeung 2021).

This section explores alternative interpretations and analyses of the results from the current experiment and modelling. First, we will address whether the participants' phonotactic judgements for a non-word vary based on the non-word's lexical category. We will then consider the possibility that the lexical support originates from a sub-lexicon or stratum. Finally, we will discuss alternative explanations that attribute the results of the current study to diachronic changes or phonetic naturalness.

5.1. Lexical category of non-words

We conducted an acceptability rating experiment on non-words to measure lexical support for an alternation pattern. One might argue that participants' ratings for a non-word may differ depending on its lexical category. In the experiment, the instruction to the participants included no indication of which lexical category the non-words belong to. Nonetheless, given that isolation forms of nouns, not verbs, in Korean are possible and frequent in natural speech, and nouns form a majority of the Korean lexicon, it is likely (or at least possible), as pointed out by an anonymous reviewer, that the participants viewed the non-words, presented in isolation, as nouns. Note that nouns are predominant in our Korean corpus: among native and Sino-Korean monomorphemes ($n = 38,037$), nouns 92.49% ($n = 35,179$), verbs and adjectives 3.54% ($n = 1,345$) and other categories 3.98% ($n = 1,513$). Then, it is possible that the current experiment measured the support from the noun lexicon, not from the entire Korean lexicon.

We have two responses to this argument. First, given that, as mentioned above, nouns form a majority of words in the vocabulary of Korean, phonotactic regularities of nouns are not significantly different from those of the Korean vocabulary as a whole. Accordingly, the Korean participants' phonotactic judgements for non-words would not be significantly different, regardless of whether they consider the non-words to be nouns or any Korean word.

In addition, more importantly, we do not think that the participants' phonotactic judgements for a non-word vary greatly depending on the assumed lexical category of the non-word. It seems that speakers' phonotactic judgements for a non-word are mainly based on the phonotactic probability of its component sequences in the entire lexicon, regardless of which lexical category the non-word belongs to.[23] Previous modelling studies on English phonotactics (Hayes & Wilson 2008; Albright 2009; Daland *et al.* 2011) show that there are high correlations between speakers' phonotactic judgements

[23] Note that this does not rule out the possibility that speakers may have and use knowledge of phonotactic differences between lexical categories, such as nouns and verbs, as demonstrated in Smith's (2016) nonce-word experiments, where English speakers were asked to identify whether non-words sound more like nouns or verbs.

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

for non-words and the corresponding predictions of the phonotactic models trained on the English lexicon, indicating that English speakers' phonotactic judgements are mainly determined by phonotactic probability in the entire English lexicon, rather than being affected by factors such as the assumed lexical category and the semantic meaning of the non-words. The speakers' judgement data used in those modelling studies were from experiments in which non-words were presented in isolation (Daland *et al.* 2011), as verbs (Albright & Hayes 2003), and as both verbs and nouns (Albright 2009: 28, fn. 17). In particular, Albright compared ratings for the same set of non-word items by two different groups of native English-speaking subjects. For one group of subjects, non-word items were presented as verbs, and for the other group, they were presented as both verbs and nouns. The correlation between the mean ratings from the two groups is quite high, 0.853, which we think indicates that the speakers' phonotactic well-formedness judgements are not significantly affected by the assumed lexical category of the non-word.

In summary, we believe that the results of the present experiment reflect the phonotactic well-formedness in the Korean lexicon, not the noun sub-lexicon, although the noun sub-lexicon is not substantially different in phonotactic regularities from the entire Korean lexicon, and thus Korean speakers' well-formedness judgements would be more or less same for nouns and typical Korean words.

5.2. Support from a sub-lexicon or stratum

Based on the results of the experiment and modelling, we have concluded that Korean vowel harmony has no lexical support. However, one might argue, as suggested by an anonymous reviewer, that the current experiment and modelling are not appropriate for measuring the lexical support for Korean vowel harmony. This argument is based on the idea that since vowel harmony alternation in Korean is attested only in verbal inflection, the lexical support for it can come from a sub-lexicon consisting of verbs and adjectives.

We think that this potential problem is part of a more general issue, that is, whether the lexical support for an alternation can come from a sub-lexicon or stratum characterised by lexical category or etymological origin of words in which the alternation is attested. We have already rejected a possibility, in §4.2, in which the lexical support comes from a lexical stratum consisting of either native or Sino-Korean words, under the assumption that ordinary Korean speakers cannot reliably distinguish between native and Sino-Korean words. Let us now consider the possibility in which the lexical support for an alternation pattern can come from words of a particular lexical category in which the alternation is attested. More specifically, should we investigate phonotactics of the verbal lexicon to measure the lexical support for Korean vowel harmony? Recall that the present study tests the research hypothesis that alternations with strong lexical support are more robust than those with no, or weak, lexical support. In comparing the robustness of a pattern and its lexical support, the generality of an alternation (which defines its robustness) needs to be measured by checking all its potential target environments across different lexical categories, and, similarly, lexical support must be measured by checking alternation-provoking sequences in the entire lexicon, not in a sub-lexicon consisting of words belonging to a particular lexical category. Accordingly, measuring the phonotactic support for Korean vowel harmony from the verb lexicon is not directly relevant to our goal, that is, testing the hypothesis in (1).

Nonetheless, for the sake of curiosity, we decided to investigate whether Korean vowel harmony is supported by phonotactics of verbs and adjectives. Following the suggestion by the anonymous reviewer, we first address this issue by investigating the participants' responses for non-words ending in /a/ or /ʌ/ which are similar in segmental composition to the frequently used verb forms inflected with the DECL.IND (etc.) suffix /-A/. We then perform a simulation of learning a MaxEnt grammar using a verbal lexicon consisting of monomorphemic verbs and adjectives.

Twenty-five non-words in the vowel harmony set end in /a/ or /ʌ/. The participants ratings for them show that disharmonic non-words (3.80, $n = 12$) are slightly higher in mean ratings than harmonic

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

ones (3.60, n = 13). The lack of vowel harmony effect is confirmed by the results of a Bayesian mixed effects linear regression model, not shown, in which the posterior probability of the proposition that disharmonic non-words are lower in ratings than harmonic ones is only 21.03%. These results are not significantly different from the results of the analysis for the entire vowel harmony set of non-words, reported in §3.2.2.

In the MaxEnt simulation using a verbal lexicon consisting of monomorphemic verbs (n = 1057) and adjectives (n = 288), all of which are native Korean, most vowel projection constraints learned were either trivial (such as *## 'no word without a vowel') or those involving vowels with highly limited distribution, [y ø ɨ], as in the results of the modelling based on the entire lexicon, presented in §4.2.2. Only three of the learned constraints with relatively low weights (0.643, 0.692, 1.658) seem to have vowel harmony effects. They require a dark vowel [u] to co-occur with some other dark vowels, but not with a light vowel [o]. Note that these constraints prohibit only one, that is, [uo], out of sixteen target vowel sequences tested in this study, shown in (21). This indicates that most of target vowel sequences are not underrepresented in the verbal lexicon, which is not significantly different from the results of the simulation using the entire lexicon, reported in §4.2.2.

In summary, the results of the present study show that Korean vowel harmony is not phonotactically supported, regardless of whether we limit ourselves to verbs or not.

5.3. A diachronic approach

The results of the current research, summarised in (44), show a tight link between phonotactics and alternations. These findings lead us to believe that the grammar itself and/or its learning algorithm includes a mechanism that integrates alternations and phonotactics. While we will not delve into the grammatical implementation of the observed alternation-phonotactics link in this study, it is notable that there are several relevant approaches from previous work that we may consider. First, in Standard Optimality Theory (Prince & Smolensky [1993] 2004), the same set of constraints is employed to explain both patterns of alternations and phonotactics, and the mechanism for alternations is in principle inseparable from that for phonotactics. In addition, Martin (2011) provides an analysis of 'grammatical leakage' (Martin's term), in which a generalisation that holds for the tautomorphemic domain leaks into the heteromorphemic domain. Specifically, Martin uses a MaxEnt learning algorithm with a Gaussian prior to explain data from English and Navajo in which categorical phonotactic constraints holding in the monomorphemic lexicon are also active, though gradiently, in the heteromorphemic domain. Martin's approach appears to be more promising for the data we discussed in this study, given that the effects of the relevant constraints are mostly gradient. Their specific formal analysis is beyond the scope of this article, and we leave it for future research.

One might attribute the correlations between alternations and phonotactics attested across languages, including those observed in this study, to the diachronic development of relevant segment sequences. For instance, as pointed out by an anonymous reviewer, Korean palatalisation used to be exceptionless both within and across morphemes in the early nineteenth century (Cho 2009, §19.6). Morpheme-internal exceptions to palatalisation then arose, which were mainly due to a subsequent historical sound change in which an underlying /i̯i/ diphthong became a monophthong /i/ (e.g., *mati* < *mati̯i* 'joint') or to foreign loans (e.g., *tʰipi* 'TV'). If these historical changes introduced a relatively small number of exceptions to the Korean lexicon, without affecting Korean palatalisation across morphemes, our finding – that the palatalisation constraint, that is, *TI, is active both within and across morphemes in Contemporary Korean – would be a consequence of the historical changes involved. Similarly, the lack or weakening of vowel harmony and laryngeal OCP effects in both alternations and phonotactics might also be attributed to some relevant historical changes.

We consider the diachronic account to be at best tentative, and possibly implausible, for the following reasons. First, only after we are informed of the exact range and magnitude of the influence exerted by the historical changes will we be able to evaluate the success of the diachronic account. Thus, we cannot

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

conclude that our findings can be explained solely based on historical changes until we know, rather than speculate, whether and exactly how much the historical changes affected alternating forms and lexical statistics. Second, as shown by Martin (2011), it is possible that the extent and impact of the historical changes may be subject to restrictions in the grammar and/or its learning algorithm. For instance, it is possible that the creation and acceptance of exceptional new words may be stochastically blocked by grammatical restrictions. In such cases, the effects of the historical changes can be considered grammatical. Finally, and most importantly, we believe that the diachronic account is too unconstrained to explain the relationship between alternations and phonotactics. According to the diachronic approach (e.g., Paster 2013), phonotactic and alternation patterns may change freely and independently of each other. Therefore, the diachronic approach predicts that cases like true derived environment patterns with a perfect mismatch between alternations and phonotactics can be attested without any restriction. This prediction is not supported by the results of the present study. None of the three patterns tested in this study, which are known to exhibit 'derived environment' effects, show a perfect mismatch between alternations and phonotactics. Instead, our research results align with previous studies (Hayes 2004; Prince & Tesar 2004; Jarosz 2006, 2011; Pater & Tessier 2006; Chong 2017, 2021) that argue that alternations with phonotactic support are easier to learn and predict that alternations with mismatching phonotactics are rare or dispreferred. Thus, it appears that the diachronic approach is not well-supported.

5.4. Phonetic naturalness

As suggested by an anonymous reviewer, the observed correlations between alternations and phonotactics in this study could be attributed to phonetic naturalness, which may shape speakers' phonotactics. For instance, the high robustness of, and high phonotactic support for, palatalisation could both be explained by its phonetic naturalness. Similarly, the low robustness of, and low phonotactic support for, vowel harmony could also be elucidated by the absence of such naturalness in vowel harmony of the type attested in Korean, which is typologically uncommon. One possible way to test this, a potential future direction for this study, would be to compare phonetically similar patterns in two different languages that differ in the robustness of their derived environment patterns.

6. Conclusion

In this study, to test the hypothesis in (1), repeated below, we conducted an acceptability rating experiment for Korean non-words and performed computational modelling of Korean phonotactic grammar.

(1) *Hypothesis for a link between phonotactics and alternations*
Alternation patterns with strong lexical support are more robust (i.e., more productive and more general) than those with no, or weak, lexical support.

From the results of the experiment and modelling, we made the following findings:

First, sequences that violate the palatalisation constraint were judged to be ill-formed by the participants and are underrepresented in the lexicon.

Second, sequences that violate the vowel harmony constraint were not judged to be ill-formed by the participants and are not underrepresented in the lexicon.

Third, sequences that violate the laryngeal OCP are either not or weakly underrepresented in the lexicon.

These findings suggest that Korean palatalisation, which is highly robust, has strong lexical support, whereas vowel harmony and laryngeal OCP, both of which are not robust, have little or weak lexical support, confirming the hypothesis in (1). Consequently, none of the three derived environment effect phenomena in Korean shows a mismatch between alternations and phonotactics. We can conclude that the results of the present study support the claim of a tight link between alternations and

https://doi.org/10.1017/S0952675725000016 Published online by Cambridge University Press

phonotactics. This suggests the presence of a grammatical mechanism and/or a learning bias responsible for integrating alternations with phonotactics.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0952675725000016.

Acknowledgements. We would like to thank the associate editor and three anonymous reviewers for their helpful and detailed comments. Our thanks also extend to Bruce Hayes, Donca Steriade, Canaan Breiss, James Whang, Haruo Kubozono, Mira Oh, Soohyun Kwon, Nayoung Park, Hoju Cha, and the audiences at Phonology Festa 16, Japan (online, 8–9 March 2021), and at the 2022 Annual Meeting on Phonology, UCLA (21–23 October 2022). We are particularly indebted to Daniel Plesniak, whose comments were invaluable in improving this article. Any remaining errors are, of course, our own responsibility. Lastly, we extend our thanks to our experimental subjects for their time and participation.

Competing interests. The authors declare no competing interests.

References

Albright, Adam (2009). Feature-based generalization as a source of gradient acceptability. *Phonology* **26**, 9–41.

Albright, Adam & Bruce Hayes (2003). Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition* **90**, 119–161.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language* **68**, 255–278.

Becker, Michael, Nihan Ketrez & Andrew Nevins (2011). The surfeit of the stimulus: analytic biases filter lexical statistics in Turkish laryngeal alternations. *Lg* **87**, 84–125.

Boersma, Paul & David Weenink (2020). Praat: doing phonetics by computer. Computer program. Available at http://praat.org.

Breiss, Canaan & Bruce Hayes (2020). Phonological markedness effects in sentence formation. *Lg* **96**, 338–370.

Bürkner, Paul-Christian (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal* **10**, 395–411.

Capodieci, Frank, Bruce Hayes & Colin Wilson (2009). *Manual: phonotactic learning program*. Ms., University of California, Los Angeles.

Cho, Hyesun (2012). Statistical learning of Korean phonotactics. *Studies in Phonetics, Phonology and Morphology* **18**, 339–370.

Cho, Young-mee Yu (2009). A historical perspective on nonderived environment blocking: the case of Korean palatalization. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: essays in honor of Paul Kiparsky*. Cambridge, MA: MIT Press, 461–486.

Cho, Young-mee Yu (2016). Korean phonetics and phonology. In Mark Aronoff (ed.) *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. Available at https://doi.org/10.1093/acrefore/9780199384655.013.176.

Choi, Myungok (1985). Pyenchik tongsauy umwunhyensangey tayhaye–p-, s-, t-pyenchiktongsalul cwungsimulo [An explanation of the phonological phenomena of irregular verbs in Korean]. *Kwukehak [Korean Linguistics]* **14**, 149–188.

Chomsky, Noam & Morris Halle (1968). *The sound pattern of English*. New York: Harper & Row.

Chong, Adam J. (2017). *On the relation between phonotactic learning and alternation learning*. PhD dissertation, University of California, Los Angeles.

Chong, Adam J. (2019). The exceptionality and derived environment effects: a comparison of Korean and Turkish. *Phonology* **36**, 543–572.

Chong, Adam J. (2021). The effect of phonotactics on alternation learning. *Lg* **97**, 213–244.

Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis & Ingrid Norrmann (2011). Explaining sonority projection effects. *Phonology* **28**, 197–234.

Do, Youngah & Ping Hei Yeung (2021). Evidence against a link between learning phonotactics and learning phonological alternations. *Linguistics Vanguard* **7**, article no. 20200127.

Franke, Michael & Timo B. Roettger (2019). *Bayesian regression modeling (for factorial designs): a tutorial*. Ms., University of Tübingen and University of Oslo. Available online at https://doi.org/10.31234/osf.io/cdxv3.

Frisch, Stefan, Janet Pierrehumbert & Michael Broe (2004). Similarity avoidance and the OCP. *NLLT* **22**, 179–228.

Goldwater, Sharon & Mark Johnson (2003). Learning OT constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.) *Proceedings of the Stockholm Workshop on Variation Within Optimality Theory*. Stockholm: Department of Linguistics, Stockholm University, 111–120.

Hayes, Bruce (2004). Phonological acquisition in Optimality Theory: the early stages. In R. Kager, J. Pater & W. Zonneveld (eds.) *Constraints in phonological acquisition*. Cambridge: Cambridge University Press, 158–203.

Hayes, Bruce & James White (2013). Phonological naturalness and phonotactic learning. *LI* **44**, 45–75.

Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *LI* **39**, 379–440.

Hayes, Bruce, Kie Zuraw, Peter Siptar & Zsuzsa Londe (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Lg* **85**, 822–863.

Ito, Chiyuki (2014). Compound tensification and laryngeal co-occurrence restrictions in Yanbian Korean. *Phonology* **31**, 349–398.
Jang, Hayeun (2017). Decay of ungrounded rules: vowel harmony of Middle Korean in Contemporary Korean. In Michael Yoshitaka Erlewine (ed.) *Proceedings of GLOW in Asia XI*, volume **2**. Cambridge, MA: MIT Working Papers in Linguistics, 87–101.
Jarosz, Gaja (2006). *Rich lexicons and restrictive grammars: maximum likelihood learning in Optimality Theory*. PhD dissertation, Johns Hopkins University.
Jarosz, Gaja (2011). The roles of phonotactics and frequency in the learning of alternations. In Nick Danis, Kate Mesh & Hyunsuk Sung (eds.) *Proceedings of the 35th annual Boston University Conference on Language Development (BUCLD)*, volume 2. Somerville, MA: Cascadilla Press, 321–333.
Jo, Jinyoung (2023). Korean vowel harmony has weak phonotactic support and has limited productivity. *Phonology* **40**, 65–100.
Jun, Jongho (2010). Stem-final obstruent variation in Korean. *Journal of East Asian Linguistics* **19**, 137–179.
Jun, Jongho (2018). Morpho-phonological processes in Korean. In Mark Aronoff (ed.) *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. Available at https://doi.org/10.1093/acrefore/9780199384655.013.241.
Jun, Jongho (2021). Morphophonological gradience in Korean n-insertion. *Glossa* **6**, Article no. 40.
Jun, Jongho & Jeehyun Lee (2007). Multiple stem-final variants in Korean native nouns and loanwords. *Eoneohag [Linguistics]* **47**, 159–187.
Kang, Hijo (2012). *Diachrony in synchrony: Korean vowel harmony in verbal conjugation*. PhD dissertation, Stony Brook University.
Kang, Hijo (2016). Hankuke yongenhwalyongeyse poinun moumcohwauy pyeniwa pyenhwa: hyencangsiwa silceysi yenkwu [Variation and change of Korean vowel harmony in verbal conjugation: an apparent time and real time study]. *Eoneohag [Linguistics]* **76**, 27–56.
Kang, Hijo & Mira Oh (2016). Dynamic and static aspects of laryngeal co-occurrence restrictions in Korean. *Studies in Phonetics, Phonology and Morphology* **22**, 3–34.
Kang, Hijo & Mira Oh (2019). The asymmetric tense consonant effects n compound and word-initial tensifications in Korean. *Studies in Phonetics, Phonology and Morphology* **25**, 3–30.
Kang, Yoonjung (2006). Neutralizations and variations in Korean verbal paradigms. *Harvard Studies in Korean Linguistics* **11**, 183–196.
Kim, Moosik (2007). 16-7 seyki kwuke hancaeuy picwungkwa ku thukcing: swunchenkimssi mich hyenphwungkwakssi hankulphyencilul taysangulo [A study on the proportion of loanword from Chinese, and its traits which was expressed at Korean of 16th century]. *Eomunlonchong [Korean Language and Literature]* **47**, 251–276.
Kim, Seoyoung (2016). *Phonological trends in Seoul Korean compound tensification*. Master's thesis, Seoul National University, Korea.
Kiparsky, Paul (1973). Phonological representations: abstractness, opacity and global rules. In Osamu Fujimura (ed.) *Three dimensions in linguistic theory*. Tokyo: TEC, 1–136.
Lee, Hoyoung (1996). *Kwukeumsenghak [Korean phonetics]*. Seoul: Thaehaksa.
Martin, Andrew (2011). Grammars leak: modeling how phonotactic generalizations interact within the grammar. *Lg* **87**, 751–770.
National Institute of the Korean Language (1999). *Phyocwun kwuke taysacen [Standard Korean dictionary]*. Seoul: Doosan Dong-A.
Nicenboim, Bruno & Shravan Vasishth (2016). Statistical methods for linguistic research: foundational ideas – part II. *Language and Linguistics Compass* **10**, 591–613.
Park, Nayoung (2020). *Hankwuke umsopayyelceyyakuy thongkyeycek haksupkwa cekhyengseng phantan [Stochastic learning and well-formedness judgement in Korean phonotactics]*. PhD dissertation, Seoul National University.
Paster, Mary (2013). Rethinking the 'duplication problem'. *Lingua* **126**, 78–91.
Pater, Joe & Anne-Michelle Tessier (2006). L1 phonotactic knowledge and the L2 acquisition of alternations. In Roumyana Slabakova, Silvina Montrul & Philippe Prévost (eds.) *Inquiries in linguistic development: in honor of Lydia White*. Amsterdam: John Benjamins, 115–131.
Pierrehumbert, Janet (1993). Dissimilarity in the Arabic verbal roots. *NELS* **23**, 367–381.
Prince, Alan & Paul Smolensky ([1993] 2004). *Optimality Theory: constraint interaction in generative grammar*. Oxford: Blackwell. Originally published in 1993 as technical report no. 2 of the Rutgers Center for Cognitive Science.
Prince, Alan & Bruce Tesar (2004). Learning phonotactic distributions. In René Kager, Joe Pater & Wim Zonneveld (eds.) *Constraints in phonological acquisition*. Cambridge: Cambridge University Press, 245–291.
R Core Team (2022). *R: a language and environment for statistical computing*. Computer software. Available at https://www.R-project.org/.
Shih, Stephanie (2017). Constraint conjunction in weighted probabilistic grammar. *Phonology* **34**, 243–268.
Smith, Jennifer L. (2016). Segmental noun/verb phonotactic differences are productive too. *Proceedings of the Linguistic Society of America* 1, Article no. 16.
Wilson, Colin & Marieke Obdeyn (2009). Simplifying subsidiary theory: statistical evidence from Arabic, Muna, Shona, and Wargamay. Ms., Johns Hopkins University.
Zuraw, Kie (2011). Predicting *sai-siot* in Korean compound nouns: phonological and non-phonological factors. Paper presented at the 21st Japanese/Korean Linguistic Conference, Seoul National University, October 2011.