

# PARADIGM UNIFORMITY IN HERITAGE KOREAN\*

YOONJUNG KANG

*University of Toronto Scarborough*

## 1 Introduction

Heritage speakers are bilingual speakers who grow up in a home where a language other than the dominant language of the larger society is spoken (Nagy 2015; Polinsky & Kagan 2007). These speakers get an early exposure to their first language (L1) at home but the exposure to data tends to be limited and they usually do not receive any formal schooling in their L1. As a result, heritage speakers' L1 abilities vary widely and often exhibit signs of incomplete acquisition (L1 gets fossilized at an intermediate stage in acquisition) or attrition (L1 abilities acquired earlier get lost as speakers switch to the societal language as their primary language) (Montrul 2002; Polinsky 2006; Polinsky 2011). Heritage language, nevertheless, is a coherent system and tends to manifest emergent traits that arguably reflect universal grammatical preferences (Polinsky 2006).

The theoretical question addressed in this paper is the status of preference for paradigm uniformity as a grammatical constraint. Paradigm leveling, whereby allophony is eliminated in an inflectional paradigm, is one of the most prominent forms of morphological change and there are opposing views regarding its underlying mechanism. Many propose that paradigm leveling is a reflection of the universal preference for paradigm uniformity (Kiparsky 1982), which the *Optimality Theory* encapsulates under terms like *Output-Output Correspondence* (OO-CORR) or *Uniform Exponence* (Benua 1997; Kenstowicz 1996; McCarthy 1998; Steriade 2000). On the other hand, others argue that the cases of paradigm leveling in attested cases of historical changes are more properly interpreted as a case of analogical extension of a dominant non-alternating paradigm, rather than as a result of general preference for non-alternation (Fertig

---

\* I would like to thank Seung-Joon Park and Samantha Kim for their assistance with data collection and analysis, and audience members at the 2010 Annual meeting of the Canadian Linguistic Association, WAFL 2016, and the University of Toronto CRC Summer Phonology/Phonetics Workshop 2016, where earlier versions of the study were presented for helpful comments. The study was funded by the University of Toronto Connaught New Staff Matching Grant.

2016; Garraett 2008) casting doubts on the purported universal preference for paradigm uniformity.

However, the lack of non-analogical paradigm leveling in historical changes is not necessarily incompatible with the view that there is a grammatical preference for uniform paradigm. It has been proposed that OO-CORR constraints are ranked high in the initial state of acquisition (Hayes 2004; McCarthy 1998; White 2017) as evidenced by special preference for non-alternation exhibited in child language (Do 2013; Tessier 2012), but these constraints get demoted and overridden as the acquisition progresses further and learners are exposed to more and more language-specific learning input. Given these two perspectives, the question of whether there is such a grammatical preference cannot be answered by examining the final outcome of acquisition. Rather, we need to examine the intermediate stages of acquisition where the initial preference is not yet completely overridden by the learned patterns from the ambient language data. Child language obviously provides relevant evidence (Do 2013; Tessier 2012) and heritage language also provides a potential source of such evidence.

In this context, Korean nouns provide an interesting test case because there is a lexical gap of non-alternating paradigm for nouns ending in a coronal obstruent. In homeland Korean, the synchronic variation that affects coronal obstruent-final nouns never shows paradigm leveling. A large body of study has demonstrated that inductive statistical learning is at play and can account for the pattern of innovation and variation in these nouns, which supports the view that paradigm leveling is just another type of analogical extension. The current paper examines whether such paradigm leveling is attested in the absence of a direct model of analogical extension in heritage Korean speakers' speech, whose speech is characterized by limited learning input and hence an increased likelihood to exhibit default grammatical preferences than the monolingual Korean speech. The paper is organized as follows. Section 2 provides a background about Korean noun morpho-phonology and lays out the specific research questions. Section 3 provides an overview of the experiment and Section 4 discusses the results and analyses. Section 5 concludes the paper.

## 2 Korean Nouns

The consonant inventory of Korean is provided in **Error! Reference source not found.**) The plosives and affricates have three-way laryngeal contrasts between lenis (/p t k ts/), aspirated (/p<sup>h</sup> t<sup>h</sup> k<sup>h</sup> ts<sup>h</sup>/) and fortis (/p' t' k' ts'/) series and coronal fricatives contrast between non-fortis (/s/) and fortis (/s') fricatives. In the coda position, the underlying laryngeal and manner contrasts of various obstruents are neutralized as summarized in **Error! Reference source not found.**) and [p t k m n ŋ l] are the only consonants that are allowed in this position. Also, no complex coda is allowed and the underlying clusters are simplified in coda position.

## (1) Consonant inventory of Korean

p p <sup>h</sup> p'	t t <sup>h</sup> t' ts ts <sup>h</sup> ts'	k k <sup>h</sup> k'	
	s s'		h
m	n	ŋ	
	l		
w	j		

## (2) Coda neutralization

/p p <sup>h</sup> p' /	→ [p] / ___ ] <sub>σ</sub>
/k k <sup>h</sup> k' /	→ [k] / ___ ] <sub>σ</sub>
/t t <sup>h</sup> t' ts ts <sup>h</sup> ts' s s' h /	→ [t] / ___ ] <sub>σ</sub>

## (3) Coda cluster simplification

CC	→ C / ___ ] <sub>σ</sub>
----	--------------------------

These phonological processes introduce alternation as illustrated by the examples in **Error! Reference source not found.**) For example, the noun ‘front’ ends in an underlying aspirated stop /p<sup>h</sup>/, which surfaces as [p<sup>h</sup>] before a vowel-initial suffix but as a lenis stop [p] when unsuffixed or followed by a consonant-initial suffix. Due to these neutralization processes, underlying contrasts among noun-final obstruents, some of them being actual lexical contrasts, are neutralized as shown in (5).<sup>1</sup>

## (4) Coda neutralization and alternation.

	in (TOP.)	#	to (‘also’)	
a. /ap <sup>h</sup> /	[a.p <sup>h</sup> -in]	[ap]	[ap-t'o]	‘front’
b. /kaps/	[kaps'-in]	[kap]	[kap-t'o]	‘price’

## (5) Neutralization of underlying contrasts

	in (TOP.)	#	to (‘also’)	
a. /hilk/	[hilk-in]	[hik]	[hik-t'o]	‘soil’
/hik/	[hik-in]	[hik]	[hik-t'o]	‘black’
b. /kaps/	[kaps'-in]	[kap]	[kap-t'o]	‘price’
/kap/	[kap-in]	[kap]	[kap-t'o]	‘pack’
c. /nat <sup>h</sup> /	[nat <sup>h</sup> -in]	[nat]	[nat-t'o]	‘individual’
/nac <sup>h</sup> /	[nac <sup>h</sup> -in]	[nat]	[nat-t'o]	‘face’
/nac/	[nac-in]	[nat]	[nat-t'o]	‘day’
/nas/	[nas-in]	[nat]	[nat-t'o]	‘sickle’

Such neutralization processes create a potential problem for learners; given a neutralized form (e.g., [nat]), there are multiple compatible inflected forms (e.g., [nas-i], [nac-in], [nat<sup>h</sup>-in], etc.). There are reasons to believe that this is indeed what happens in Korean; innovations and variations, both synchronic and diachronic, are widely attested in Korean nouns that are best

<sup>1</sup> Allophonic rules such as intersonorant voicing of lenis obstruents, palatalization of /s/ before /i/, or allophonic rhotocization of /l/ are not reflected in our transcription.

explained as results of reanalysis based on the unsuffixed form as a base (Albright 2008; Hayes 1998; Jun 2010; Kang 2003; Ko 1989). Importantly, the innovations reflect the statistical distribution of noun-final consonants in the language (Albright 2008; Jun 2010; Kang 2003; Ko 1989) such that the new variants take after the most frequent noun type for the given base. For most nouns, attested reanalyses eliminate the existing alternation in the paradigm as in **Error! Reference source not found.a)** (Kenstowicz 1996; Ko 2006). For coronal obstruent-final nouns, however, the pattern is quite different. The innovation is towards other coronal obstruents, most frequently to /s/ but never to /t/, as shown in **Error! Reference source not found.b)**. Notably, this variation pattern in Homeland Korean is a reflection of a peculiar gap in the Korean lexicon; namely, there are no productively used /t/-final nouns to serve as a model for analogical paradigm leveling for coronal obstruent-final nouns.

(6) Variation in noun-final consonants in Homeland varieties of Korean

a. Labial and dorsal obstruents and clusters

Original	>	Reanalysis	
[mulip] ~ [mulip <sup>h</sup> -in]	>	[mulip] ~ [mulip-in]	‘knee, TOP.’
[puʌk] ~ [puʌk <sup>h</sup> -in]	>	[puʌk] ~ [puʌk-in]	‘kitchen, TOP.’
[tak] ~ [talk-in]	>	[tak] ~ [tak-in]	‘chicken, TOP.’

b. Coronal obstruents

[pat] ~ [pat <sup>h</sup> -in]	>	[pat] ~ [pac <sup>h</sup> -in], [pas-in], *[pat-in]	‘field, TOP.’
[k’ot] ~ [k’oc <sup>h</sup> -in]	>	[k’ot] ~ [k’os-in], *[k’ot-in]	‘flower, TOP.’

While the variation in nominal inflection is thoroughly described for the homeland adult Korean, studies on the morpho-phonology of Child Korean or Heritage Korean are far more sparse. (Do 2013) examined the production of consonant-final nouns in Child Korean and found that the youngest speakers in her study (4;6 ~ 5;3) produced all coronal obstruent-final nouns as [t]-final leveling out alternation (e.g., [pat-e] ‘field, LOC.’ cf. /pat<sup>h</sup>/; [nat-i] ‘day, NOM.’ cf. /nac/). The intermediate group (5;1 ~ 7;3) inflected coronal obstruent-final nouns as [s]-final but tend to use alternative wording or morphological structure to avoid producing an alternation. The oldest group (8;0 ~ 8;2) were adult-like in their noun production. Do (2013)’s study is informative in showing that the younger speakers do exhibit an emergent preference for the non-alternating paradigm, not directly attributable to the learning input, lending support for the status of paradigm uniformity as a universal preference. The study also shows that nominal inflections are acquired quite late with children producing paradigm-leveling for coronal obstruent nouns past their 5<sup>th</sup> birthday.

Our study examines noun paradigms in Heritage Korean in Toronto. Specifically, we ask the following questions; a) do heritage speakers produce non-alternating paradigms more than native speakers in general?; b) do heritage speakers produce non-alternating paradigms for coronal obstruent-final nouns, which are not directly supported by the learning data?; c) how does the preference for non-alternating paradigms change with an increase in proficiency, which we presume is correlated with an increase in exposure to the learning input?

## 3 Methods

### 3.1 Participants

Heritage Korean speakers in this study are defined as speakers of Korean who were born in Canada, or were born in Korea and came to Canada at or before the age of 5, and were raised by Korean-speaking caregivers. Native Korean speakers are defined as speakers of Korean from the Seoul Metropolitan area who were born in Korea, lived there at least until the age of 16 and were raised by Korean-speaking caregivers. A number of the participants who came to Canada between the ages of 6 and 15, and hence did not meet the criteria of either speaker groups answered the recruitment ad and participated. We collected and analyzed the data from this intermediate group of speakers as well.<sup>2</sup> Nine heritage, eight intermediate and nine native speakers were initially recruited from the campus of the University of Toronto Scarborough in 2008 and then 20 additional heritage speakers were recruited from the Korean communities in Toronto and Vancouver in 2014.

Following the main experiment, speakers also completed a word translation task and filled out a background questionnaire. The word translation task was designed to gage the speakers' level of lexical knowledge in Korean and it also serves as a general measure of grammatical proficiency (Polinsky 2006). In this task, participants were given a list of Korean words and were asked to translate or explain their meaning in English. A total of 80 Korean nouns divided into three frequency categories were included (21 low-, 39 mid-, and 20 high-frequency words). The low-frequency words were selected from those with a frequency of occurrence below 99 in a word frequency list compiled from a 1.5 million word corpus (Cho 2002), while the mid-frequency words ranged between 100 and 499, and the high-frequency words were 500 or above. A translation response was categorized as acceptable if it exactly matched the meaning or hinted at a closely related concept.

### 3.2 Methods and Stimuli

The data were elicited through a simple sentence completion production task. In this task, the experimenter pronounced the target noun in isolation form, where a potential contrast in final obstruent is neutralized, and prompted the participants to complete a sentence using the provided noun combined with a vowel-initial suffix. In completing the sentence, the participants had to combine the noun with a vowel-initial suffix, revealing their analysis of the underlying final consonant of the noun. In the first round, the participants were prompted to produce all the target nouns with the nominative marker, /-i/, and in the second round, the same target nouns were used, except the suffix was a topic marker, /-in/. Some examples are provided in **Error! Reference source not found.**) An audio recording was made of the sessions and the speakers' responses were categorized into different response types as explained below.

---

<sup>2</sup> Also, two of the speakers born in Canada reported traveling back to Korea for a considerable amount of time—one speaker lived in Korea between 6 and 12 years of age and another speaker went back to Korea for an extended period every year until 15 years of age—and these speakers were also grouped as the Intermediate group.

## (7) Examples of a noun prompt and a frame sentence

## a. Nominative marker

Experimenter: [pam] ‘chestnut’  
 Participant: [\_\_\_\_-i k<sup>h</sup>ita] ‘chestnut-nom. big’

## b. Topic marker

Experimenter: [pam] ‘chestnut’  
 Participant: [\_\_\_\_-in k<sup>h</sup>ita] ‘chestnut-top. big’

Seven different frame sentences were used to make the sentence semantically natural and to make the task not overly monotonous. For each new frame sentence, a couple of practice items were provided before the test items to ensure that the participant understood the task. The breakdown of the nouns by the types of final consonants are provided in Table 1. The first type is nouns that do not alternate. These nouns end in either a sonorant consonant (/m/, /n/, /ŋ/ or /l/) or a non-coronal lenis stop (/p/ or /k/). We expect that these nouns will be produced correctly without alternation by all speakers and that these nouns, in effect, serve as a control condition. We will refer to this type as “TYPE A” (n=44). The second type consists of nouns that end in a consonant cluster (/lk/, /ps/, or /ks/) or a laryngeally marked non-coronal stop (/p<sup>h</sup>/, /k<sup>h</sup>/, or /kʔ/). These nouns should alternate, according to the normative standard pronunciation, but many of these nouns are often produced without alternation in homeland Korean. We refer to this type as “TYPE B” (n=14). The third type includes nouns that end in a coronal obstruent (/s/, /ts/, /tsh/, or /th/) and these nouns are the main focus of our study. These are alternating nouns but unlike the second type, the innovation in homeland Korean does not eliminate alternation but replaces one type of alternation with another. We refer to this last type as “TYPE C” (n= 27). In total, 85 nouns were presented and each speaker produced each noun twice, once with a nominative marker and once with a topic marker and a total of 7819 noun productions (85 nouns \* 2 suffixes \* 46 speakers – 1 omission due to an experimenter error) were analyzed for the study.

**Table 1.** Types of target nouns in the sentence completion task

Type	N	Final Consonants	Alternation	Innovation in Homeland Korean	Examples
A	44	/m,n,ŋ,l,p,k/	No	-	/nun/ ‘eye’, /pap/ ‘rice’
B	14	/lk,ps,ks,p <sup>h</sup> ,k <sup>h</sup> ,kʔ/	Yes	Leveling	/pakʔ/ ‘outside’, /hilk/ ‘dirt’
C	27	/s,ts,tsh,tʰ/	Yes	Alternation	/os/ ‘clothes’, /kʔotsh/ ‘flower’

### 3.3 Analyses

Noun-suffix combinations that participants produced were transcribed and categorized into one of the following four response types: CORRECT, INNOVATION, ALTERNATIVE, and ERROR. These response types are illustrated with representative examples in **Error! Reference source not found.-Error! Reference source not found.** For each example, the first column represents the underlying representation (UR) of the intended noun in the standard pronunciation. In cases of innovation, the new innovated UR is provided after an arrow. The second column shows the isolation forms of nouns given as prompts to the participant. The third column represents the suffixed forms produced by the participant. Note that the transcription reflects all phonemic distinctions but not positional allophonic changes. A response is categorized as CORRECT if the

target noun is produced with the correct underlying consonant as expected by the standard orthographic form and the application of standard phonological rules. Depending on the noun type, the correct form may involve alternation as in **Error! Reference source not found.b)** or no alternation as in **Error! Reference source not found.a)**.

## (8) CORRECT

## a. No alternation

UR	Isolation	Suffixed	
/hani/	[hani]	[hani-i]	‘sky-NOM’
/jaksok/	[jaks’ok]	[jaks’ok-i]	‘promise-NOM’

## b. Alternation

UR	Isolation	Suffixed	
/kaps/	[kap]	[kaps’-in]	‘price-TOP’
/pasas/	[pasat]	[pasas-i]	‘mushroom-NOM’

The second type, INNOVATION, refers to cases where the pronunciation deviates from the standard form but is a possible phonological reanalysis of the isolation form. INNOVATION may eliminate an alternation as in **Error! Reference source not found.a)**, introduce an alternation where there was none before as in **Error! Reference source not found.b)**, or replace one type of alternation with another as in **Error! Reference source not found.c)**. Some of the innovative forms attested in our study are commonly used in homeland Korean (e.g., [mulip-in], [tak-i], or [pis-i]) while others are not. Particularly notable is the fact that the leveling of TYPE C nouns (e.g., [kat] ~ [kat-in]) is frequently attested in the heritage speakers’ production while completely absent in the native control speakers’ production. So, we can also classify the innovations into two types depending on whether a particular innovation for a particular lexical item is attested in the native speakers’ production (NATIVE INNOVATION) or only found in the heritage speakers’ production (HERITAGE INNOVATION). We will discuss the distribution of response types in detail in the result section.

## (9) INNOVATION

## a. No alternation (= Paradigm leveling)

UR	Isolation	Suffixed	
/mulip <sup>h</sup> /→/mulip/	[mulip]	[mulip-in]	‘knee-TOP’
/talk/→/tak/	[tak]	[tak-i]	‘chicken-NOM’
/kAt <sup>h</sup> /→/kAt/	[kAt]	[kAt-in]	‘exterior-TOP’

## b. New alternation

/nuns’Ap/→/nuns’Ap <sup>h</sup> /	[nuns’Ap]	[nuns’Ap <sup>h</sup> -in]	‘eyebrow-TOP’
-----------------------------------	-----------	----------------------------	---------------

## c. Different alternation

/nip <sup>h</sup> /→/nips/	[nip]	[nips’-in]	‘marsh-TOP’
/pits <sup>h</sup> /→/pis/	[pit]	[pis-i]	‘debt-NOM’

The next two types of responses, ALTERNATIVE and ERROR, are almost exclusively found in heritage speakers’ production. The ALTERNATIVE category refers to forms that fail to reveal the speaker’s analysis of the underlying form. There are several different ways this situation may arise. First, the speakers sometimes replace the target noun with another more commonly used noun with a similar meaning, as in **Error! Reference source not found.a)**. Second, the speakers may delete the suffix leaving the noun-final consonant in the same coda position, as in **Error!**

**Reference source not found.**b). Suffix deletion is a very common phenomenon in spoken Korean and suffix deletion does not necessarily make the sentence ungrammatical. In this experiment, however, suffix deletion was only found in the heritage speakers' speech, never in the native control speakers' production. Third, the speakers may use a consonant-initial allomorph of the suffix which keeps the noun-final consonant in coda position, as in **Error! Reference source not found.**c). The nominative and the topic marker each have two allomorphs; the vowel-initial allomorphs ([-i] and [-in]) occur when the noun ends in a consonant, as was the case with our target nouns, while the consonant-initial allomorphs ([-ka] and [-nin]) occur when the noun ends in a vowel. Fourth, the speakers do not resyllabify the noun-final consonant onto the onset of the following vowel, either by inserting a short artificial pause between the noun and the suffix as in **Error! Reference source not found.**d) or glottalizing the final consonant as in **Error! Reference source not found.**e). ALTERNATIVE productions were quite common in heritage speakers' speech (16.6%), particularly for lower proficiency speakers. ERROR refers to forms that are not possible reanalyses of the neutralized coda. For example, there is no (morpho-)phonological rule of Korean that supports an alternation between [k] and [s] in **Error! Reference source not found.**a) or between [k] and [t] in the examples in **Error! Reference source not found.**b). This last category is quite infrequent comprising only 1.3% of the total HERITAGE productions.

## (10) ALTERNATIVE

UR	Isolation	Suffixed	
a. /pak' / → /pak'ats <sup>h</sup> /	[pak]	[pak'ats <sup>h</sup> -in]	'outside-TOP'
b. /tʌts <sup>h</sup> /	[tʌt]	[tʌt-Ø (k <sup>h</sup> ida)]	'trap (is big)' (intended: 'trap-NOM')
c. /ap <sup>h</sup> /	[ap]	[ap-k'a]	'front-NOM'
d. /k'ots <sup>h</sup> /	[k'ot]	[k'ot#-i]	'flower-NOM'
e. /mos/	[mot]	[motʔ-in]	'nail-TOP'

## (11) ERROR

a. /hiɪk/	[hiɪk]	[his-in]	'soil-TOP'
b. /puʌk <sup>h</sup> /	[puʌk]	[puʌt-in]	'kitchen-TOP'

Finally, heritage speakers often produce forms with a coronal stop [t] or [t<sup>h</sup>] with a nominative suffix [-i], failing to apply the affrication rule that affects coronal stop + [i] suffix sequences (/t, t<sup>h</sup>/ > [ts, ts<sup>h</sup>] / \_\_\_ + [i, j]). Examples are provided in **Error! Reference source not found.**). While these forms are errors in that they are unexpected and unattested for the native control speakers, these errors are different in nature from the errors in **Error! Reference source not found.**) in that the former reflects acceptable reanalyses of the coda [t] in isolation form except that the affrication rule is not applied. Also note that the affrication rule is a non-derived environment blocking rule that only applies across a stem-suffix sequence and [ti] and [t<sup>h</sup>i] sequences are phonotactically licit in Korean (Cho 2009), posing a special challenge for acquisition (Pater & Tessier 2003). So, these examples are classified either as CORRECT if an originally /t<sup>h</sup>-final noun is produced with final [t<sup>h</sup>] as in **Error! Reference source not found.**a) (n=4), or as INNOVATION if a coronal-obstruent final noun is produced with final [t] (n=133) as in **Error! Reference source not found.**b) or [t<sup>h</sup>] and the noun is not originally /t<sup>h</sup>-final (n=6) as in **Error! Reference source not found.**c).

## (12) No affrication in Heritage Korean

UR	Isolation	Suffixed	
a. /kʌtʰ/	[kʌt]	[kʌtʰ-i] (cf. Standard: [kʌtʰ-i])	‘exterior-NOM’
b. /mas/ → /mat/	[mat]	[mat-i]	‘taste-NOM’
c. /pitsʰ/ → /pitʰ/	[pit]	[pitʰ-i]	‘light-NOM’

Heritage speakers produce other types of errors that native speakers do not produce, such as attaching a wrong suffix, as in **Error! Reference source not found.**), or stacking two allomorphs of nominative suffix as in **Error! Reference source not found.**). These errors, while indicative of a lack of mastery of noun paradigm, may still reveal speakers’ analyses of the underlying consonant. So, these suffixal errors were categorized as one of the four response types discussed in **Error! Reference source not found.**)-**Error! Reference source not found.**), based on how the final consonant was realized. For example, **Error! Reference source not found.a)** and **Error! Reference source not found.a)** are categorized as CORRECT, as the form is pronounced with the correct final consonant even though the form contains an error in the suffix. **Error! Reference source not found.b)** and **Error! Reference source not found.b)** are categorized as INNOVATION, as the final consonant differs from the standard form but are phonologically acceptable reanalyses of the neutralized coda. **Error! Reference source not found.c)** and **Error! Reference source not found.c)** are categorized as ALTERNATIVE because the final consonant remains in the coda and does not reveal the speakers’ analyses of the underlying representation. With this background, we now turn to the results.

## (13) Wrong suffix

UR	Isolation	Suffixed	
a. /pakʰ/	[pak]	[pakʰ-e]	‘outside-LOC’ (intended: ‘outside-NOM’)
b. /patʰ/ → /pas/	[pat]	[pas-i]	‘field-NOM’ (intended: ‘field-TOP’)
c. /kilim/	[kilim]	[kilim-to]	‘picture-ALSO’ (intended: ‘picture-NOM’)

## (14) Nominative allomorph stacking

a. /mun/	[mun]	[mun-i-ka]	‘door-NOM-NOM’
b. /jʌpʰ/ → /jʌp/	[jʌp]	[jʌp-i-ka]	‘side-NOM-NOM’
c. /moktsʰʌts/	[moktsʰʌt]	[moktsʰʌt#-i-ka]	‘uvula-NOM-NOM’

### 3.4 Speaker Groups

There is a wide range of individual variation in proficiency in the heritage speaker population and before we provide an aggregate result of heritage speakers as a group, we will motivate the division of the heritage speakers into appropriate proficiency groups. For this initial classification, we tabulate the percentage of non-native-like responses each speaker produced. A response is considered non-native-like if it is an ERROR response, an ALTERNATIVE response, or an INNOVATION response that is never attested in the native speakers’ production (=HERITAGE INNOVATION). We divided heritage speakers into low- and high-proficiency groups by k-means clustering of the percentage of non-native-like responses. Table 2 summarizes the mean and the range of non-native response rates, ages of arrival in Canada, ages at the time of study, and correct response rates in the translation task.

**Table 2.** Speaker groups

<b>Group</b>	<b>NATIVE</b>	<b>INTERMEDIATE</b>	<b>HERITAGE-HIGH</b>	<b>HERITAGE-LOW</b>
N	9 (5F, 4M)	8 (6F, 2M)	18 (8F, 10M)	11 (3F, 8M)
Native-like response rate	1.00 (0.99-1.00)	0.98 (0.94-1.00)	0.91 (0.72-0.97)	0.44 (0.31-0.60)
Correct translation rate	0.88 (0.71-0.97)	0.84 (0.58-0.96)	0.53 (0.31-0.89)	0.25 (0.17-0.38)
Age of Arrival (AOA)	16.9 (16-20)	10.7 (6-15) <sup>3</sup>	0.9 (0-5)	0.2 (0-1)
Age <sup>4</sup>	21.6 (20-24)	20.4 (19-23)	22.5 (16-32)	20.4 (17-31)

## 4. Results

### 4.1 Preference for Non-alternating Paradigms

We examine the response patterns by the noun type and by the speaker group in view of the questions we posed in section 1. Do heritage speakers produce more non-alternating variants than native speakers? Do heritage speakers produce non-alternating paradigms for coronal obstruent-final nouns, which are absent in the learning input? To quantify the degree of preference for non-alternating paradigms, in this section we set aside the ALTERNATIVE and ERROR responses, and examine the CORRECT and INNOVATION responses only, where we can determine the speakers' UR (re)analyses for the target noun. We will return to take a closer look at the other response types in section 4.2.

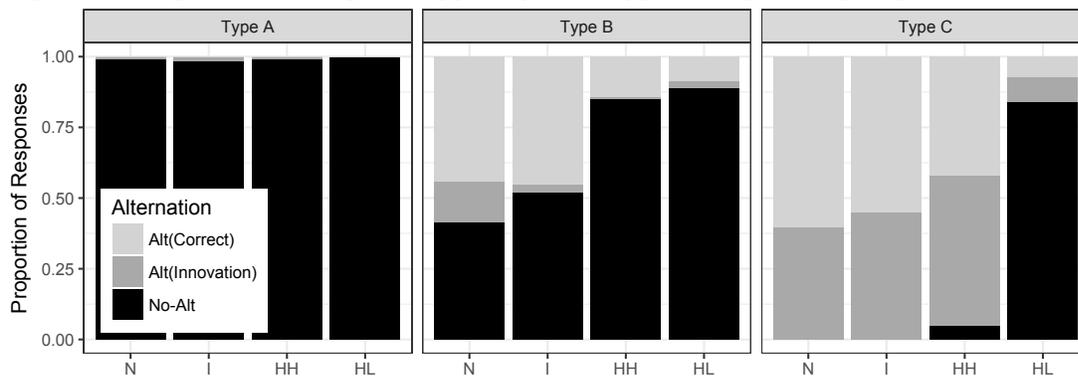
Table 3 and Figure 1 summarize the proportion of alternation and non-alternation responses by the noun type and by the speaker group. For statistical tests, we use mixed-effects logistic regression models using the *lme4* package (Bates et al. 2015) in *R* (R\_Core\_Team 2015) to determine whether speaker groups differ in their preference towards non-alternating paradigms. The response variable is ALTERATION (alternating or non-alternating, collapsing over CORRECT and INNOVATION types) and the predictor variable is speaker GROUP (NATIVE, INTERMEDIATE, HERITAGE-HIGH AND HERITAGE-LOW). The factor GROUP is backward difference coded to compare one level against the next. A by-speaker random intercept is included as a random effect. The output of the statistical models is summarized in Table 4. Follow-up pair-wise comparisons of groups were conducted using the *phia* package (De Rosario-Martinez et al. 2015).

**Table 3.** Non-alternation rate by noun type and speaker group

<b>Group</b>	<b>NATIVE</b>	<b>INTERMEDIATE</b>	<b>HERITAGE-HIGH</b>	<b>HERITAGE-LOW</b>	<b>TOTAL</b>
Type A	0.99 (n=792)	0.98 (n=702)	0.99 (n=1,549)	1.00 (n=673)	0.99 (n=3,716)
Type B	0.41 (n=251)	0.52 (n=219)	0.85 (n=473)	0.89 (n=164)	0.69 (n=1,107)
Type C	0.00 (n=485)	0.00 (n=429)	0.05 (n=933)	0.84 (n=252)	0.12 (n=2,099)

<sup>3</sup> Two intermediate speakers who were born in Canada (AOA=0) but lived in Korea for a considerable duration of time are not included in this calculation.

<sup>4</sup> Two of the heritage speakers (one Heritage-Low and one Heritage-High speakers) did not respond to the question about their age in the background questionnaire but they are estimated to be in a similar age group as the rest of the participants.

**Figure 1.** Proportion of response types by noun type and speaker group**Table 4.** Mixed effects logistic regression models to predict preference for non-alternation

	Type A		Type B		Type C	
	$\beta$	$p$	$\beta$	$p$	$\beta$	$p$
Intercept	<b>5.335</b>	<b>&lt;0.001</b>	<b>1.063</b>	<b>&lt;0.001</b>	-18.002	0.298
GROUP: INT. VS. NAT	-0.627	0.393	0.490	0.318	1.097	0.970
GROUP: HER.HIGH VS. INT	0.879	0.173	<b>2.012</b>	<b>&lt;0.001</b>	30.374	0.365
GROUP: HER.LOW VS. HER.HIGH	0.832	0.349	0.220	0.646	<b>6.390</b>	<b>&lt;0.001</b>

For TYPE A nouns, the response is almost always correct (3684/3716=0.99) across all four speaker groups and a post-hoc test shows that the speaker groups were statistically indistinguishable from one another ( $p > 0.1$ ). This is as expected given that TYPE A nouns do not alternate (e.g., /nun/ [nun] ~ [nun-i] ‘eye’; /mok/ [mok] ~ [mok-i] ‘neck’), which is argued to be the default preference, and there are no dominant alternating paradigms that support innovation towards an alternation for these nouns.

TYPE B nouns alternate in the standard pronunciation (e.g., /talk/ [tak] ~ [tak-i] ‘chicken’) but they are frequently produced without alternation (0.41) by the NATIVE group (/tak/ [tak] ~ [tak-i] ‘chicken’), presumably due to analogy to the dominant non-alternating paradigms. The INTERMEDIATE speakers produce non-alternating forms at a rate comparable to the NATIVE group (0.52) but the HERITAGE speakers, both low and high proficiency, produce non-alternation forms significantly more often (0.85 and 0.89). Also notable is the fact that while NATIVE speakers often produce alternating forms that are themselves innovations, changing one form of alternation to another (0.14), INTERMEDIATE or HERITAGE speakers rarely produce such innovative alternations (INTERMEDIATE: 0.03; HERITAGE-HIGH: 0.01; HERITAGE-LOW: 0.02).

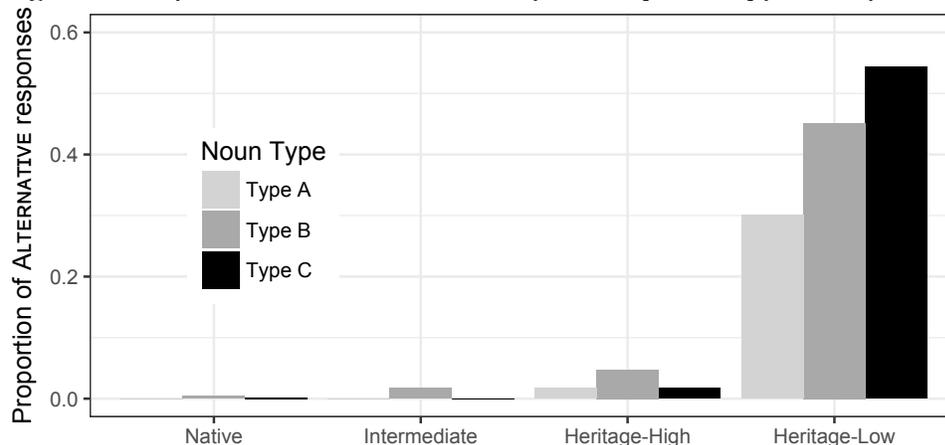
For TYPE C nouns, NATIVE and INTERMEDIATE groups do not produce any non-alternation form. Recall that TYPE C nouns end in a coronal obstruent and alternate in the standard form (e.g., /pic/ [pit] ~ [pic-i] ‘debt’) but unlike TYPE B nouns, there is no dominant paradigm to induce analogical paradigm leveling. Instead, the frequent innovation is towards the [t] ~ [s] alternation paradigm (/pis/ [pit] ~ [pis-i] ‘debt’), which is the most frequent coronal obstruent-final noun type. HERITAGE SPEAKERS, in contrast, do produce non-alternating forms for TYPE C and this is especially the case for the low-proficiency group (HERITAGE-HIGH: 0.05; HERITAGE-LOW: 0.84). When you examine the individual patterns, nine out of 18 high proficiency heritage speakers produced non-alternating forms while all of the 11 low proficiency heritage speakers produced non-alternating forms.

To summarize, we found that heritage speakers do prefer non-alternating paradigms more frequently than native speakers. In the case of TYPE B nouns, the preference was exhibited as a higher rate of paradigm leveling. TYPE C nouns provide a more striking contrast; non-alternation is categorically banned in the native production but is frequently attested in the heritage production. We also found that the two heritage speaker groups behaved differently; while HERITAGE-LOW speakers produced non-alternations across all noun types, including TYPE C nouns, HERITAGE-HIGH speakers showed a transitional pattern, preferring non-alternations for Type B nouns more than native controls but producing alternations more reliably for Type C nouns.

## 4.2 Avoidance of Alternation

In this section, we consider to what extent the Heritage speakers' preference for non-alternation is a reflection of their grammar as opposed to an artifact of the experimental design. In other words, could it be that these low-proficiency speakers are mechanically adding suffixes to the prompt in the sentential completion task, which would give the appearance of preference for non-alternation rather than accessing their grammar in their production? In her study of Child Korean morpho-phonology, (Do 2013) observed that the younger speakers not only produced more non-alternations than the older speakers, but they also frequently used alternative lexical items or morpho-syntactic constructions to avoid the target structure involving alternating paradigms. Acquisition literature shows that this is a common strategy found in Child language; children avoid words that contain complex phonological structures they have not yet mastered, indirectly revealing their grammatical preferences (Byun & Tessier 2006). Inspired by these previous findings, we examined the ALTERNATIVE productions by heritage speakers more closely. Recall that this category includes forms where the speakers produced an alternative lexical or morpho-syntactic structure that avoids revealing their UR analyses. If these errors are a general function of proficiency, we expect that they are attested equally likely across all noun types. On the other hand, if these errors are in fact strategies to avoid a marked structure they have not mastered, i.e., alternating paradigms, they should be more common in noun types that heritage speakers have difficulties with. Figure 2 provides a breakdown of the proportion of ALTERNATIVE responses by the noun type and by the speaker group. Only HERITAGE-LOW speakers produced a substantial number of ALTERNATIVE responses and their distribution is not even across the three noun types, with TYPE C inducing more ALTERNATIVE responses than TYPE B, which in turn induces more ALTERNATIVE responses than TYPE A. This difference is found to be statistically significant based on a mixed-effects logistical regression (with a by-subject random intercept and slope). This finding provides a support for the argument that these heritage speakers indeed have a grammatical preference for non-alternating paradigms and for those cases where their grammatical preference and the target structure do not align, they employ an alternative production strategy that avoids producing the problematic structure. The conflict between the grammatical preference and the target structure is particularly severe for Type C nouns where the target structure strictly bans non-alternating paradigms, and this conflict gives rise to a higher rate of ALTERNATIVE productions.

**Figure 2.** Proportion of ALTERNATIVE responses by noun type and speaker group



### 4.3 Maximum Entropy Grammar

In this section, we present a sketch of Maximum Entropy Grammar models (Jäger 2004) of our data as implemented in *Praat* (Boersma & Weenink 2017). For our simulation, a representative subset of the production data was considered—/m/-final nouns representing Type A, /p<sup>h</sup>-final nouns representing Type B, and /s/-final nouns representing Type C and our analysis employs the following constraints listed in (15).

(15) Constraints

- a. OO-CORR: “Do not alternate.”
- b. IO-CORR: “Do not alter the input.”
- c. T→S: “Coronal obstruent-final nouns should alternate: [t]# ~ [s]V”
- d. REALIZE: “Morphemes have an appropriate surface exponence.”
- e. \*STRUC: “Avoid producing structures.”

OO-CORR prohibits alternation and IO-CORR prohibits changes from the input to the output. We include an anti-correspondence constraint, T→S, which requires nouns that end in [t] in an unsuffixed form to alternate, specifically to be realized as [s] in prevocalic positions. This constraint is an acquired constraint inducted from the statistical distribution of nouns in Korean (Albright 2008; Hayes 1998; Jun 2010; Kang 2003). A full analysis of the full data should include additional anti-correspondence constraints to be weighted differently reflecting the frequency of each alternation type in different suffix contexts but for space reasons, here we present a simplified version of the analysis employing T→S, which is the most prominent one to affect the noun paradigms. REALIZE is a cover constraint that penalizes an alternative lexical or morpho-syntactic structure as an exponent of the target noun-suffix combination. \*STRUC is a general constraint against structures, which motivates reductions or changes in the target structure and counter-balances REALIZE which prohibits omission of structures. The breakdown of the production data used in the current simulation is summarized in Table 5. NATIVE and INTERMEDIATE groups are combined as they behave comparably.

**Table 5.** Input data for the Maximum Entropy Grammar simulation

Response types	NATIVE/INTERMEDIATE			HERITAGE-HIGH			HERITAGE-LOW		
	Type A	Type B	Type C	Type A	Type B	Type C	Type A	Type B	Type C
NO ALTERNATION	238 (100%)	79 (33%)	0	249 (99%)	185 (76%)	13 (4%)	118 (77%)	70 (46%)	84 (39%)
ALTERNATION	-	159 (67%)	339 (100%)	-	43 (18%)	338 (95%)	-	11 (7%)	13 (6%)
ALTERNATIVE	0 (0%)	0 (0%)	1 (0%)	3 (1%)	14 (6%)	6 (2%)	36 (23%)	71 (47%)	119 (55%)

We illustrate our analysis using the constraint weights for the NATIVE/INTERMEDIATE speaker group as examples in **Error! Reference source not found.** In the derivation of each target word (a stem + suffix combination), we assume that along with the UR, the unsuffixed form of the noun serves as a base for the OO-CORR constraint. We consider two candidates (no alternation and alternative output) for Type A nouns and three candidates (no alternation, alternation, and alternative output) for Type B and Type C nouns. Figure 3 summarizes the constraint weights in the three simulated grammatical models for the three speaker groups and

Table 6 provides a summary of the output distribution predicted by the model, which shows a close fit with the input data.

(16) Maximum Entropy Grammar of NATIVE speakers

a. Type A noun: /cip-in/ ‘house, TOP.’

/pom + in/ (Base: [pom])	REALIZE	T→S	IO-CORR	OO-CORR	*STRUC	H	exp(H)
	16.328	12.087	7.183	6.491	3.672		
a. pom-in					-1	-3.672	100%
b. ∅	-1					-16.328	0%

b. Type B noun: /ap<sup>h</sup>-in/ ‘front, TOP.’

/ap <sup>h</sup> + in/ (Base: [ap])	REALIZE	T→S	IO-CORR	OO-CORR	*STRUC	H	exp(H)
	16.328	12.087	7.183	6.491	3.672		
a. ap-in			-1		-1	-10.856	33%
b. ap <sup>h</sup> -in				-1	-1	-10.163	67%
c. ∅	-1					-16.328	0%

c. Type C noun: /so-in/ ‘clothes, TOP.’

/os + in/ (Base: [ot])	REALIZE	T→S	IO-CORR	OO-CORR	*STRUC	H	exp(H)
	16.328	12.087	7.183	6.491	3.672		
a. ot-in		-1	-1		-1	-22.943	0%
b. os-in				-1	-1	-10.163	100%
c. ∅	-1					-16.328	0%

Figure 3. Proportion of response types by noun type and speaker group

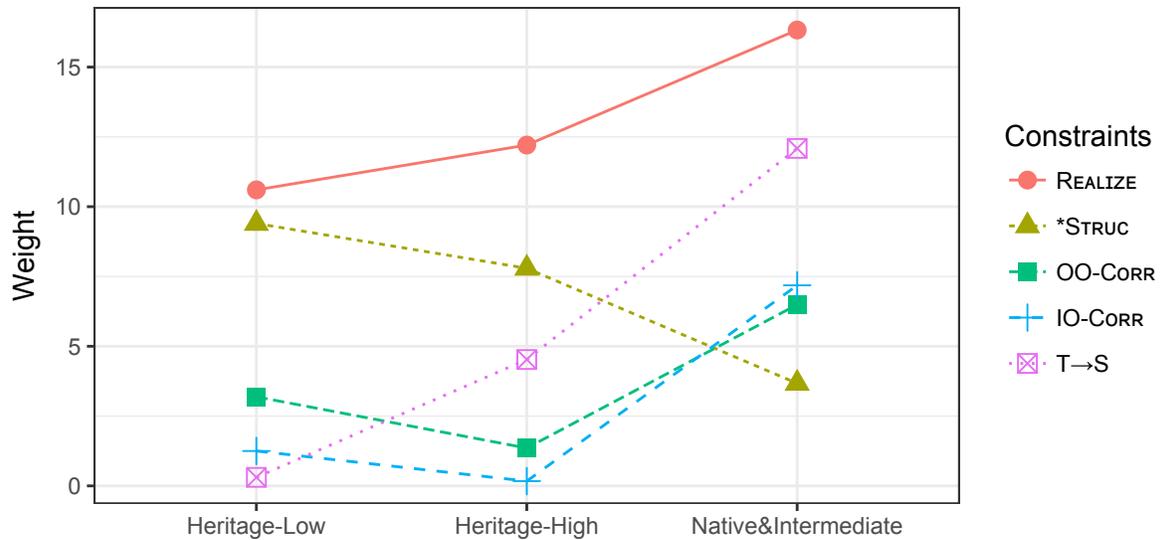


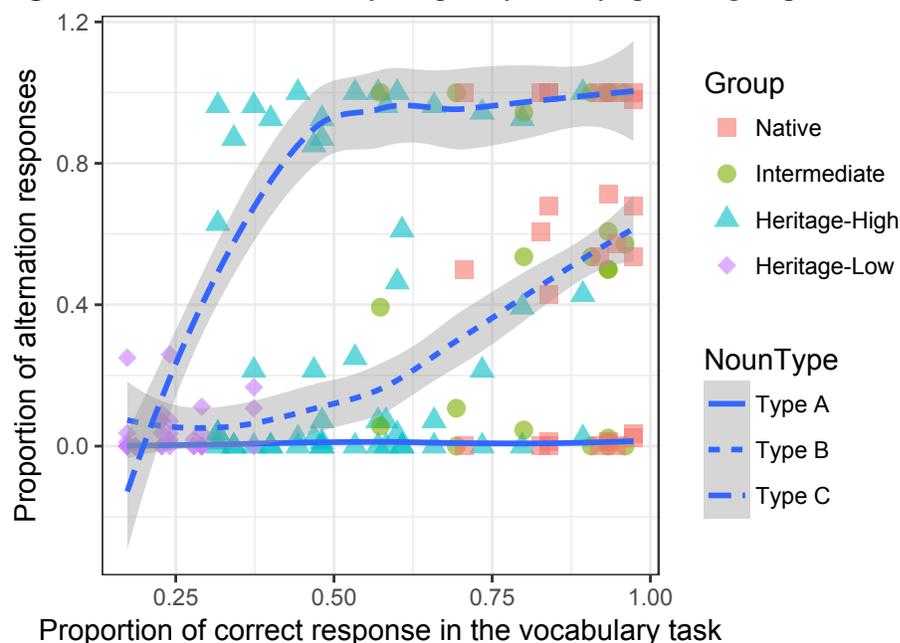
Table 6. Output distribution of the Maximum Entropy Grammar simulation

Response types	NATIVE/INTERMEDIATE			HERITAGE-HIGH			HERITAGE-LOW		
	Type A	Type B	Type C	Type A	Type B	Type C	Type A	Type B	Type C
NO ALTERNATION	100%	33%	0%	99%	76%	3%	77%	46%	38%
ALTERNATION	-	67%	100%	-	23%	92%	-	7%	7%
ALTERNATIVE	0%	0%	0%	1%	1%	4%	23%	48%	54%

It is instructive to look at the change in relative weights of the constraints across the grammars rather than the absolute weights. For HERITAGE-LOW speakers, REALIZE and \*STRUC have comparable rankings accounting for the frequent ALTERNATIVE realizations. The relative ranking of the two constraints diverges as we move from the HERITAGE-LOW to the HERITAGE-HIGH and the NATIVE & INTERMEDIATE groups, reflecting the more accurate realization of the target stem-suffix structure in the latter two groups. The choice between the alternation and no alternation outputs is accounted for by the relative ranking of OO-CORR, IO-CORR, and T→S. For the HERITAGE-LOW group, OO-CORR outranks the other two constraints and non-alternation is the dominant output across all noun types. For the HERITAGE-HIGH group, T→S is promoted above OO-CORR, which continues to outrank IO-CORR. This ranking produces the output where paradigm uniformity is enforced only in Type B nouns, where the anti-faithfulness constraint T→S is not relevant. Finally, in the NATIVE/INTERMEDIATE group, the OO-CORR constraint is demoted further and is ranked below IO-CORR and alternation is the majority output in both Type B and Type C nouns.

#### 4.4 Lexical Knowledge

In this section, we examine individual variation among the heritage speakers and see how their preference for paradigm leveling is correlated with their lexical knowledge, which is an indirect measure of exposure to the learning data and grammatical development (Polinsky 2006). Figure 4 plots individual speakers across all four speaker groups by their accuracy in vocabulary translation (x-axis) and their proportion of alternation responses for each noun type in the sentence completion task (y-axis). Type A nouns are produced without alternation by all speaker groups regardless of the speakers' level of vocabulary knowledge. These are nouns where the default preference (no alternation) and the learning data poses no conflict. Type B and Type C nouns are produced more accurately, i.e., with more alternation, by speakers with a higher level of vocabulary knowledge. This is expected under the view that the alternations are acquired patterns that develop as acquisition progresses. However, the developmental trajectory of the two noun types are different; the alternation in Type C nouns are acquired rather early, with speakers with vocabulary accuracy as low as 50% reaching the native speaker-like alternation rate but Type B nouns develop much later and require more vocabulary knowledge to approach a native-like production rate. These distinct developmental timelines indicate that the two nouns types are governed by different sets of rules/constraints that arise at different stages of acquisition. One important difference is that the evidence for alternation is categorical for Type C nouns while the evidence is variable and not categorical for Type B nouns. Future studies will simulate the exact mechanism of rule induction using the Minimal Generalization Algorithm (Albright & Hayes 2003) and examine to what extent the developmental trajectory of the noun (non-)alternation is derivable from the statistical induction of the input learning data, including the early preference for paradigm leveling. Specifically, we will examine if the paucity of learning input and the smaller vocabulary size of the heritage speakers (and for that matter child Korean speakers) can account for the seemingly emergent preference for paradigm leveling without explicitly including a grammatical constraint that promotes paradigm uniformity. Such a simulation study will provide a stronger test of the universal status of preference for paradigm uniformity.

**Figure 4.** Word translation by frequency and by speaker group

## 5. Conclusion

In this paper, we examined the production of consonant-final noun paradigms by Heritage Korean speakers. Specifically, we were interested in the status of paradigm uniformity as a universal grammatical preference and examined if and how the heritage speakers produce alternating noun paradigms. We found that the heritage speakers show a stronger preference for paradigm uniformity than native controls. In particular, they produce non-alternating paradigms not supported by the learning data in a straightforward way, namely /t/-final nouns. This preference for paradigm uniformity, however, is weakened (= they produce more alternation) as lexical knowledge increases. Alternation patterns supported by the learning data in a categorical manner (Type C) are accurately acquired before those that are variable (Type B). The findings of the study are compatible with Paradigm Uniformity as default grammatical preference that is ranked high in earlier stages of acquisition and is demoted as acquisition progresses. However, further studies involving careful modeling of the inductive mechanism that simulates the trajectory of lexical acquisition and alternation patterns is necessary to provide a more stringent test of the universal preference for paradigm uniformity.

## References

- Albright, Adam. 2008. Explaining universal tendencies and language particulars in analogical change. In Jeff Good (eds.), *Language Universals and Language Change*. Oxford: Oxford University Press.

- Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90. 119-161.
- Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai & Gabor Grothendieck. 2015. lme4 package, version 1.1-10 [Computer software]. Available from <https://cran.r-project.org/web/packages/lme4/>.
- Benua, Laura. 1997. Transderivational identity: phonological relations between words. Ph.D. dissertation, University of Massachusetts Amherst.
- Boersma, Paul & David Weenink. 2017. Praat: doing phonetics by computer [Computer program].
- Byun, McAllister & Anne-Michelle Tessier. 2006. Motor influences on grammar in an emergentist model of phonology. *Language and Linguistics Compass* 10. 431-452.
- Cho, Nam-Ho. 2002. *Hyentay kwuke sayong pinto cosa: hankwuke haksupyong ehwi sencengul wihan kicho cosa* [A survey of word frequency in Contemporary Korean]. Seoul, Korea: The National Institute of the Korean Language.
- Cho, Young-Mee Y. 2009. A historical perspective on nonderived environment blocking: The case of Korean palatalization. In Kristin Hanson & Sharon Inkelas (eds.), *The nature of the word: studies in honor of Paul Kiparsky*. Cambridge, MA: MIT Press.
- De Rosario-Martinez, Helios, John Fox & R\_Core\_Team. 2015. phia package, version 0.2-1. Available from <https://cran.r-project.org/web/packages/phia/>.
- Do, Young Ah. 2013. Biased learning of phonological alternations. Ph.D. dissertation, Massachusetts Institute of Technology.
- Fertig, David. 2016. Mechanism of paradigm leveling and the role of universal preferences in morphophonological change. *Diachronica* 33. 423-460.
- Garraett, Andrew J. 2008. Paradigm uniformity and markedness: historical convergence and universal grammar. In Jeff Good (eds.), *Explaining linguistic universals*, 125-143. Oxford: Oxford University Press.
- Hayes, Bruce. 1998. On the richness of paradigms, and the insufficiency of underlying representations in accounting for them. Handout for talk, Stanford University, April 2, 1998.
- . 2004. Phonological acquisition in Optimality Theory: the early stages. In René Kager, Wim Zonneveld & Joseph Pater (eds.), *Fixing priorities: constraints in phonological acquisition*, 158-203. Cambridge: Cambridge University Press.
- Jäger, Gerhard. 2004. Maximum entropy models and stochastic Optimality Theory. *Rutgers Optimality Archive ROA-625*. <http://roa.rutgers.edu>.
- Jun, Jongho. 2010. Stem-final obstruent variation in Korean. *Journal of East Asian Linguistics* 19. 137-179.
- Kang, Yoonjung. 2003. Sound changes affecting noun-final coronal obstruents in Korean. In W. McClure (eds.), *Japanese/Korean Linguistics*, 128-139.
- Kenstowicz, Michael. 1996. Base identity and uniform exponence: Alternatives to cyclicity. In Jacques Durand & Bernard Laks (eds.), *Current trends in phonology: models and methods*, 365-394. Salford: University of Salford Press.
- Kiparsky, Paul. 1982. Explanation in Phonology. In Stanley Peters (eds.), *Goals in linguistic theory*, Reprinted in *Explanation in phonology* (Dordrecht: Foris, 1982), chapter 1985. Englewood Cliffs, N. J.: Prentice-Hall.

- Ko, Heejeong. 2006. Base-Output Correspondence in Korean Nominal Inflection. *Journal of East Asian Linguistics* 15. 195-243.
- Ko, Kwang-mo. 1989. Cheyen kkuthuy pyenhwa t>s ey tayhan saylowun haysek [Explaining the noun-final change t>s in Korean]. *Enehak* 11. 3-22.
- McCarthy, John J. 1998. Morpheme structure constraints and paradigm occultation. In M. Catherine Gruber, Derrick Higgins, Kenneth Olson & Tamra Wysocki (eds.), *Proceedings of the Chicago Linguistic Society 35. Vol. II: The panels*, 123-150. Chicago: CLS.
- Montrul, Silvina. 2002. Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition* 5. 39-68.
- Nagy, Naomi. 2015. A sociolinguistic view of null subjects and VOT in Toronto heritage languages. *Lingua* 164, Part B. 309-327.
- Pater, Joe & Anne-Michelle Tessier. 2003. Phonotactic knowledge and the acquisition of alternations. *The Proceedings of Proceedings of the 15th International Congress on Phonetic Sciences*, Barcelona, Spain.
- Polinsky, Maria. 2006. Incomplete acquisition: American Russian. *Journal of Slavic Linguistics* 14. 191-262.
- . 2011. Reanalysis in adult heritage language: new evidence in support of attrition. *Studies in Second Language Acquisition* 33. 305-328.
- Polinsky, Maria & Olga Kagan. 2007. Heritage Languages: In the ‘Wild’ and in the Classroom. *Language and Linguistics Compass* 1. 368-395.
- R\_Core\_Team. 2015. R: A Language and Environment for Statistical Computing. Version 3.2.1. Vienna, Austria: R Foundation for Statistical Computing.
- Steriade, Donca. 2000. Paradigm uniformity and the phonetics-phonology boundary. In Janet Pierrehumbert & Michael Broe (eds.), *Acquisition and the lexicon*, 313-334. Cambridge: Cambridge University Press.
- Tessier, Anne-Michelle. 2012. Testing for OO-faithfulness in the acquisition of consonant clusters. *Language acquisition* 19. 144-173.
- White, James. 2017. Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a P-map bias. *Language* 93.