The Base of Korean Noun Paradigms: Evidence from Tone*

Youngah Do (Georgetown University), Chiyuki Ito (Tokyo University of Foreign Studies), and Michael Kenstowicz (Massachusetts Institute of Technology)

This paper reports and analyzes the tonal patterns that emerge in South Kyengsang monosyllabic nouns that exhibit two well-known analogical changes in stem shape, one involving coronal obstruent codas and the other stems with an underlying cluster. By the first change, underlying and orthographic /nac^h/ 'face' inflects as *nat*, *nac^h-ɨl* (conservative) or *nas-ɨl* (innovative); and by the second underlying /talk/ 'chicken' inflects as *tak*, *talk-ɨl* (conservative) or *tak-ɨl* (innovative). We find that many such nouns with a high-low tonal pattern change to high-high when inflected with the segmentally innovative stem. We propose that this tonal change supports the model of Korean noun paradigms proposed in Albright (2008) and Do (2013) in which the citation form serves as the base for the construction of the suffixed forms. If the base is a neutralization site, then learners select the alternant in which they have the greatest confidence of scoring a correct hit when undoing the neutralization.

1. Introduction

The citation form of a Korean noun ending in a coronal obstruent is a site for the neutralization of manner of articulation contrasts in which underlying /t, t^h, c, c^h, s/ are all realized as an unreleased stop [-t̚]. This alternation is motivated by Kim (1972)'s Principle of Closure that requires a coda consonant to bear the feature [-continuant], resulting in a phonetically unreleased plosive.

(1) Neutralization of coronal obstruent codas

Citation	nat̚	nat̚	nat̚	nat̚
Nominative	nas-i	nac^h-i	nac-i	nac^h-i
Accusative	nas-ɨl	nat^h-ɨl	nac-ɨl	nac^h-ɨl
	'sickle'	'piece'	'day'	'face'

Among the five possible alternants realized before the vowel-initial case inflections, *-t* and *-s* stand out. The former is found in just a single lexical item in the current standard language, while *-s* is the most frequent, accounting for over 72% in simplex nouns and 53% as the second member of a compound (Ito 2010: 361).

* A preliminary version of this paper was presented as a poster at the 87th Annual Meeting of the Linguistic Society of America, Boston, MA and the NINJAL International Conference on Phonetics and Phonology 2013, NINJAL (National Institute for Japanese Language and Linguistics), Tokyo.

Table 1. Lexical distribution of coronal obstruent codas

Coda	Simplex	Compound
/t/	1 (1%)	0 (0%)
/t^h/	17 (12%)	395 (16%)
/c/	8 (6%)	89 (4%)
/c^h/	14 (10%)	676 (27%)
/s/	103 (72%)	1,321 (53%)

Various researchers (Albright 2002a, 2008, Kang 2003, Ito 2010, Jun 2010, Sohn 2012) have observed that the alternation between [-t̄] and *-t*, *-t^h*, *-c*, *-c^h*, *-s* is unstable in nouns and many speakers substitute a consonant different from the one that is etymologically expected (and typically also reflected in the orthography). Among the *-t*, *-t^h*, *-c*, *-c^h*, *-s* alternants, *-s* is the most common substitute. It may be viewed as the default, surpassing Yang's (2005) threshold requirement, at least if based on simplex nouns, which are the topic of this investigation. Also, *-s* is the form that is uniformly substituted in loanwords: *k^hɛt*, *k^hɛs-i* 'cat', *t^hik^het*, *t^hik^hes-i* 'ticket' (Kenstowicz and Sohn 2001). Albright (2008) interprets this extension of the *-t̄* ≈ *-s* alternation as evidence for a model of morphophonological acquisition in which the learner selects a single slot of the paradigm as the base for all alternations depending on which one provides the greatest overall predictability for all phonological alternations.[1] In case this slot is a neutralization site for some phonological alternation, then the learner chooses the alternant in which he has the most confidence of scoring a correct "hit" among the various possible outcomes. Confidence is a composite of both type and token frequency. The confidence-based interpretation of the Korean coronal stems has been extended by Jun (2010) to account for variations in the next two most frequent coronal stem alternants *-c^h* and *-t^h* on the basis of the particular inflectional suffix. His major finding is that *-t^h* tends to be chosen before the dative/locative suffixes *-e(sə),* while *-c^h* is preferred before suffixes beginning with *-i,* such as the accusative *-il* and topic marker *-in*. Sohn (2012: 59-61) analyzes the latter change as a progressive extension of palatalization from the nominative *-i*.

Most of the research on the coronal coda alternation has focused on the standard Seoul dialect, which has lost the tonal distinctions of Middle Korean (15-16th c., MK). The Kyengsang dialects of southeastern Korea have preserved the MK tonal distinctions. In monosyllabic nouns, there are three inflectional tonal patterns illustrated in (2).[2] See Lee and Zhang (2014) for recent discussion.

[1] The allomorphy found in the accusative *-ril* ≈ *-il* and topic *-nin* ≈ *-in* suffixes neutralizes the contrast between l-final and n-final vs. vowel-final stems, rendering these case forms less informative than the citation form. See Albright (2008) for details.

[2] The transcription system for the examples in this paper is as follows. MK diphthongs have been monophthongized in Contemporary Korean (CK). Some symbols such as ㄹ, ㅇ are transcribed differently depending on environment. High tones are marked by the acute accent while a low tone syllable is unmarked. Other Korean words in the body of the paper follow the Yale Romanization system, except for some authors' names, which follow their customary spellings.

ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ	ㅅ	ㅆ	ㅇ	ㅈ	ㅉ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
k	k*	n	t	t*	r/l	m	p	p*	s	s*	Ø/ŋ	c	c*	c^h	k^h	t^h	p^h	h

(2) Inflectional tonal patterns of monosyllabic nouns[3]

	H(H)	H(L)	L(H)
Citation	múl	súl	tŏn
Nominative	múl-í	súl-i	ton-í
Accusative	múl-ɨl	súl-ɨl	ton-ɨl
	‘water’	‘wine’	‘money’

The citation form is a neutralization site for tone (Chang 2007, Kenstowicz et al. 2007; see Appendix D for further discussion of this point). In particular, nouns exhibiting a high tone (H) in citation alternate with either a double-H in inflection (e.g. *múl-í*) or with a simple H(L) (L = low tone) pattern (e.g. *súl-i*). The neutralization of the H(H) and H(L) classes in the citation form may be a recent phenomenon, given that earlier studies (e.g. C-K. Kim 1975) report that the double-H inflected stems were realized with a mid tone in isolation (e.g. *mūl*), while H(L) stems were realized with a high tone (e.g. *súl*). The coronal coda alternations, on the other hand, have been documented at least since MK (the 15th century).[4]

Given that the H(H) and H(L) tonal patterns are neutralized in the citation form, a learner of the Kyengsang dialect faces the same dilemma as in (1) if the citation form is the base: should the inflected form be H(H) or H(L)? Moreover, since the H(H) class is twice as large as H(L) in monosyllabic native nouns (see discussion below), there is a comparable difference in reliability between the two alternations, such that H(H) is more likely to lead to a correct hit than H(L), if the learner starts from the ambiguous isolation form. We might therefore expect H(H) to attract lexical items from the H(L) class similar to the way in which the *-s* has done for the segmental alternation. In addition, to the extent that there are other clues from the segmental phonology (“islands of reliability”, Albright 2002c) as to which accent class a given stem belongs, we might expect to see speakers employing them as well when faced with the task of undoing the neutralization.

In this paper, we report and discuss data from South Kyengsang tonal inflection to explore these questions. In section 2, we review the results of the data collected and analyzed in Do et al.’s (2014) survey of five contemporary South Kyengsang speakers. The relative sizes of the three accent classes and their correspondences with MK are established. In section 3, we report and analyze the results of a survey of 12 South Kyengsang speakers focused specifically on the accent assigned to monosyllabic nouns with an obstruent coronal coda. We collected two pieces of information: first, the tonal pattern that is assigned when these nouns are inflected with the coronal consonant that is found in the spelling,

---

	ㅏ	ㅐ	ㅑ	ㅓ	ㅔ	ㅕ	ㅖ	ㅗ	ㅘ	ㅙ	ㅚ	ㅛ	ㅜ	ㅝ	ㅞ	ㅟ	ㅠ	ㅡ	ㅢ	ㅣ	ㆍ	ㆍㅣ
CK	a	ɛ	ja	ə	e	jə	je/e	o	wa	wɛ	we	jo	u	wə	we	wi	ju	ɨ	ɨj/i	i		
MK	a	aj	ja	ə	əj	jə	jəj	o	wa	waj	oj	jo	u	wə	wəj	uj	ju	ɨ	ɨj	i	ʌ	ʌj

[3] Monosyllabic loanwords show a fourth type, with a falling tone in both citation and inflection (Kenstowicz and Sohn 2001, N-J. Kim 1997, M. Kim 1997).

[4] In the 15th century, *-t* and *-t^h* were neutralized to *-t* while *-s*, *-c*, and *-c^h* were neutralized to *-s*. The *-t* vs. *-s* distinction started to be lost in the 16th century and the merger was completed in the 18th century. See K-M. Lee (1972) and Ito (2010) for details.

and second the tonal pattern in any innovative form when the nouns are inflected with a consonant that diverges from the standard orthographic form. A robust generalization emerges in which H(L) stems that inflect with a conservative coda conforming with the orthography change to H(H) when inflected with an innovative coda consonant. The lexical frequency and speaker age parameters of this change are presented (section 4). We then turn in section 5 to a similar survey and analysis of the inflectional accent associated with the handful of nominal stems with underlying coda clusters such as /talk/ 'chicken' which inflect either with preservation of the cluster (as in conservative *talk-i*, *talk-il*) or with a leveling of the isolation form (as in innovative *tak-i*, *tak-il*). A similar tendency is found in which the accent changes from H(L) to H(H) when the segmentally innovative stem form is inflected. In section 6, we return to the coronal coda alternation and its accentual correlates. We note certain correlations between particular coronal coda consonants and the inflectional accent pattern that appear as frequency biases in the lexicon and are projected in novel word experiments. These islands of reliability are analyzed in terms of lower confidence statistics. The paper then discusses whether the frequent co-occurrence of the segmental change in coronal codas and the accent change from H(L) to H(H) observed in section 2 can be attributed to lexical schemata correlating the H(L) vs. H(H) accents with particular codas. No statistically reliable result emerges here; hence, it is tentatively concluded that the change from H(L) to H(H) with innovative codas reflects the speaker's knowledge of the overall frequency of H(H) in the lexicon comparable to the frequency of *-s* as the default coronal coda obstruent. This in turn is taken as support for the model of acquisition of Korean noun inflection proposed in Albright (2008) and further developed in Do (2013), in which the isolation form serves as the base for nonautomatic morphophonological alternations. Section 6 then discusses the accentual correlates of the innovative CVC stems in the same terms; in this case the proper interpretation of the accent change for this segmental analogy is less clear. Section 7 compares the segmental and accentual analogies discussed in the paper, noting that the former have proceeded further. This is attributed to a larger range of neutralization sites compared to the accent changes. Section 8 is a summary and conclusion. Appendices A-C contain the experimental materials used in this study, and Appendix D reports pitch tracking data elicited from six South Kyengsang speakers showing a (near) merger of the H(L) and H(H) accents in the isolation form.

2. Background

As part of a more general study of the South Kyengsang tonal classes (Do et al. 2014), five native South Kyengsang speakers were polled for their tonal class assignment for c. 1,900 native nouns, including c. 470 monosyllables. We show the size of the three tonal classes for monosyllables summed over these five speakers in Table 2. (Not every word was known to each speaker and hence the number of responses is less than 2,350). The H(H) class is approximately twice as large as H(L).

Table 2. Accent distribution of monosyllabic native nouns

Tonal class	Number	Ratio
H(H)	1,133	53%
H(L)	548	26%
L(H)	450	21%
Totals	2,123	

A second relevant fact can be gleaned from Table 3 below. It shows the correspondence rates between the South Kyengsang monosyllables and their MK cognates (for nouns that have an MK cognate whose tone is known). Again, these data represent nouns collected from the five South Kyengsang speakers in Do et al. (2014). The shaded numbers indicate the regular phonological correspondences. The point of interest here is the exceptions to the regular development. The percentage of words derived from the MK L class that should have ended up in the H(L) class but instead appear in the H(H) class is almost three times larger than the percentage of MK H words that should have ended up in the H(H) class but instead appear in the H(L) class: 21% vs. 8%. In other words, the H(H) class has attracted a larger proportion of words from the H(L) class than vice versa. This makes sense, given that the H(H) class is the largest class. If a speaker fails to hear the tone properly and must guess as to which class a noun belongs, then H(H) is a safer choice, given that it is approximately twice as large as H(L).

Table 3. MK–South Kyengsang (SK) correspondences (monosyllabic native nouns)

MK \ SK	H(H)	H(L)	L(H)	Totals	H(H)%	H(L)%	L(H)%
H	693	67	57	817	84%	8%	7%
L	68	231	20	319	21%	72%	6%
R	55	31	179	265	21%	12%	68%
Totals	816	329	256	1,401			

3. Accent in coronal coda nouns

We now turn to the interaction between the segmental and tonal alternations. If the extension of the *-t̚* ≈ *-s* alternation is evidence that the speaker is taking the citation form as the base from which the inflected form is derived, and if the citation form serves as the base for all alternations, then we predict that for nouns with coronal obstruents in the H(L) class, speakers should also tend to change the accent to H(H), since it represents a more reliable choice than H(L).

In order to investigate this question, 12 South Kyengsang (Pusan) speakers (seven females, five males, age range 26 to 59) were polled for the 15 monosyllabic nouns in Table 4 ending in a coronal obstruent (*-c*, *-c*^{h}, *-t*^{h}) that belonged to the L tonal class in MK, and hence are expected to inflect as H(L) in South Kyengsang.

Table 4. List of native nouns with a coronal obstruent coda (*-c*, *-c^h*, *-t^h*)[5]

Inflected	Citation		MK form	MK accent
pjət^h	pjət̚	'sunshine'	pjət^h	L
tac^h	tat̚	'anchor'	taT	L
mut^h	mut̚	'land'	mut^h	L
nac^h	nat̚	'face'	nʌc^h	L
sot^h	sot̚	'iron pot'	sot^h	L
juc^h	jut̚	'Korean traditional game'	zuS	L
oc^h	ot̚	'lacquer'	os	L
pat^h	pat̚	'field'	pat^h	L
toc^h	tot̚	'sail'	tosk	L
kət^h	kət̚	'surface'	kəc^h	L
k*oc^h	k*ot̚	'flower'	koc	L
mit^h	mit̚	'bottom'	mit^h	L
suc^h	sut̚	'charcoal'	susk	L
kjət^h	kjət̚	'neighborhood', 'side'	kjət^h	L
pic	pit̚	'debt'	pit	L

The questionnaire (Appendix A) was distributed to South Kyengsang speakers by e-mail. In the questionnaire, the speakers were asked to read a sentence in which each noun with a coronal coda appeared in the accusative case form with the suffix *-ɨl*. They were asked to report the pronunciation that they prefer most as well as its accent pattern by choosing among H(L), H(H), L(H), etc. (e.g. /jəp^h-ɨl/ 'side' + acc., [jə.p^{h}ɨl], H(H)). They were then asked whether they use an alternative form of the stem other than the one they already reported, and if so, then to read the same sentence with this form and report its accent pattern (e.g. /jəp^h-ɨl/, [jə.pɨl], H(L)).[6] See Appendix B for a sample response.

The results from our subjects appear in Table 5 below. The "conservative" row shows the number of cases that were inflected according to the standard etymological coronal indicated in the spelling, while the "innovative" row indicates a non-etymological alternant at variance with the spelling—in the vast majority of cases *-s*. The H(L) > H(L) column indicates the number of choices which retained

[5] The T and S in the MK form indicate words whose exact coda is unknown (T = *-t* or *-t^h* and S = *-s*, *-c*, *-c^h*, or *-sk*) since they are not attested in an inflected form in MK texts. In the survey, *kjət^h* 'neighborhood' was tested separately from *kjət^h* 'side'; the former is an abstract derivative of the latter. They showed almost the same patterns in coda pronunciation and accent: among 48 cases (conservative coda, conservative tone, innovative coda, innovative tone × 12 speakers), 43 cases agreed (90%). Thus, we include only the result of *kjət^h* 'neighborhood' in our results; it was elicited first in the experiment.

[6] Although the speakers were instructed to report the most preferred form first and any alternative form next, the one they reported first was always the one based on the orthography. This may be because the sample question given in the instructions for the questionnaire showed a segmentally conservative form followed by an innovative alternant: /jəp^h.ɨl/ 'side' + acc., [jə.p^{h}ɨl], H(H), /jəp^h.ɨl/, [jə.pɨl], H(L). The fact that all the speakers reported using the conservative forms is somewhat surprising, given that many South Kyengsang speakers use the innovative forms more frequently in actual speech, at least for some words. This may be an experimental effect in which speakers were biased to report the conservative orthographic forms as the "correct" form.

the etymologically expected H(L) tonal contour, while H(L) > H(H) indicates the responses in which the H(H) tonal pattern was assigned.[7]

Table 5. Variation of coronal obstruent codas and tonal patterns

	H(L) > H(L)	H(L) > H(H)	Totals
Conservative	167 (85%)	29 (15%)	196
Innovative (-t > -s)	51 (27%)	139 (73%)	190
Totals	218	168	386

We see that in the conservative pronunciation the vast majority of speakers (85%) have retained the etymologically expected H(L) inflection; 15% have the H(H) contour, which approximates the 21% change we have seen in our more general survey for all monosyllables summarized in Table 3. The striking finding is that when our speakers inflected the coronal coda noun with the innovative coda (principally *-s*), then 73% also changed the tonal pattern to H(H). Pearson's Chi-square test with the Yates' continuity correction showed a highly significant result ($\chi^2 = 131.3238$, df = 1, $p < 2.2e\text{-}16$). The magnitude of this change is much larger than the 21% change from our general survey for all monosyllables seen in Table 3 and suggests that an innovation at the segmental level tends to be accompanied by a tonal change. This is expected under the assumption that the learner is constructing the inflected form on the basis of the citation form for both alternations. If he starts from the citation form, then he must undo the neutralizations and reason probabilistically as to which segmental and tonal alternant is most likely to result in a correct hit for the inflected form. The *-s* is the most frequent segmental alternant, and H(H) is the most frequent tonal alternant, and so the fact that they tend to frequently occur together now makes sense. If this reasoning is correct, it supports Albright's (2008) 'single-base hypothesis' that the learner constructs the rest of a paradigm by starting from a particular paradigm cell for all morphophonological alternations.

However, due to the idiosyncrasies of the South Kyengsang lexicon, an alternative, more "product-based" (Phillips 2006) explanation is also possible. Most monosyllabic nouns with coda *-s* in the conservative (standard) form belong to the South Kyengsang H(H) accent class, reflecting a bias that was present in Middle Korean. We show these below.

Table 6. Monosyllabic native nouns with a coda *-s*[8]

Standard forms		South Kyengsang accent	MK form	MK accent
jəs	'taffy'	H(H)	jəs	H
pis	'comb'	H(H)	pis	H
pus	'brush'	H(H)	put	H
kas	'leaf mustard'	H(H)	kas	H

[7] For innovative forms a few speakers assigned both H(H) and H(L) options for a few words, ten in all. These were counted twice, once as H(H) and once as H(L). Some speakers reported that innovative forms are not possible for the word *pic* 'debt'.

[8] The South Kyengsang accent column shows the (assumed) most conservative accent class taken from the five consultants in Do et al. (2014) or from previous studies (He 1955, Ramsey 1978).

kas	'hat'	H(H)	kat	H
kis	'feather'	H(H)	cic^{h}	H
kis	'collar'	H(H)	kic	H
kos	'place'	H(H)	kot	H
cəc	'pickled fish'	H(H)	cəs	H
mas	'taste'	H(H)	mas	H
mos	'pond'	H(H)	mos	H
mos	'nail'	H(H)	mot	H
nas	'sickle'	H(H)	nat	H
os	'clothes'	H(H)	os	H
kus	'exorcism'	H(H)	kus	H
t*ɨs	'mind'	H(H)	ptɨt	H
pjəs	'cockscomb'	H(L)	pjəs/pjəc^{h}	H
t^{h}as	'fault'	H(L)	tas/t^{h}as	L
tos	'mat'	H(L)	tosk	L
pəs	'friend'	L(H)	pət	R
cas	'pine nuts'	L(H)	cas	R
cis	'behavior'	L(H)	cɨz	R
nes	'four'	L(H)	nəjh	R
ses	'three'	L(H)	səjh	R

Thus, when the speaker inflects a noun with the innovative variant in *-s*, then the word could fall under the lexical schema that "monosyllables in *-s* inflect as H(H)". In other words, the co-occurrence of the tonal and segmental changes has two possible interpretations. First, it may be due to the same mechanism in which the citation form is the base for two separate analogical changes to the higher type-frequency defaults of *-s* and H(H). Alternatively, it may be due to two different mechanisms that take place sequentially: segmental changes such as *-t^{h}* > *-s* occur first based on the coda merger in citation form and then the tonal patterns are restructured based on local associations ("islands of reliability", Albright 2002c) between the stem-final consonant and the accent such as *-s* ≈ H(H). For the latter change, no appeal to the citation form is needed; just the knowledge that *-s* final nouns inflect as H(H) with high probability. The choice between these two alternative interpretations depends on how well each does in predicting the experimental results as well as more generally the form in which the speaker's statistical knowledge of the lexicon is encoded. We return to these questions in section 6.

Finally, we note that our data is also consistent with Jun's (2010) generalization that *c^{h}* is favored over *t^{h}* before suffixes beginning with *-ɨ* such as the accusative. For stems with underlying *-c^{h}*, 95 innovating responses changed to *-s* and only one to *-t^{h}*; but for stems with underlying *-t^{h}*, 46 changed to *-s* while 26 appeared as *-c^{h}*. And among the latter, 16 take the innovative H(H), while 12 retain the H(L) accent.

In sum, our speakers exhibit a striking connection between the tonal and segmental inflection for the monosyllabic coronal stems. A shift from the conservative (orthographic standard) segmental inflection to the innovative one (principally in *-s*) significantly correlates with a shift from the etymologically expected H(L) inflection to H(H).

4. Frequency

Kang (2003) and Sohn (2012) find that the innovative -*t*̄ -> -*s* inflection tends to affect lower frequency words before it affects higher frequency ones. This is the expected profile of a lexical/analogical change, which adjusts the distribution of lexical items with respect to two or more entrenched word classes like the strong and weak verbs of English (Phillips 1984, 2006; Bybee 2001). Such changes are more likely to affect less frequently used words, since the speaker has less evidence as to which class a given word belongs compared to more frequently used words. We therefore expect a similar outcome with the H(L) to H(H) tonal change. Figure 1 below shows a fairly good correlation between the log-frequency of our words from Table 4 in the King Sejong corpus (Kim and Kang 2000) and retention of the H(L) inflectional tone in the presence of the -*t*̄ -> -*s* segmental change. The vertical axis indicates the number of responses where a given word appeared as H(L) (excluding the H(H) ~ H(L) answers) in the company of the segmental change and the horizontal axis indicates the word's relative frequency in the King Sejong corpus. The data suggest that the more frequent the word, the more likely it is to retain the H(L) accent; conversely, the less frequent the word, the more likely it is to change to the H(H) pattern in inflection. A linear regression analysis with the number of H(L) responses as the dependent variable and log frequency as the independent variable shows a highly significant result (estimate = 2.1033, std. error = 0.4651, t-value = 4.523, p = 0.000573).

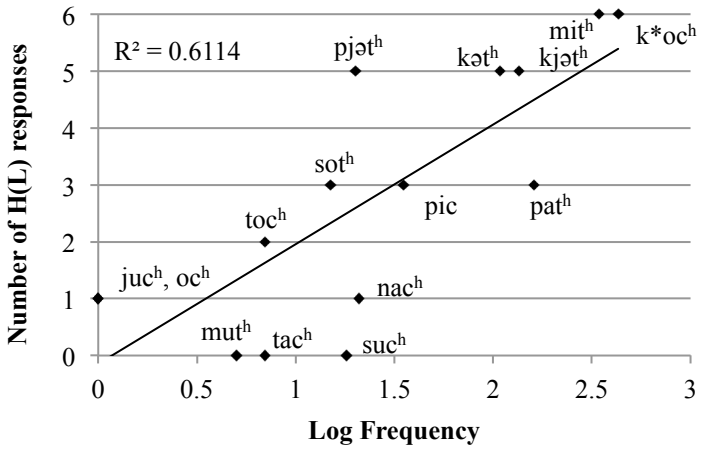


Figure 1. Retention of the conservative tonal pattern and frequency (monosyllabic native nouns with a coronal obstruent coda)

Figure 2 shows a strong correlation between speaker age and the substitution of H(H) for H(L) (again excluding the H(H) ~ H(L) answers).[9] The vertical axis indicates the number of items in our poll for which a particular speaker retains the H(L) inflectional accent in the company of an innovative coronal coda, and the horizontal axis is the speaker's age. The older the speaker, the more the H(L) pattern is retained; conversely, the younger the speaker, the more likely the H(H) pattern will

[9] Due to the fact that two pairs of speakers had the same age and an identical number of H(L) responses, the figure does not show 12 datapoints.

be assigned to the inflected form. A linear regression analysis with the number of H(L) responses as the dependent variable and speaker age as the independent variable shows a significant result (estimate = 0.25675, std. error = 0.05006, t-value = 5.129, p = 0.000445). However, this result should be interpreted cautiously given the small sample size, which moreover has a preponderance of younger speakers.

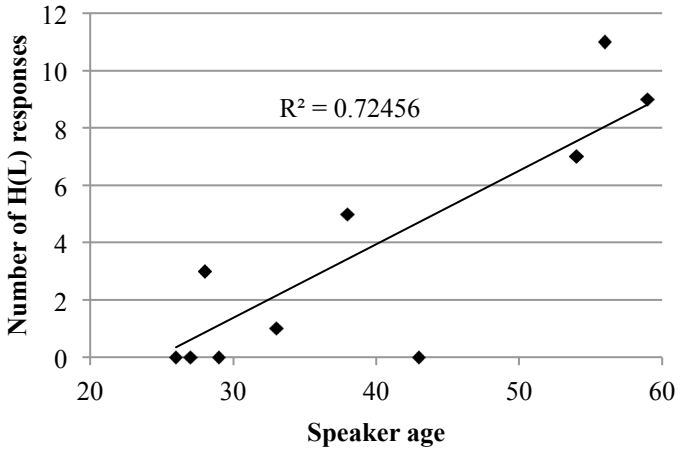


Figure 2. Retention of the conservative tonal pattern and speaker age (monosyllabic native nouns with a coronal obstruent coda)

5. CVCC nouns

In this section, we turn to another alternation in the inflection of Korean nouns in which the citation form is the locus of an analogical extension in its segmental phonology, and explore its tonal reflexes in the South Kyengsang dialect. Korean has a handful of monosyllabic noun stems ending in a cluster. The cluster shows up in inflection, while it is simplified in the citation form to conform to the language's CVC syllable template. For many speakers, there is an alternative inflected form based on the CVC shape appearing in citation (Kenstowicz 1997). We indicate these two possible inflections as "conservative" and "innovative" in (3).

(3) Segmental alternations of cluster codas

Citation			Nominative	Accusative	Locative
kap	'price'	Conservative	kaps-i	kaps-ɨl	kaps-e
		Innovative	kap-i	kap-ɨl	kap-e

If the inflected form is being derived from the citation form, then we predict that H(L) CVCC nouns which level the isolation alternant (e.g. *kap*) should tend to shift to the South Kyengsang H(H) tonal pattern when they are inflected, since the segmental change indicates that the speaker is taking the citation form as the base, and H(H) is more likely to lead to a correct hit compared to H(L).

In order to test this prediction, we polled 15 South Kyengsang speakers (eight females, seven males, age range 21 to 83) for the 5 CVCC stems in Table 7 below that belonged to the Middle Korean L class (= South Kyengsang the H(L) class). The same methodology as used in section 3 was employed, except that the alternation of the CVCC stems was tested based on the nominative case form. See Appendix C for the list of sentences used for the experiment.

Table 7. Monosyllabic native nouns with a cluster coda

Inflected	Citation		MK form	MK accent
talk	tak̚	'chicken'	tʌlk	L
hɨlk	hɨk̚	'soil'	hʌlk	L
nəks	nək̚	'soul, ghost'	nəks	L
saks	sak̚	'pay, wages'	saks	L
salk	sak̚	'wildcat'	sʌlk	L

The results appear in Table 8 below. Row CVCC > CVCC indicates the conservative inflection that retains the cluster reflected in the orthographic standard language form, while CVCC > CVC indicates the innovative inflection that levels the CVC alternant appearing in citation. The H(L) > H(L) column designates an inflected form that retains the etymologically expected H(L) tonal contour, while H(L) > H(H) marks a change to the larger H(H) class.

Table 8. Variation of cluster codas and tonal patterns

	H(L) > H(L)	H(L) > H(H)	Totals
CVCC > CVCC	67 (92%)	6 (8%)	73
CVCC > CVC	38 (52%)	35 (48%)	73
Totals	105	41	146

When the noun is inflected with the original CVCC cluster alternant, only 8% of responses show the innovative H(H) tonal contour. But when the noun is inflected with the innovative CVC stem alternant found in the citation form, then the proportion of words inflecting with the H(H) contour jumps to 48%, almost six times larger. Thus, once again, our main hypothesis is supported: a change of tonal class tends to accompany a change in the segmental shape of the stem when it is inflected. This coincidence makes sense under the hypothesis that in constructing the inflected form, the learner proceeds from the citation form. When that form is a neutralization site for tone, then the learner selects the inflectional alternant that is most likely to lead to a correct hit.[10]

[10] In actuality our test of CVCC stems included four other words *moks* 'share', *kops* 'double', *kaps* 'price', and *c*^{h}*ilk* 'arrowroot', which have the CVC alternates *mok*, *kop*, *kap*, *c*^{h}*ik*, respectively (note that for the word 'double', *kop* is the standard form and *kops* is dialectal). These words are excluded from the discussion here because the MK accents of *moks* 'share' and *kop* 'double' are unattested, and the MK accents of *kaps* 'price' and *c*^{h}*ilk* 'arrowroot' were H, which corresponds with the South Kyengsang H(H) class. The MK accent of the former two words was probably L, given that they tend to appear with the MK L class reflex in North Kyengsang, South Kyengsang, and Yanbian Korean. In fact, for our speakers they

Is appeal to the citation form really necessary here or could the tonal change result from a product schema over the output? As seen in Table 7, all five CVCC stems terminate in *-k* in the simplified form and hence we might expect them to inherit any tonal preferences that *-k* final stems exhibit in the lexicon. The *-k* final monosyllabic stems in our corpus from Do et al. (2014) approximate the distribution of the H(H) and H(L) classes for all monosyllables: H(H) (63%) vs. H(L) (35%) and show a strong dispreference (2%) for the L(H) rise class (as do most obstruents). Thus, here as well, the distribution of the innovative inflected forms with respect to the three tonal classes seen in (2) approximates the distribution of monosyllabic *-k* final forms in the lexicon as a whole. Consequently, rather than appealing to the general distribution of the H(H) vs. H(L) accent patterns, speakers could be referring to the particular stem-final consonant in deciding what accent pattern to assign to an innovating form. Taking this alternative seriously raises the questions of what are the statistics of the lexicon concerning the relation between the coda consonant and the tonal patterns, do speakers evidence knowledge of these relations, and how is this knowledge encoded? We touch on these questions in the next section.

6. Coda consonant – inflectional tone correspondences

In order to address the more general question of speakers' knowledge of the statistics of the lexicon concerning the inflectional tonal pattern vis-à-vis the stem-final consonant, we bring the results of a more general study to bear on this question. Do et al. (2014) conducted a novel word experiment with 43 South Kyengsang speakers that investigated the tonal assignments for 92 nonce words (40 disyllables, 40 trisyllables, and 12 monosyllables). The subjects were shown the nonce word in a sentence frame in the accusative case on the computer screen in Hangul and were asked to choose among several possible tonal patterns for the word. For monosyllables, there were four choices: H(H), H(L), L(H), and other. Among the monosyllables, eight terminated in an obstruent and four in a sonorant. The eight obstruent stem nonce words were *nuk-, kək-, poc^h- nəp-, map-, tus-, tes-,* and *cus-*. The findings for obstruent consonants are shown below. They evidence a general preference for H(H), except in the case of the aspirate *-c^h* (= *poc^h-*)

---

alternated in a manner similar to the other MK L class words discussed in this section: conservative (*moks*, *kops*) = H(H) 0, H(L) 30, innovative (*mok*, *kop*) = H(H) 15, H(L) 15. On the other hand, *kaps* 'price' and *c^{h}ilk* 'arrowroot' are different in that both conservative and innovative forms appeared with H(H) at a much higher ratio: conservative cluster forms (*kaps*, *c^{h}ilk*) = H(H) 15, H(L) 15; innovative forms (*kap*, *c^{h}ik*) = H(H) 30, H(L) 0. Interestingly, in the CVCC ≈ CVC analogy for *kaps* and *c^{h}ilk*, the segmentally conservative forms show an irregular accentual correspondence for MK H tone as H(L). This may be due to analogical change based on an association between the coda cluster and the South Kyengsang H(L) class. In fact, the MK form of contemporary *moks* 'share' was *mok* (no cluster, accent information unknown) and the MK verbal stem that was the cognate of contemporary *kop/kops* 'double' was *kòp-* 'to double' (no cluster). This suggests that these stems changed their coda from simplex CVC to a cluster (*-ks*, *-ps*) based on the CVCC–H(L) association mentioned above. See section 6 for relevant discussion.

Table 9. Results of a novel word experiment (monosyllabic words with an obstruent coda)

Accent \ Coda	p	s	c^h	k	p%	s%	c^h%	k%
H(H)	56	77	15	54	65%	60%	35%	62%
H(L)	7	28	16	21	8%	22%	37%	24%
L(H)	21	16	3	10	24%	12%	7%	12%
Other	2	8	9	1	2%	6%	21%	2%

To what extent do these novel word preferences correlate with the monosyllabic South Kyengsang lexicon? Table 10 shows the frequencies of these codas for the three tonal classes summed across the five South Kyengsang speakers in Do et al.'s (2014) survey. They also show the general bias for H(H) except in the case of -c^h, where H(L) is the more frequent type. The correlation between the distributions in Tables 9 and 10 (excluding "Other" in Table 9) is significant (spearman's rank correlation rho, S = 66, p = 0.005253).

Table 10. Frequency of some monosyllabic codas in the South Kyengsang lexicon

Accent \ Coda	p	s	c^h	k	p%	s%	c^h%	k%
H(H)	34	120	25	93	69%	72%	39%	63%
H(L)	15	30	36	51	31%	18%	56%	35%
L(H)	0	16	3	3	0%	10%	5%	2%

Furthermore, there are indications of a more general connection between an aspirate coda and the H(L) tonal pattern in South Kyengsang: -p^h appears with H(H) 45% vs. H(L) 55%, and -t^h appears with H(H) 32% vs. H(L) 67%.[11] Also several MK words in the L class, which, we recall, is regularly reflected as H(L) in Kyengsang, have irregularly changed their coda consonant to an aspirate in inflection: MK *òs* > South Kyengsang *óc^h-ɨl* 'lacquer', MK *kòc* > South Kyengsang *k*óc^h-ɨl* 'flower', MK *njàp* > South Kyengsang *jə́p^h-ɨl* 'side', MK *sàp* > South Kyengsang *sə́p^h-ɨl* 'brushwood'. For these cases, it looks as if the correlation between an aspirate coda and H(L) tone has resulted in a change in the laryngeal feature of the consonant.[12]

We might assume that these segmental changes as well as the accent choices in the novel words are based on the strength of the correlation between coda type and accent class: if a segment such as -*s* or -c^h is highly overrepresented in a certain accent class, then words with that coda may change to that accent class, even if the accent class itself is not a high type-frequency class as a whole. In order to pursue this point, we investigated the confidence scores for each coda-accent correlation from the monosyllable data collected in Do et al.'s (2014) survey of five South Kyengsang speakers by using corrected lower confidence limit statistics (Mikheev 1997, Albright and Hayes 2002, Albright 2002a, 2002b, 2008). The confidence value α is 0.75. The coda-accent correlation is based on the

[11] Monosyllabic native nouns with a coda -k^h are not attested.

[12] It is worth observing that the lexical bias of aspirated consonants for a following low tone in inflection stands in contrast to a more general raising of F0 after aspirates seen at the beginning of the word in Korean including Kyengsang (Kenstowicz and Park 2006, Lee and Jongman 2012).

contemporary Korean (conservative) coda consonant and the MK accent class, since this presumably represents the conservative stage of the current South Kyengsang dialect before the coronal coda and inflectional accent changes investigated in section 3 entered the language (= Pre-Kyengsang). For example, there are 9 words with a coda *-p*, among which 7 words correspond with the MK H class. The corrected lower confidence limit of this correlation (coda *-p* and the H class) is: (Hit/Scope) − (Standard Error × Two-tailed z value) − (0.5/Scope) = (7/9) − (0.14 × 1.15) − (0.5/9) = 0.563. Table 11 shows the confidence scores of each correlation calculated this way. Complex codas are excluded here.

Table 11. Confidence scores of coda-accent correlation

Pre-Kyengsang	Hit	Scope	Corrected lower confidence limit
V-H	61	90	0.616
s-H	21	29	0.611
ŋ-L	11	15	0.569
p-H	7	9	0.563
l-H	38	62	0.534
m-H	27	43	0.531
n-H	9	13	0.507
k*-L	1	1	0.500
c^h-L	8	12	0.468
t^h-L	7	11	0.424
k-H	9	19	0.316
k-L	9	19	0.316
m-R	16	43	0.276
l-R	17	62	0.201
c-H	2	3	0.187
p^h-H	3	6	0.182
p^h-L	3	6	0.182
c^h-H	4	12	0.135
V-R	16	90	0.126
V-L	13	90	0.096
s-R	5	29	0.074
t^h-H	3	11	0.073
l-L	7	62	0.059
n-R	3	13	0.058
ŋ-R	3	15	0.048
s-L	3	29	0.021
m-L	0	43	-0.012
k-R	1	19	-0.033
ŋ-H	1	15	-0.041
c^h-R	0	12	-0.042
n-L	1	13	-0.047
t^h-R	1	11	-0.054
p-L	1	9	-0.065
p-R	1	9	-0.065

p^h-R	0	6	-0.083
c-L	1	3	-0.146
c-R	0	3	-0.167

As can be seen in Table 11, *s-H* (= Kyengsang H(H)) shows the highest confidence score (0.611) among all coronal codas, which can help to explain the analogical changes to H(H) when the coda is realized as [s]. On the other hand, c^h-*L* (= Kyengsang H(L)) and t^h-*L* also show relatively higher confidence scores (0.468, 0.424), again supporting the tendency in the novel word experiment for the nonce words with coda -c^h to appear with H(L).

There is also some evidence that the segment-accent correlations have been generalized based on natural feature classes rather than individual segments. Table 12 shows the confidence scores for aggregated codas based on the laryngeal features relevant for Korean phonology: aspirate, lax, and sonorant (tense consonants do not appear in this position, except for *k**). Aspirate codas show a high confidence score for the L class (= Kyengsang H(L)), while lax and sonorant codas show high confidence scores for the H class (= Kyengsang H(H)). Also, the sonorant coda-R (= Kyengsang L(H)) correlation is relatively higher (0.211) compared to the aspirate coda-R (-0.022) or lax coda-R (0.064).

Table 12. Laryngeal features of coda and accent

	Hit	Scope	Corrected lower confidence limit
Aspirate-H	10	29	0.226
Aspirate-L	18	29	0.500
Aspirate-R	1	29	-0.022
Lax-H	37	57	0.568
Lax-L	13	57	0.155
Lax-R	7	57	0.064
Sonorant-H	136	223	0.570
Sonorant-L	32	223	0.114
Sonorant-R	55	223	0.211

Although these correlations may be inflated due to particular codas having higher confidence scores such as *s-H*, c^h-*L*, there is evidence from our South Kyengsang data that these segment-accent correlations have been generalized by speakers. For example, Table 13 shows the words with a coda -p^h in the Do et al.'s (2014) data plus data from two additional speakers. Words derived from the MK L accent class have the expected H(L) inflection. But when the aspirate coda is found in a stem whose MK H tone should yield H(H) inflection, quite a few (shaded) words have shifted to the H(L) inflection. Given that the confidence score for p^h-*L* (=Kyengsang H(L)) is 0.182 (based on the calculation in Table 11), it is unlikely that these changes are due to an originally strong correlation between -p^h and H(L). Rather, it appears that the relatively high aspirate-coda-L correlation contributed by -t^h and -c^h has been extended to these -p^h words. Based on our contemporary South Kyengsang data, the confidence score for the p^h-*H(L)* correlation is 0.447, which has dramatically increased from the estimated conservative state (0.182) based on the expected development from MK.

See Albright and Do (2013) for experimental results indicating that speakers encode the statistics of the lexicon in terms of natural phonological classes as opposed to individual segments.

Table 13. Accent of monosyllabic native nouns with a coda *-p^h*. SK1-7 indicate the South Kyengsang speakers.

Word	Gloss	MK form	MK accent	SK1	SK2	SK3	SK4	SK5	SK6	SK7
ap^h	front	alp^h	L	H(L)	H(L)	H(L)	H(L)	H(L)	H(L)	H(L)
jəp^h	side	njəp	L	H(L)	H(L)	H(L)	H(L)	H(L)	H(L)	H(L)
səp^h	brushwood	səp	L	H(L)			H(H)		H(L)	
ip^h	leaf	nip^h	H	H(H)	H(H)/H(L)	H(H)	H(L)	H(H)	H(H)	H(L)
cip^h	straw	tip^h	H	H(L)	H(H)/H(L)	H(H)	H(H)	H(H)	H(H)	H(H)
sup^h	forest	sup^h	H	H(L)	H(L)	H(L)	H(H)	H(H)	H(L)	H(L)

In sum, a variety of evidence indicates that aspirate coda consonants in monosyllables are correlated with the H(L) inflectional accent, while coda *-s* is correlated with H(H).

We now return to the data from section 3 concerning the accent class that is assigned when the coronal coda consonant is changed from the conservative to the innovative form. Our main hypothesis is that there will be a statistically significant bias to the H(H) inflectional accent in the segmentally innovative forms. A subsidiary question is whether the accent assigned in the segmentally innovative stems reflects the general default H(H) for monosyllables as a whole or is the product of correlations between particular coda consonants and tone. If the lexical biases documented in this section are active, we predict a significant difference in accent based on whether the innovative coda is *-s* vs. *-c^h* or *-t^h*. The tables below summarize the results, broken down by coda type. As seen in Table 14, when speakers responded with a conservative coda, most words appear with the regular H(L) class regardless of the coda type (80-100%). However, when speakers responded with an innovative coda consonant, most words switched the accent class to H(H), as seen in Table 15. Among innovative codas, *-s* shows the highest ratio of H(H) (76%) while aspirated codas appeared at lower rates (60% and 67%).

Table 14. Accent distribution when pronounced with a conservative coda

Coda \ Accent	H(H)	H(L)	Totals	H(H)%	H(L)%
c^h	12	88	100	12%	88%
t^h	17	67	84	20%	80%
c	0	12	12	0%	100%
Totals	29	167	196	15%	85%

Table 15. Accent distribution when pronounced with an innovative coda

Coda \ Accent	H(H)	H(L)	Totals	H(H)%	H(L)%
c^h	18	12	30	60%	40%
t^h	2	1	3	67%	33%
s	119	38	157	76%	24%
Totals	139	51	190	73%	27%

In order to assess the significance of these differences, we ran several statistical tests over our data. First, to test the difference between the conservative and innovative tonal patterns, a mixed effects logistic regression model was run over the data tabulated in Table 5 using the glmer function from the lme4 package (Bates et al. 2014) in R (R Development Core Team 2011). Accent class (H(L) vs. H(H), with the former as the baseline) was the dependent variable and coda consonant type (conservative (reference) vs. innovative) was the independent variable. Item and subject were set as random intercepts and random slopes. The results are shown in Table 16. Overall, the H(H) accent class is significantly dispreferred, as the negative estimate of the intercept indicates. As expected, the distribution of the accent classes was significantly affected by the coda consonant's status as conservative or innovative: the preference for the H(H) class is encouraged when coda consonant is in the innovative form.

Table 16. Result of a logistic regression model (conservative vs. innovative, coronal codas)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-2.389	0.447	-5.344	9.11e-08	***
Type-Innovative	4.178	0.699	5.977	2.27e-09	***

Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1.

Table 17 shows the accent distribution for segmentally innovative codas, taking into account whether the corresponding conservative coda form has already changed its accent class from H(L) to H(H). We expect that if the speaker has already changed the accent class in the segmentally conservative form then he should do so as well in the segmentally innovative inflection. This is largely true, with only three H(L) innovative coda responses for words whose corresponding conservative coda form is H(H). Also, the ratio of innovative coda H(H) responses is higher when the corresponding conservative coda form has already changed to H(H) compared to when the conservative coda form retains the H(L) class (90% vs. 70% as a whole).

Table 17. Accent distribution when pronounced with an innovative coda based on the tonal patterns when pronounced with a conservative coda

	Conservative: changed		Conservative: not changed		Totals	changed-H(H)%	not changed-H(H)%
	H(H)	H(L)	H(H)	H(L)			
c^h	4	1	14	11	30	80%	56%
t^h	1		1	1	3	100%	50%
s	23	2	96	36	157	92%	73%
Totals	28	3	111	48	190	90%	70%

For the 190 innovative coda datapoints in Table 17, a mixed effects logistic regression model was run with accent class (H(L) vs. H(H), the former as baseline) as the dependent variable, and coda type (c^h-t^h (reference) vs. *s,* c^h and t^h are aggregated for simplicity) as well as the conservative form's accent type (changed to H(H) or not (reference)), the log token frequency of each item (numerical), and speaker age (numerical) as independent variables. Item and subject were set as random intercepts. As seen in Table 18, coda type was not significantly different. On the other hand, the conservative accent pattern (changed to H(H) or not) was slightly significant: when the conservative form has already changed to H(H), the innovative form tends to appear with H(H) more. Also, the log token frequency of each item and speaker age were strongly significant.

Table 18. Result of a logistic regression model (innovative, coronal codas)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	7.37974	1.56526	4.715	2.42e-06	***
Coda-s	0.21377	0.61466	0.348	0.72801	
Conservative accent-changed	1.58439	0.79353	1.997	0.04586	*
Log token frequency	-1.60642	0.40564	-3.960	7.49e-05	***
Speaker age	-0.09763	0.02791	-3.497	0.00047	***

Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1.

Thus, whether the coda is -*s* vs. -c^h, -t^h does not appear to have had a significant effect on our 12 speakers' choices of accent as H(H) vs. H(L) in the segmentally innovative forms. Stated more cautiously, the character of the coda consonant does not have the effect we might expect if the inflectional accent was determined by lexical schemata such as "coda-*s* co-occurs with H(H)" and "coda-aspirate co-occurs with H(L)". Given this, we tentatively conclude that the preponderance of H(H) accent found in the innovative codas reflects the overall bias of H(H) over H(L) in the South Kyengsang monosyllables, and is not based on the local association (-*s* ≈ H(H)) that is applied after the segmental analogical changes take place. This conclusion is consistent with the stronger thesis of Albright (2008) that the base of Korean noun inflection is the isolation form despite the fact that this form of the paradigm is a neutralization site for coronal obstruents in all Korean dialects and a site of merger of the H(L) and H(H) accent types in the South Kyengsang dialect.

Next, the difference in tonal distributions between the conservative cluster codas (CVCC) and the innovative simplex codas (CVC) was tested in the same way based on the data in Table 8. As shown

in Table 19, the difference between conservative vs. innovative coda type is highly significant: the innovative CVC codas tend to appear with H(H) more frequently. Thus, here as well, a segmentally innovative inflection for the stem tends to be accompanied by the H(H) accent.

Table 19. Result of a logistic regression model (conservative vs. innovative, cluster codas)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-2.4382	0.4349	-5.607	2.06e-08	***
Type-Innovative	2.3544	0.4899	4.805	1.54e-06	***

Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1.

For the 73 innovative CVC cluster codas in Table 8, a mixed effects logistic regression model was run with accent class (H(L) vs. H(H), the former as baseline) as the dependent variable and the conservative form's accent type (changed to H(H) or not (reference)), the log token frequency of each item (numerical), and speaker age (numerical) as independent variables. Item and subject were set as random intercepts. The result (Table 20) shows that speaker age is slightly significant, but unlike in the case of coronal codas, whether or not the accent has changed in the segmentally conservative form and log token frequency are not significant.[13]

Table 20 Result of a logistic regression model (innovative, cluster codas)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-2.16331	1.01570	-2.130	0.0332	*
Conservative accent-changed	0.70037	0.92470	0.757	0.4488	
Log token frequency	0.01527	0.39688	0.038	0.9693	
Speaker age	0.04507	0.01824	2.471	0.0135	*

Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1.

Moreover, in the case of cluster codas, the effect of speaker age shows the opposite tendency compared to coronal codas: as seen in Figure 3 below, the older the speaker, the lower the ratio of H(L) responses in innovative forms.[14] At present, it is unclear why the opposite tendency is observed between coronal and cluster codas.

[13] Essentially the same result was obtained when we ran the same model based on all nine words collected in the experiment.

[14] Since the number of words is very small in the cluster coda test, Figure 3 shows the "ratio" of H(L) responses, not the "number" of H(L) responses.

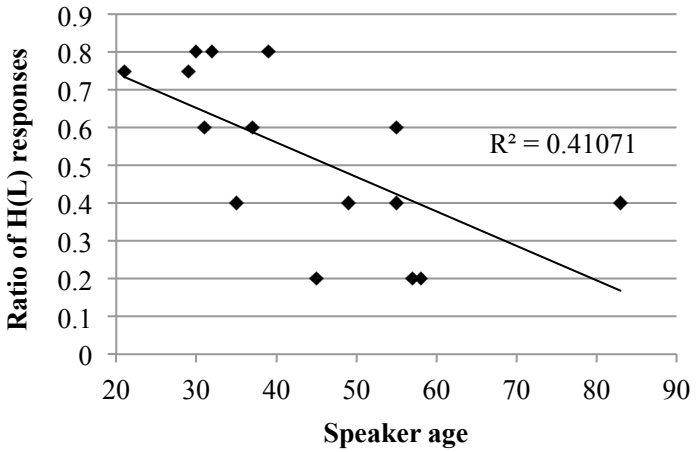


Figure 3 Retention of the conservative tonal pattern and speaker age (5 CVCC stems)

To summarize the results of this section, there is strong evidence that the inflectional tone tends to change from H(L) to H(H) when there is a change in the segmental shape of the stem. This correlation was observed for both the coronal coda and the cluster coda stems. A secondary question was whether the accentual change could be attributed to statistically based lexical schemata connecting the stem-final segments with a particular accent. Here our results are less conclusive. The key confidence scores are repeated in Table 21, along with the general default rules Ø > -íl and Ø > -ìl (*X-H* and *X-L*, respectively).

Table 21. Relevant confidence scores of coda-accent correlation

Pre-Kyengsang	Hit	Scope	Corrected lower confidence limit
s-H	21	29	0.611
X-H	188	321	0.552
c^h-L	8	12	0.468
t^h-L	7	11	0.424
k-H	9	19	0.316
k-L	9	19	0.316
X-L	70	321	0.190
c^h-H	4	12	0.135
t^h-H	3	11	0.073
s-L	3	29	0.021

We see that Ø > -íl / s__ is the most reliable rule (based on the 0.611 of *s-H*), rising a few points above the general default rule Ø > -íl for *X-H* (0.552) that assigns H(H) regardless of segmental shape (for monosyllables). The higher score of the latter compared to c^h*-L*, t^h*-L*, etc. could help to explain why the rules based on aspirates that favor H(L) are not significantly active in our data, as shown in Table 18. On the other hand, for the innovative CVC alternants from CVCC stems ending in *-k*, there is no difference in confidence between rules assigning H(H) vs. H(L) (both 0.316). This fact is consistent with the essentially equivalent 48% H(H) vs. 52% H(L) response ratio seen in Table 8, and might

suggest that the general rule Ø > -ɨl (*X-H*) may not be strongly affecting the distribution here, even though its confidence score is higher than *k-H and k-L*.

In any case, due to the relatively small number of datapoints for the accentual correlates of the two analogical changes discussed here (coronal coda 190, cluster coda 73), testing with a larger pool of subjects will be necessary before we can assert that inflectional rules based on the confidence scores in Table 21 are responsible for the overall pattern of data as well as any differences between these two segmental analogical changes.

7. Segmental analogy and tonal analogy

The discussion so far has shown that two analogical changes—segmental and tonal changes—are observed in South Kyengsang nouns and that they frequently co-occur. Table 22 (Tables 5 and 8 are aggregated) shows the number of responses in each combination: segment-no change (= conservative) ≈ tone-no change (= conservative), segment-no change (= conservative) ≈ tone-change (= innovative), segment-change (= innovative) ≈ tone-no change (= conservative), segment-change (= innovative) ≈ tone-change (= innovative). As can be seen, the case in which both the coda and tone are conservative as well as the case in which both the coda and tone are innovative are overrepresented, while the case in which the coda is conservative and the tone is innovative as well as the case in which the coda is innovative and the tone is conservative are underrepresented. This suggests a strong connection between the segmental change and the tonal change.[15]

Table 22. Segmental and tonal analogy

Segment \ Tone	No	Yes	Totals	No	Yes
No	234	35	269	1.43	0.33
Yes	89	174	263	0.56	1.68
Totals	323	209	532		

Still, the segment-tone correlation is not a perfect rule, and there are quite a few exceptions to this correlation. Also, the segmental change is observed by all speakers for every word, while the tonal change is not as strong: 263 segmental changes vs. 209 tonal changes. This tendency is observed in the difference between "segment (no)—tone (yes)" 35 vs. "segment (yes)—tone (no)" 89. Given these discrepancies, we infer that the segmental changes and tonal changes do not unfold completely in

[15] Another case showing the co-occurrence of segmental and tonal changes is observed in compound nouns. In South Kyengsang, underlying LH(L) + H(H) compounds appear as LHL while LH(L) + H(L) compounds appear as LHH as a rule. When the second member ends with a coronal coda and the coda changes from the segmentally conservative to innovative form, a change of compound accent is also observed for some speakers: *ka.ɨl* 'autumn' + *pjət^h* 'sunshine' → [ka.ɨ́l.pjə́.t^h-ɨl] ~ [ka.ɨ́l.pjə́.c^h-ɨl] ~ [ka.ɨ́l.pjə.s-ɨl]; *so.kɨm* 'salt' + *pat^h* 'field' → [so.kɨ́m.pá.t^h-ɨl] ~ [so.kɨ́m.pa.c^h-ɨl] ~ [so.kɨ́m.pa.s-ɨl]; *tal.lɛ* 'wild rocambole' + *k*oc^h* 'flower' → [tal.lɛ́.k*ó.c^h-ɨl] ~ [tal.lɛ́.k*o.s-ɨl]. These examples show that when the coronal coda changes, the compound accent changes as if the underlying accent of the second member is H(H).

tandem but rather constitute two separate analogical changes; also, the segmental changes have proceeded further than the tonal changes.

If this is true, then why have the segmental changes proceeded further than the tonal changes? A possible answer to this question is a difference in the degree of neutralization between the segmental patterns vs. the H(H) vs. H(L) tonal patterns. Segmental neutralization always occurs unless the word is followed by a vowel-initial suffix. Thus, neutralization occurs not only when the word is used in isolation (e.g. *k*oc*^{h} 'flower' → [k*ot]), but also when it forms the first member of a compound noun (e.g. *kət*^{h} 'surface' + *usim* 'laugh' →[kə.tu.sim] 'affected laugh') as well as the first member of a phonological phrase (e.g. *k*oc*^{h} 'flower' + *əps.ta* 'not exist' →[k*o.təp.ta]). On the other hand, the tonal distinction between H(H) and H(L) is neutralized when a word is used in isolation (both H(H) and H(L) appear as H),[16] but is retained when the word constitutes the first member of a compound noun (e.g. H(H) + H(L) → HH: *non* 'rice field' + *pat*^{h} 'field' → [nón.pát]; H(L) + H(L) → LH: *k*oc*^{h} 'flower' + *pat*^{h} 'field' → [k*ot.pát]) or the first member of a phonological phrase (e.g. H(H) + HL → HHL: *kus* 'exorcism' + *po.ko* 'see, gerund' → [kút.pó.ko]; H(L) + HL → LHL: *k*oc*^{h} 'flower' + *po.ko* 'see, gerund' →[k*ot.pó.ko]). In sum, tonal distinctions neutralize less frequently than segmental (coda) distinctions (Table 23), which may help to explain why the segmental analogy was more frequently observed in our speakers' responses compared to the tonal analogy.[17]

Table 23. Neutralization of segment and tone in various environments

	Simplex, isolation	First member of compound	Phrasal level
Segment	Yes	Yes	Yes
Tone	Yes	No	No

8. Summary and conclusion

In this paper, we investigated the tonal reflexes of a couple of well-known segmental analogical changes in Korean noun inflection in the South Kyengsang dialect. For monosyllabic stems ending in a coronal obstruent that neutralize to [-t̄] in the citation form, there is a tendency to substitute a non-etymological consonant when the noun is inflected, chiefly *-s* or *-c*^{h}. We found that when a H(L) noun adopts the innovative inflection in *-s* then there is a statistically significant tendency to change its tonal pattern to H(H). We also found the same tendency in stems with an underlying cluster when they are inflected with the CVC stem shape based on the citation form. We interpreted this finding as support for the model of paradigm learning proposed by Albright (2002, 2008) and explored further in Do (2013), in which the citation form of the noun serves as the base for all phonological alternations in the Korean nominal. In the South Kyengsang dialect, a monosyllable in high tone unpredictably

[16] Still, there may be slight acoustic differences between H(H) and H(L) in their isolation forms, which are not perceptually salient. See Appendix D.

[17] As mentioned in section 1, in fact, tonal neutralization between H(H) and H(L) in isolation forms may be a relatively recent event, given that some previous research (C-K. Kim 1975) reports a tonal difference. The neutralization of coronal codas, on the other hand, has been documented since MK.

alternates with two possible inflectional tones as H(H) or H(L). The H(H) pattern is about twice as frequent as H(L). If the citation form is the base for the rest of the paradigm, then the speaker must guess as to the inflected tonal pattern, with H(H) being a safer choice due to its greater type frequency. We also noted a correlation between the particular coronal coda and the tonal pattern. Most nouns ending in orthographic *-s* belong to H(H), while those ending in *-c*^{h} and *-t*^{h} tend to belong to H(L). We discussed the possibility that the change to H(H) inflection when the coronal coda inflects with *-s* might be due to lexical schemata that associate the H(H) and H(L) accent classes with particular segment types. A mixed-effects logistic regression model found that this factor does not reach statistical significance with our data, and so we tentatively concluded that the overall type frequency of the H(H) pattern in monosyllables was responsible for the change of tone under a change of coda. On the other hand, it is less clear whether the H(H) pattern emerging in the CVC variants of the cluster stems reflects this general default H(H) for monosyllables or is sensitive to the consonant *-k* that appears in the innovative form of these stems. We also observed that the segmental change has proceeded further than the tonal change, which is probably due to the fact that the former has more neutralization sites than the latter. Thus, while both segmental and tonal analogical changes frequently co-occur, they are two separate developments based on the isolation form, both motivated by the same mechanism.

Future research into this question should examine more South Kyengsang speakers for the connections documented here as well as explore other accent-segmental correlations in the tonal dialects of Korean. Tracking the development of the coronal coda variants and their tonal correlates in first language acquisition is also critical to gain a better understanding of these phenomena.

Appendix A: Questionnaire for this study (coronal codas)

이름 :
나이: 만

*출생 후 만 20 세 까지 부산/경남 지방 이외에서 1 년 이상 거주 경험이 있으신 분은 이 실험에 참여 하실 수 없습니다.

아래의 문장을 자연스러운 부산 사투리로 읽는다고 생각해 주세요. 밑줄 친 부분을 어떤 발음으로 읽으시겠습니까? 예를 들어, '옆을 지나다.' 의 경우, 부산 사람들은 [여플], [여블] 등 다양한 방식으로 이 단어를 읽곤 합니다. 본인이 가장 선호하는 발음을 한글로 표기해 주세요. 이때, 어떤 엑센트로 이 단어를 읽으시겠습니까? 예를 들어, [여플]의 경우 고저(여플), 저고 (여플), 고고(여플) 등으로 발음 될 수 있는데요, 본인이 위에서 선택한 한글발음을 할 때의 엑센트를 한글 표기 옆에 표시해 주세요.

예. 옆을 [여플], (고고)

볕을 쬐다.
닻을 올리다.
곁을 지키다.

물을 걷다.
낯을 가리다.
솥을 들다.
윷을 던지다.
옻을 만지다.
밭을 걷다.
돛을 달다.
겉을 보다.
꽃을 받다.
밑을 메우다.
숯을 갈다.
곁을 지나다.
빚을 지다.

이제, 위에서 선택한 발음 이외에 다른 발음으로도 밑줄 친 단어를 읽을 수 있는지 생각해 보세요. 예를 들어, 위의 예에서, [여플]을 선택한 경우, [여플] 이외에 다른 발음도 자연스러운 부산 사투리로 말씀하실때 가능 합니까? 가능하시다면 그 발음과 그때의 엑센트를 표시해 주세요.

예. 옆을 [여블] (고저)

두개 이상의 다른 발음이 가능하다고 생각하는 경우 모두 표기 하셔도 좋습니다.

예. 옆을 [여블] (고저), [여쁠] (고저)

Appendix B: Sample response by a subject (innovative case)

볕을 쬐다.	벼슬	고저
닻을 올리다.	다슬	고고
곁을 지나다.	겨슬	고저
물을 걷다.	무슬	고고
낯을 가리다.	나슬	고고
솥을 들다.	소슬	고고
윷을 던지다.	유슬	고고
옻을 만지다.	오슬	고고
밭을 걷다.	바슬	고저
돛을 달다.	도슬	고고
겉을 보다.	거슬	고저
꽃을 받다.	꼬슬	고저
밑을 메우다.	미슬	고저
숯을 갈다.	수슬	고고
곁을 지키다.	겨슬	고저
빚을 지다.	비슬	고고

Appendix C: List of sentences used for the test of CVCC stems

닭이 까맣다.
몫이 많다.
값이 비싸다.
흙이 검다.
넋이 나갔다.
삯이 적다.
곬이 크다.
칡이 보인다.
삶이 달린다.

Appendix D: Acoustic analysis of monosyllabic nouns

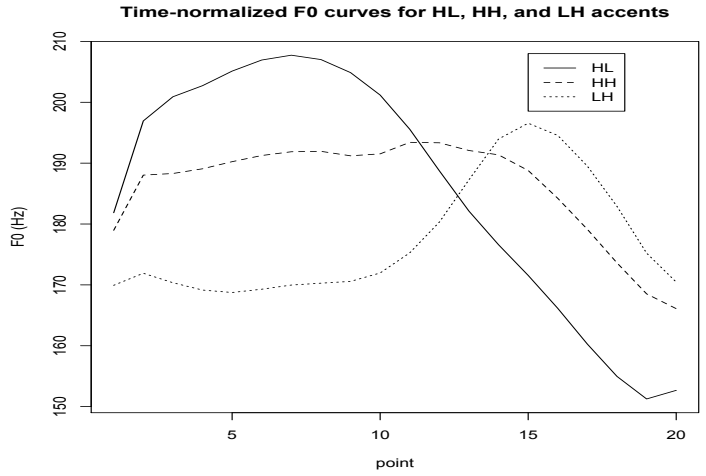
In order to corroborate the presumed merger of the H(L) and H(H) tonal classes in the isolation (citation) form, we collected and analyzed a corpus of 27 monosyllabic nouns recorded by six female South Kyengsang speakers ranging from 26 to 60 years of age. The words were balanced for tonal class (H(L), H(H), L(H)) and segmental shape ((C)V, (C)VC; obstruent vs. sonorant onset and coda), except for one combination: there is no word in the South Kyengsang lexicon with a zero-onset and an obstruent coda that belongs to the L(H) class, indicated with row Ø in the table below. Instead, we used the word *sok* 'inside' for this case. The 27 words were randomized and were recorded in isolation in two separate lists, first as a citation form (no suffix) and then inflected with a suffix, primarily the nominative *(-i ≈ -ka*) or in a few cases the copular *-ta*. The lists were read twice by each speaker. Note that the "Accent" column below is the expected (conservative) accent and that speakers did not necessarily pronounce each word with these expected accent patterns. Statistical analysis was done based on the accent class that each speaker employed for the inflected form.

Word	Gloss	Accent	Coda	Onset	inflected form	
an	inside	H(H)	Son	Ø	an-i	inside + NOM
pal	foot	H(H)	Son	Lax	pal-i	foot + NOM
mul	water	H(H)	Son	Son	mul-i	water + NOM
ip	mouth	H(H)	Obs	Ø	ip-i	mouth + NOM
pap	rice	H(H)	Obs	Lax	pap-i	rice + NOM
mək	ink	H(H)	Obs	Son	mək-i	ink + NOM
i	louse	H(H)	Ø	Ø	i-ka	louse + NOM
pɛ	boat	H(H)	Ø	Lax	pɛ-ka	boat + NOM
mo	corner	H(H)	Ø	Son	mo-ka	corner + NOM
ɨn	silver	H(L)	Son	Ø	ɨn-i	silver + NOM
tɨŋ	the back	H(L)	Son	Lax	tɨŋ-i	the back + NOM
mal	horse	H(L)	Son	Son	mal-i	horse + NOM
ap^h	front	H(L)	Obs	Ø	ap^h-i	front + NOM
kuk	soup	H(L)	Obs	Lax	kuk-i	soup + NOM
mok	neck	H(L)	Obs	Son	mok-i	neck + NOM

u	good	H(L)	Ø	Ø	u-ta	good + copula
po	wrapping cloth	H(L)	Ø	Lax	po-ka	wrapping cloth + NOM
mu	nothing	H(L)	Ø	Son	mu-ka	nothing + copula
il	work	L(H)	Son	Ø	il-i	work + NOM
pɛm	snake	L(H)	Son	Lax	pɛm-i	snake + NOM
mal	language	L(H)	Son	Son	mal-i	language + NOM
Ø		L(H)	Obs	Ø	Ø	
cas	pine nuts	L(H)	Obs	Lax	cas-i	pine nuts + NOM
nes	four	L(H)	Obs	Son	nes-i	four + NOM
o	five	L(H)	Ø	Ø	o-ta	five + copula
kɛ	dog	L(H)	Ø	Lax	kɛ-ka	dog + NOM
nɛ	stream	L(H)	Ø	Son	nɛ-ka	stream + NOM
sok	inside	L(H)	Obs	s	sok-i	inside + NOM

The results for the inflected forms showed a clear contrast among the three tonal classes in a manner that accords with previous research (Chang 2007, Kenstowicz et al. 2007). Time normalized pitch curves over the data pooled from all six speakers for the H(L), H(H), and L(H) accent contours appear in (i) below. They were constructed over 20-interval spans using a Praat script (Xu 2009). The F0 contours differ in both shape and scaling, with three F0 heights on the first syllable spaced c. 20 Hz apart. The H(L) curve reaches its peak in roughly the midpoint of the first syllable and then falls sharply. The H(H) accent has a gradual and shallow rise until roughly the vowel onset of the second syllable. The L(H) contour is flat across the first syllable and then rises sharply to roughly the midpoint of the second syllable. The scaling of the L(H) peak is comparable to the peak of the H(H) accent and lower than H(L).

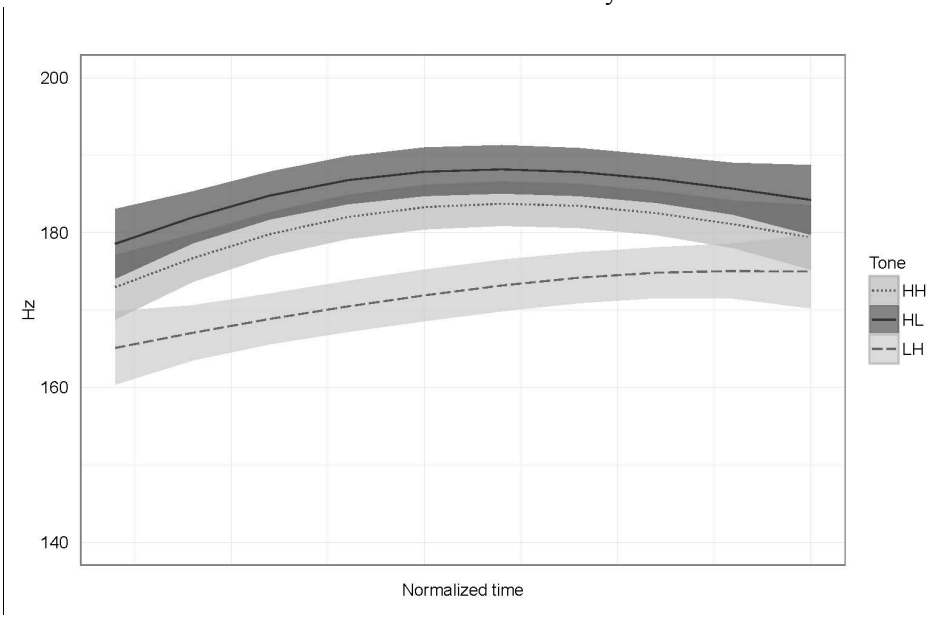
(i)



For the isolation forms, the time normalized pitch curves over a 15-interval span appear in (ii). For the first half of the syllable, the LH (rise) maintains a c. 20 Hz distance from H(H) comparable to that found in the suffixal context displayed in (i). On the other hand, the space between the H(L) and H(H) contours is compressed considerably, as shown by the overlapping 95% confidence intervals. Nevertheless, there was a small but consistent difference of 5-10 Hz in peak value for all six speakers. Also, the duration of the vowel in the isolation forms was measured and showed means of 195 ms

(H(L)), 224 ms (H(H)), and 292 ms (L(H)). A mixed effects linear regression model with H(H) as baseline and item and subject as random intercepts found a robust duration difference for L(H) ($p = 0.0003$) and an almost significant difference for H(L) ($p = 0.0549$). It thus appears that there is a near merger (in the sense of Labov 1994) of the H(L) and H(H) tonal contours in the isolation form rather than a complete neutralization.

(ii) Time normalized F0 contours of isolation forms of monosyllabic stems[18]



Our results for the unsuffixed monosyllables differ somewhat from those reported in Chang (2007), which we summarize here. In her study, six South Kyengsang speakers recorded the two words *nán* (H(L)) 'orchid' and *nám* (H(H)) 'other people' in the sentence frame *ice X andá* 'Now I know X'. There were ten repetitions per speaker. Chang found that the F0 peak was higher in H(H) *nám* for three speakers, lower for one speaker, and no different for the other two. In each case, the magnitude of the differences was small. Chang found no significant difference in peak location for the H(H) and H(L) contours; also the duration of the monosyllables with the H(L) accent was longer than for H(H).

At the very least, we may conclude that the H(L) vs. H(H) contrast is much less secure in the unsuffixed, bare stem compared to the suffixed form. And given the differences between our results and Chang's, it seems fair to conclude that there is no reliable and consistent difference between the two accents in this context. The accent changes discussed in the body of the paper can thus be safely considered to be triggered by this (near) merger in the isolation form.

References

[18] We thank Chingting Chuang for help in constructing this chart.

Albright, Adam. 2002a. *The Identification of Bases in Morphological Paradigms*. Los Angeles: UCLA Ph.D. dissertation.

Albright, Adam. 2002b. The lexical bases of morphological well-formedness. In Sabrina Bendjaballah, Wolfgang U. Dressler, Oskar E. Pfeiffer and Maria D. Voeikova (Eds.), *Morphology 2000: Selected papers from the 9th Morphology Meeting*, 5-15. Vienna, 24-28 February 2000.

Albright, Adam. (2002c). Islands of reliability for regular morphology: Evidence from Italian. *Language* 78(4). 684-709.

Albright, Adam. 2008. Explaining universal tendencies and language particulars in analogical change. In Jeff Good (Ed.), *Language Universals and Language Change*, 144-181. Oxford: Oxford University Press.

Albright, Adam and Bruce Hayes. 2002. Modeling English Past Tense Intuitions with Minimal Generalization. In Maxwell, Michael (Ed.), *Proceedings of the 2002 Workshop on Morphological Learning*, 58-69. Association of Computational Linguistics, Philadelphia: Association for Computational Linguistics.

Albright, Adam and Youngah Do. 2013. Featural overlap facilitates learning of phonological alternations. The 87th Annual Meeting of the Linguistic Society of America, Boston, MA.

Bates, Douglas, Martin Maechler, Ben Bolker and Steven Walker. 2014. Lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-6. Online: http://cran.r-project.org/web/packages/lme4/index.html.

Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.

Chang, Seung-Eun. 2007. *The Phonetics and Phonology of South Kyungsang Korean Tones*. Austin Texas: University of Texas Ph.D. dissertation.

Do, Youngah. 2013. *Biased Learning of Phonological Alternations*. Cambridge, MA: MIT Ph.D. dissertation.

Do, Young Ah, Chiyuki Ito, and Michael Kenstowicz. 2014. Accent classes in South Kyengsang Korean: Lexical drift, novel words, and loanwords. *Lingua* 148. 147-182.

He, Wung. 1955. Pangcem yenkwu [Study on tone marks]. *Tongpang Hakci* 2. 37-194.

Ernestus, Miriam and Harald Baayen. 2003. Predicting the unpredictable: interpreting neutralized segments in Dutch. *Language* 79. 5-38.

Ito, Chiyuki. 2010. Analogy and lexical restructuring in the development of nominal stem inflection from Middle to Contemporary Korean. *Journal of East Asian Linguistics* 19. 357-383.

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 (Ed.), *Japanese/Korean Linguistics* 12. 128-139. Stanford: CSLI.

Kenstowicz, Michael. 1996. Base identity and uniform exponence: alternatives to cyclicity. In J. Durand and B. Laks (Eds.), *Current Trends in Phonology: Models and Methods*, 363-394. Salford: University of Salford.

Kenstowicz, Michael and Hyang-Sook Sohn. 2001. Accentual adaptation in North Kyungsang Korean. In Michael Kenstowicz (Ed.), *Ken Hale: A Life in Language*, 239-270. Cambridge, MA: MIT Press.

Kenstowicz, Michael and Chiyoun Park. 2006. Laryngeal features and tone in Kyungsang Korean: a phonetic study. *Studies in Phonetics, Phonology and Morphology* 12. 247-264.

Kenstowicz, Michael, Hyesun Cho and Jieun Kim. 2007. Contrasts, mergers, and acquisitions in Kyungsang accent. *Toronto Working Papers in Linguistics* 28. 107-122.

Kim, Chin-Wu. 1972. Two phonological notes: A-sharp and B-flat. In Michael Brame (Ed.), *Contributions to Generative Phonology*, 155-170. Austin: University of Texas Press.

Kim, Cha-kyun. 1975. Kyengsangto pangen.ui sengcohyeng [Tonal types in South Kyengsang Korean]. *Ehak Yenkwu* 11-2. 119-137.

Kim, Hung-Gyu and Beom-Mo Kang. 2000. Frequency analysis of Korean morpheme and word usage. Technical report, Institute of Korean culture. Seoul: Korea University. http://www.sejong.or.kr.

Kim, Michael. 1997. Korean monosyllables: Floating V-slot and edge bracket. *Harvard Studies in Korean Linguistics* 7. 138-152.

Kim, No-Ju. 1997. *Tones, Segments, and their Interaction in North Kyungsang Korean*. Columbus, Ohio: Ohio State University Ph.D. dissertation.

Labov, William. 1994. *Principles of Linguistic Change*. Oxford: Blackwell Publishers.

Lee, Ki-Moon. 1972. *Kwukesa kaysel [History of Korean Language]*. Seoul: Tower Press. (First published in 1961.)

Lee, Hyunjung and Allard Jongman. 2012. Effects of tone on the three-way laryngeal distinction in Korean: an acoustic and aerodynamic comparison of the Seoul and South Kyungsang dialects. *Journal of the International Phonetic Association* 42-2. 145-169.

Lee, Hyunjung and Jie Zhang. 2014. The nominal pitch accent system of South Kyungsang Korean. *Journal of East Asian Linguistics* 23, 71-111.

Mikheev, Andrei. 1997. Automatic rule induction for unknown-word guessing. *Computational Linguistics* 23. 405-423.

Phillips, Betty. 1984. Word frequency and the actuation of sound change. *Language* 60. 320-342.

Phillips, Betty. 2006. *Word Frequency and Lexical Diffusion*. New York: Palgrave Macmillan.

Ramsey, S. Robert. 1978. *Accent and Morphology in Korean Dialects*. Seoul: Tower Press.

Sohn, Hyang-Sook. 2012. Random patterning in paradigms of noun-final coronal obstruents: palatalization, reanalysis and analogical leveling. *Studies in Phonetics, Phonology and Morphology* 18. 51-84.

Xu, Yi. 2013. Prosody Pro. www.phon.ucl.ac.uk/home/yi/ProsodyPro/

Yang, Charles. 2005. On productivity. *Linguistic Variation Yearbook* 5. 265-302.