Accent Classes in South Kyengsang Korean: Lexical Drift, Novel Words, and Loanwords

Youngah Do (Georgetown University), Chiyuki Ito (Tokyo University of Foreign Studies) and
Michael Kenstowicz (Massachusetts Institute of Technology)
June 2014

1. Introduction

Over the past decade, the investigation of various kinds of lexical frequency effects has emerged as a prominent line of research in generative phonology. Some of the more influential studies include the following. Albright and Hayes (2003) found that several segmental factors correlate with whether an English verb inflects its past tense through suffixation or internal stem change and that subjects project these 'Islands of Reliability' to their responses in novel word experiments. Ernestus and Baayen (2003) report that speakers of Dutch inflect a novel verb terminating in an ambiguous voiceless consonant with a morpheme alternant that reflects the relative type frequency of the particular voiced vs. voiceless alternatives in the lexical inventory of Dutch verbs. Becker et al. (2011) demonstrate a similar effect with Turkish voicing. They also point to a statistically reliable connection between the voicing alternation and the height of the preceding vowel. However, their subjects did not demonstrate any cognizance of this factor in their wug test responses. The authors take this result to indicate that grammatical knowledge has a strong UG basis. Generalizations that cannot be expressed by natural constraints (phonetically motivated and or with strong cross-linguistic credentials) will in principle be unlearnable. Hayes et al. (2009) tested a similar effect of vowel height in Hungarian vowel harmony and found that the less natural pattern did nevertheless show up in their wug test, but to a smaller and less reliable degree. From this they conclude that the acquisition of UG-motivated sound patterns has an analytic advantage compared to less natural ones. Zuraw (2000, 2010) found a similar effect in Tagalog. In the native lexicon, the nasal substitution alternation at the prefix-stem boundary applies with greater frequency and reliability to stems beginning with a voiceless stop compared to the corresponding voiced one: e.g. /paŋ-RED-pighatíʔ/ 'grief' > [pa-mi-mighatíʔ] 'being in grief' but /maŋ-RED-bigkás/ 'pronounce' > [mam-bi-bigkás] 'reciter'. This voicing difference also appeared in her subjects' wug test data as well as in loanword adaptation. Zuraw proposed a formal analysis of the phenomenon according to which the full-word forms of the language are listed as such in the lexicon but that in the course of learning the vocabulary of Tagalog, the weights of the relevant constraints such as *N-[-voice] are adjusted to reflect the voicing disparity based primarily on the frequency of the alternations for the voiced vs. voiceless stem-initial consonants. Faithfulness to the listed word (USE LISTED) ensures that for any given lexical item the appropriate alternant (coalesced N vs. N-C) is generated. But when such faithfulness is not at play, as in the adaptation of a loanword or the production of a novel word, then the speaker relies on the lower ranked/weighted markedness constraints to choose between the coalesced N vs. uncoalesced N-C alternants. Under this approach, the alternating forms are listed rather than generated; but the speaker's knowledge of the frequencies of the alternations is encoded in the weights of the grammatical constraints, rather than in counts or vectors of the lexical items themselves. Zhang et al. (2011) have utilized this approach to describe the productivity of the various alternations comprising the Taiwanese tone circle.

Another related line of research is concerned with the role of frequency in the diffusion of sound change and restructuring through the lexicon. It is well known that lexical variants with higher type frequency tend to attract words from lower type frequency classes. For example, many English verbs that originally inflected the past tense through stem vowel change have shifted to the suffixing class (weak verbs), while changes in the opposite direction are rare. But token frequency has also been claimed to be relevant to phonological change (Bybee 2001, Phillips 1984, 2006). For an innovating sound change, words with higher token frequency are more likely to change before lower frequency words do so, since the speaker has more opportunity to hear the novel, innovating pronunciation in the more frequently used words. On the other hand, analogical changes, which adjust the distribution of lexical items with respect to two or more entrenched word classes such as the English strong and weak verbs, are more likely to apply to less frequent words, since the speaker has less evidence for the class membership of such words. For example, Phillips (1984) found that the stress retraction that distinguishes verbs from the corresponding noun in such pairs as *pròtést* vs. *prótèst* tended to generalize to less frequent words first. However, Sondereggar and Niyogi (2013) put this result in some perspective: in a list of 110 such verb-noun pairs, they find that only about 12% have changed over the past 300 years; the overwhelming majority are stable. For the words that have changed, they find that there is indeed a correlation with frequency, as Phillips claimed, but caution that this result must be viewed with some reserve, given the small number of items involved.

In this paper, we address the questions of lexical diffusion and the native speaker's tacit knowledge of the statistics of the lexicon through a detailed study of the nominal accent classes in the South Kyengsang dialect of Korean. This dialect (in contrast to the standard Seoul dialect) has preserved the lexical accent distinctions of Middle Korean (15th-16th century, MK). For example, it is possible to find minimal triples as *mál-i* 'horse' nominative, *mál-í* 'unit of measure' nom., *mal-í* 'language' nom. and *káci* 'kind', *kácí* 'branch', *kací* 'eggplant'. Nevertheless, there are statistical biases in the distribution of these contrasting accent patterns as a function of syllable shape and weight. The existence of such biases was apparent to students of MK accent and formed the basis for internal reconstructions of the language back to an earlier stage with predictable final accent (Ramsey 1991, Whitman 1994). Another relevant (and puzzling) fact is that when Kyengsang speakers adapt an English loanword, the locus of English stress is ignored and a default accent is assigned, which is equivalent to a bimoraic trochaic foot at the right edge of the word: *remón* 'lemon', *tomíno* 'domino'. Kenstowicz and Sohn (2001) attributed this phenomenon to the speaker taking recourse to UG to assign an unmarked accent pattern: a bimoraic trochee at the right edge of the word. H-J. Kim (2012), inspired by the research mentioned above, in particular Zuraw (2000, 2010), suggests that UG plays little role here. Rather, the Kyengsang speaker assigns accent based on the action of UG markedness constraints, whose weights are determined by the frequency of the accent patterns in the native Kyengsang lexicon. See Kubozono (2006) for a similar interpretation of the Latin Stress Rule like accent assigned to English loanwords in Japanese and Davis et al. (2012) for a more general discussion of cases of accentual adaptation of loanwords.

Our paper is structured as follows. Background sections (2 and 3) describe the data collection and regular accent correspondences between MK and the Kyengsang dialects. We then present a detailed portrait of the accentual contrasts found in native monomorphemic noun stems, focusing on the South Kyengsang dialect. Since the relevant statistical biases differ depending on stem size, words are described in three successive blocks: monosyllables, disyllables, and trisyllables (sections 4, 5 and 6, respectively). For each block, the type frequencies of the contrasting accent classes are established based on our corpus. We then examine the regularity of the accentual correspondences from the known MK source to its reflex in our corpus. It is shown that larger accent classes systematically tend to attract words from less populated classes (the type-frequency effect). We also show that less frequently used words are more likely to change their accent class affiliation compared to more frequently used ones (the token-frequency effect). For each stem length, we then establish certain syllable shape and weight biases. Most of these have a natural phonetic basis. But at least one is more phonologically arbitrary. However, it is plausibly traced back to a natural tendency at an earlier stage of the language. In section 7, we present the results of a novel word experiment designed to test whether speakers are aware of the syllable shape and weight factors that bias a word to a particular accent class in the native lexicon. We find that both types of factors show up in the experiment. But the experiment uncovered two additional disparities between the lexicon and the experimental results. A statistically reliable but unnatural weight constraint fails to emerge, while an aspiration effect in initial position is extended to medial position in the subjects' responses. We then briefly compare our statistical and experimental results for the South Kyengsang dialect to those of H-J. Kim (2012) for North Kyengsang and review the accent patterns that emerge in loanwords (sections 8 and 9). Section 10 returns to the question of the extent to which we can legitimately claim that the biases showing up in the novel word experiment are rooted in the statistics of the lexicon. Using the Maxent Grammar Tool (Hayes 2009), weighted-constraint grammars are proposed for the disyllabic and trisyllabic stems. It is shown that these two stem classes have quite different constraint structures, particularly with regard to syllable weight. We find that the best results occur when the wug test results are simulated by a grammar that has been trained on the lexical trisyllables. This suggests that for words with longer stem length, the accent is more predictable; or, stated differently, accent is more contrastive in the shorter disyllables. One possible interpretation of this state of affairs is that the majority of trisyllabic words have accent assigned by rule with a relatively small number of exceptions, while in disyllables the accents are listed. When tasked with assigning accent to a novel word, it is the rules/constraints enforcing the quantity-sensitive trochaic accent in the lexical trisyllables that are activated rather than more detailed matching of the statistics of the lexicon. The final section summarizes the study and suggests several avenues of future study.

2. Data Collection

Five South Kyengsang dialect speakers who were born and raised in the Pusan and Changwen area furnished the data for this study. Our corpus consists of c. 1,900 native (non-Sino-Korean) nouns taken from two standard Korean dictionaries (*Kwuklip kwuke yenkwuwen* 1999, and Kadowaki et al. 1993) as well as words gathered through interviews with our consultants. They were assembled into a 500-page booklet with spaces for marking the tonal pattern of each noun in its citation and inflected (nominative/accusative) forms. For example, for monosyllables the three attested patterns are High(-High) (*múl*, *múl-í*, *múl-ɨ́l* 'water'), High(-Low) (*súl*, *súl-i*, *súl-ɨl* 'wine'), and Rise/Low-High (*tǒn*, *ton-í*, *ton-ɨ́l* 'money').[1] For three of our subjects (CP, MA, KM), the items in the booklet were elicited

[1] The transcription system for the example data in this paper is as follows. MK falling-sonority diphthongs have been monophthongized in Contemporary Korean (CK), as shown below. Some symbols such as ㄹ, ㅇ are transcribed differently depending on environment. For the MK accent, low pitch (= no dot), high pitch (= one dot) and rising pitch (= two dots) are transcribed with grave accent (`), acute accent (´), and both accents (˘) over the vowel, respectively. Other Korean words in the body of the paper follow the Yale Romanization system, except for some authors' names, which follow their customary spelling.

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

orally one by one and marked by one of the authors. The other two subjects (YD, YJ) went through the booklet marking each item by themselves, after being shown examples of the various accent patterns. The data were then tabulated into excel files and classified and counted for analysis. Some of the words we solicited were rare or otherwise unknown to our speakers. This explains the fact that the corpus size varies among our consultants. Nevertheless, the basic generalizations hold over all five speakers. For each of the nouns we collected, we also recorded its cognate MK form (if known) taken from a database derived from the analysis of over thirty MK texts by the second author (cf. Ito 2013).

3. Historical Development

MK texts were written in the alphabet (Hangul) devised by King Sejong in the 15th century. During the 15th-16th centuries, many texts marked the tone of each syllable by a series of side-dots next to the strokes composing the syllable block. One dot denoted a H(igh) tone, two dots a R(ising) tone, and no dots a L(ow) tone. The R tone was restricted to the initial syllable of the stem as a rule, while H and L tones could be found on initial, medial, and final syllables contrastively. The tones of the inflectional suffixes were predictable from the tonal category of the stem and their odd-even position in the string of syllables after the first high-pitch (accented syllable) in the phonological phrase. There was one distinctive pitch peak per word. See He (1955), Ceng (1971, 1972), Hayata (1974, 1999), W-C. Kim (1973), Kadowaki (1976), Ramsey (1978), S-O. Lee (1978), Fukui (1985, 2013), and others for discussion of MK accent. Some examples appear in Table 1. X indicates that the tone is predictable and hence is not specified.

Table 1: MK accent

Monosyllable			Disyllable		
H	íp	'mouth'	HX	káci	'branch'
L	cìp	'house'	LH	mərí	'head'
R	nǔn	'snow'	LL	pòrì	'barley'
			RX	sǎrʌm	'person'

According to the reconstruction in Kenstowicz et al. (2007), the first change in the evolution of the modern tonal dialects was the addition of a high tone to the final syllable of a stem lacking a pitch peak. This marked the final step in the transition to a pitch accent system in which each word had one (and only one) pitch peak (culminativity), grouping Kyengsang with Ancient Greek and Common Slavic, compared to Tokyo Japanese and Basque, which have a large class of unaccented words on the surface.[2] The H-insertion threatened to neutralize the contrast between toneless/all low stems like *pòrì* 'barley' and final H stems like *mərí* 'head'. In the North Korean Hamkyeng dialects, this neutralization occurred. But in the South Korean Kyengsang dialects, the original (non-inserted) H was retracted one syllable by a general sound change, thus avoiding a merger between the inserted

	ㅏ	ㅐ	ㅑ	ㅓ	ㅔ	ㅕ	ㅖ	ㅗ	ㅘ	ㅙ	ㅚ	ㅛ	ㅜ	ㅝ	ㅞ	ㅟ	ㅠ	ㅡ	ㅢ	ㅣ
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

2 See however Lee and Davis (2010) for a different interpretation.

high tone and original high tones on final syllables. When there was no preceding syllable (i.e. in monosyllables), then neutralization occurred in Kyengsang as well.

Table 2: Accent of MK, Hamkyeng and Kyengsang

MK	Hamkyeng	Kyengsang	
pòrì	porí	porí	'barley'
mə̀rí	mərí	mə́ri	'head'
cìp	cíp	cíp	'house'
íp	íp	íp	'mouth'

However, the Kyengsang H-retraction seen in *mə̀rí* > *mə́ri* 'head' threatened the distinction with the original initial-H class such as MK *káci* 'branch'. This merger was also avoided by doubling the tone of the former initial H to create an HH class that was absent from the original MK inventory of stem types. This H doubling also applies to inflected monosyllabic stems from the original MK H class and creates a regular phonological alternation, as seen in 'mouth' below.

Table 3: Accent of MK, Hamkyeng and Kyengsang (with inflected forms)

MK	Hamkyeng	Kyengsang	
pòrì	porí, pori-ká	porí, porí-ka	'barley'
mə̀rí	mərí, mərí-ka	mə́ri, mə́ri-ka	'head'
káci	káci, káci-ka	kácí, kácí-ka	'branch'
cìp	cíp, cip-í	cíp, cíp-i	'house'
íp	íp, íp-i	íp, íp-í	'mouth'

On this account, the evolution of accent from MK to Kyengsang took the form of a push-chain, as depicted in (1). Other interpretations are possible. For our purposes here, it is the regularity of the accentual correspondences that is significant rather than the diachronic direction or sequencing of the sound changes.

(1) Chain-shift in development of Kyengsang accent

LL > LH	pòrì > porí
LH > HL	mə̀rí > mə́ri
HX > HH	káci > kácí

The Kyengsang dialects split into South and North variants in terms of their reflexes of the MK Rise tone. In the South, the Rise was retained on monosyllables and is associated with a noticeable phonetic vowel lengthening. But in polysyllables, it split into a LH sequence. This produces regular tonal alternations in the inflection of monosyllables. The H component of the (former) rise regularly undergoes tone doubling in inflection: *săr*ʌ*m* 'person' > *sarám* > *sarám-éke* (-*eke* = dative). In North Kyengsang the vowel associated with the MK Rise was lengthened and the tone was simplified to H, which then doubled like the original word-initial H: *săr*ʌ*m* 'person' > *să:r*ʌ*m* > *sá:ram* > *sá:rám*.

Table 4: The reflex of MK Rise tone in South and North Kyengsang

MK	South Kyengsang	North Kyengsang	
pə̌m	pə̌:m, pəm-í, pəm-éké	pə́:m, pə́:m-í, pə́:m-éke	‘tiger’
sǎrʌm	sarám, sarám-í, sarám-éke	sá:rám, sá:rám-i, sá:rám-eke	‘person’

In sum, the South Kyengsang lexicon displays three tonal classes for monosyllables and four for disyllables. It is noteworthy that the isolation form is a neutralization site for a final-syllable H, which splits into two classes when an inflection is added. We denote these classes as H(H) and H(L) for monosyllables and LH(L) and LH(H) for disyllables. See Lee and Davis (2009) and Lee and Zhang (2014) for recent synchronic analyses of the South Kyengsang nominal accent patterns.

Table 5: South Kyengsang tonal class (monosyllable, disyllable)

Monosyllable				Disyllable			
H(H)	múl	múr-í	‘water’	HH	kúrɨ́m	kúrɨ́m-i	‘cloud’
H(L)	súl	súr-i	‘wine’	HL	mánɨl	mánɨr-i	‘garlic’
R	tǒn	ton-í	‘money’	LH(L)	parám	parám-i	‘wind’
				LH(H)	sarám	sarám-í	‘person’

From this point on, we will use the following accentual transcription system for simplicity. Note that H(H) and H(L), which are distinguished as σ̋ and σ́, respectively, appear with identical surface tonal contours in isolation forms (Kenstowicz et al. 2007). The same is true for LH(L) and LH(H): σ σ́ vs. σ σ̋.

Table 6: Accentual transcription system of South Kyengsang

Monosyllable		Disyllable	
H(H)	műl ‘water’	HH	kúrɨ́m ‘cloud’
H(L)	súl ‘wine’	HL	mánɨl ‘garlic’
R	tǒn ‘money’	LH(L)	parám ‘wind’
		LH(H)	sara̋m ‘person’

Having sketched the general accentual sound changes that occurred in the development of Middle Korean into the contemporary Kyengsang dialects, we now turn to a detailed portrait of the native nominal accent classes and the factors that have biased particular lexical items to depart from the expected MK-SK correspondences. These are discussed successively in terms of three prosodic shapes: monosyllables, disyllables, and trisyllables.

4. Monosyllables

As we have just seen, Kyengsang monosyllables reflect three MK tonal classes: High (> H(H)), Low (> H(L)), and Rise. In MK, their numerical distribution was H > L > R. Table 7 shows the frequencies of the monosyllabic words collected from our South Kyengsang speakers across the three classes.

Table 7: Accent distribution of South Kyengsang monosyllabic words

speaker / accent class	YD	CP	MA	YJ	KM	Totals	Ratio
H(H)	233	214	196	212	278	1,133	53%
H(L)	148	135	120	82	70	555	26%
R	91	104	71	104	80	450	21%
Totals	472	453	387	398	428	2,138	

The speakers differ in the number of words they recognize/use. But the distribution reflects the MK one well, with half of the items falling into the double H(H) class and the remaining half split between the alternating H(L) and R classes, with a small bias in favor of the former. This distribution suggests that the three classes as a whole have largely maintained their relative proportions to one another, even if particular lexical items have shifted from one class to another.

Next we see the agreement rate among our speakers. This is based on the words for which all five speakers responded. (Recall that speakers differed in the words they knew/used). The Agree row indicates the number of words to which all five speakers assigned the same accent. Disagree indicates the number of words where this is not the case. The agreement rate varies somewhat depending on the word size. The overall agreement rate is 64%, indicating quite a bit of variance among our South Kyengsang speakers.

Table 8: Agreement rates

	Monosyllable	Disyllable	Trisyllable
Agree	213 (67%)	334 (59%)	169 (74%)
Disagree	106 (33%)	230 (41%)	59 (26%)

4.1 Middle Korean–South Kyengsang correspondences: monosyllables

Table 9 shows the distribution of the South Kyengsang monosyllables across the three MK tonal classes that are the source of the current South Kyengsang tonal classes. ‘SK attested’ indicates the number of words in our corpus that have an attested MK source whose tone is known. They form 69% of the entire corpus of monosyllables (= 322/468). ‘MK attested’ indicates the number of monosyllabic noun stem words that are attested in MK (Ito 2013). As can be seen, the H class was largest in MK at almost twice the size of the L, which in turn was slightly larger than the R class. The ‘SK all’ column (repeated from Table 7) shows the relative proportion of the three accent classes in the current South Kyengsang lexicon, regardless of whether or not they have a known MK cognate, as well as whether or not they show the regular correspondences with MK. We see that the MK H class has slightly lost population (56% → 53%) and the R class has slightly gained (18% → 21%), while the L class remains about the same (26% → 26%).

Table 9: Accent distribution of the South Kyengsang monosyllables across the three MK tonal classes

MK accent	MK attested	SK attested	SK accent	SK all
H	286 (56%)	188 (58%)	H(H)	1,133 (53%)
L	132 (26%)	71 (22%)	H(L)	555 (26%)
R	89 (18%)	63 (20%)	R	450 (21%)
Totals	507	322		2,138

4.2 Type Frequency

Table 10 summarizes the correspondences between lexical items in our corpus whose MK tonal value is known, summed across our five South Kyengsang speakers. Each data point represents a particular MK-South Kyengsang pair for one of our speakers. It allows us to assess the regularity of the accent correspondences and to see if particular classes attract or repel lexical items given a common starting point.

Table 10: Correspondences between MK and South Kyengsang (monosyllabic). Regular correspondences are shaded. The same representation is used in other correspondence tables below.

MK \ SK	H(H)	H(L)	R	Totals
H	720 (83%)	90 (10%)	55 (6%)	865
L	81 (24%)	239 (70%)	22 (6%)	342
R	69 (23%)	33 (11%)	192 (65%)	294
Totals	870	362	269	1,501

(2) Examples of regular correspondences

a. MK H: ı̋ ‘louse’, k*ɛ̋ ‘sesame’, t^{h}ə̋ ‘site’, c^{h}ı̋m ‘saliva’, nűn ‘eye’, pa̋l ‘foot’, s*a̋l ‘rice’, pűs ‘writing brush’, p^{h}űl ‘grass’

b. MK L: áp^h ‘front’, súc^h ‘charcoal’, mít^h ‘bottom’, pák* ‘outside’, mók ‘neck’, wí ‘upside’

c. MK R: kě ‘crab’, nɛ̌ ‘stream’, sǔm ‘breath’, mǎl ‘language’, pə̌l ‘bee’, tǒl ‘stone’, cǒŋ ‘servant’

The first thing to note is that the regular correspondences predicted by the sound changes (shaded) mentioned in section 3 predominate. Second, the relative degree of regularity of the expected correspondences (MK H (83%) > L (70%) > R (65%)) reflects the type frequency of the classes in MK (H (56%) > L (26%) > R (18%)) seen in Table 9. Third, among the irregular correspondences, there is a bias against shifting to the R class. That is, items from the MK H class (> Kyengsang H(H)) are more likely to shift to Kyengsang H(L) (10%) than to R (6%), and items from the MK L class are more likely to shift to Kyengsang H(H) (24%) than to R (6%). This difference makes sense, in that the H(H) and H(L) classes are identical in their isolation form and so are ambiguous as to the tonal contour (H(H) vs. H(L)) of the corresponding inflected form. Items from the MK R class are distinct from the other two in both the isolation (H vs. R) and inflected (L(H) vs. H(H) and H(L)) forms. Moreover, the isolation form of R has a marked LH rising tone that other things being equal make it an unattractive model. In addition, as we shall see, it is associated with a particular syllable shape. Thus, the Rise class is the smallest as well as the most marked phonologically, owing to its complex

LH tone. As seen in Table 10, when lexical items abandon this class, there is a 23% vs. 11% preference for the H(H) class that approximates the 53% vs. 26% frequency of these two classes in the overall corpus. A final disparity of note is that the percentage of MK L class words that shift to the Kyengsang H(H) class is almost 2.5 times larger than the percentage of MK H class words that shift to the Kyengsang H(L) class: 24% vs. 10%. In other words, the H(H) class is more likely to attract words from the H(L) class than vice versa.

In sum, we see that the H(H) class is favored by the South Kyengsang monosyllables. It has the largest population in the lexicons of both MK and South Kyengsang, as well as the most regular correspondences with MK. It attracts items from the other two classes. The R class is a deflector. It has the lowest percentage of regular correspondences, having lost more items than the other two classes.

4.3 Token frequency

In addition to type frequency, it is claimed that token frequency also plays a role in phonological change (Phillips 1984, 2006, Bybee 2001). More frequently used words have more opportunity to undergo an innovating sound change as well as to serve as models for the transmission of the change from one speaker to another. Less frequently used words are more susceptible to analogical changes when there are two (or more) possible lexical variants for a word: the learner has less opportunity to hear the word in general as well as in a crucial inflectional context that might determine its mode of alternation. We thus expect that items with regular correspondents are more likely to be drawn from the more frequently used words, while those that have been reclassified due to misidentification or inflectional ambiguity will be drawn from the less frequently used words. Our data allows us to test this hypothesis.[3]

To do so, we classified the monosyllabic words into high and low frequency groups as follows. First, for each monosyllable in our corpus with a known MK cognate, we identified its token frequency in the King Sejong corpus (Kim and Kang 2000). We then determined the median in the distribution of such words. Words lying above the median were classified as “high frequency words” and words lying below the median were classified as “low frequency words”.[4] Table 11 and Figure 1 show how the monosyllables from the five South Kyengsang speakers are partitioned by the low vs. high frequency groups. For example, among the 711 South Kyengsang H(H) items descended from the MK H class that were collected from our speakers, 290 fall into the low frequency group, while 421 fall into the high group. A similar data pattern is exhibited by the MK L and R classes. The percentages in parentheses show the proportions for each MK accent class with respect to the high vs. low frequency grouping of its correspondences among our five South Kyengsang speakers. Quite consistently, more regular correspondences are drawn from the high frequency groups than from the low frequency groups: 93% vs. 74% (MK H), 80% vs. 62% (MK L), and 78% vs. 51% (MK R). And in the complementary irregular correspondences, more correspondences are drawn from the low frequency groups than from the high frequency groups. A Pearson’s Chi-square test with the Yates continuity correction partitioning the 1,450 words into regular vs. irregular MK-South Kyengsang

[3] See Kang (2003), Albright (2008), J-H. Jun (2010), and Sohn (2012) for discussion of the role of token frequency in the generalization of the t≈s alternation in the inflection of Korean nominal stems ending in a coronal obstruent.

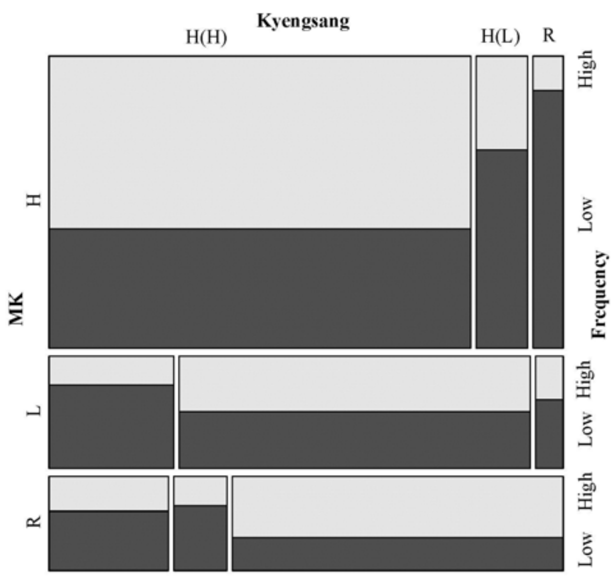
[4] We thank Adam Albright for making available a transcribed Excel version of the King Sejong corpus. This corpus is based on contemporary written Korean and so can only give an approximation to the frequency of words at earlier stages of the language.

development pitted against high vs. low frequency group found the connection to be statistically significant: χ^2 = 87.6729, df = 1, p < 2.2e-16.

Table 11: Frequency and correspondence with MK (monosyllabic)

SK / MK	H(H)	H(L)	R	Totals
H (Low)	290 (74%)	59 (15%)	45 (11%)	394
H (High)	421 (93%)	28 (6%)	6 (1%)	455
L (Low)	60 (32%)	115 (62%)	11 (6%)	186
L (High)	21 (15%)	113 (80%)	7 (5%)	141
R (Low)	41 (33%)	20 (16%)	63 (51%)	124
R (High)	24 (16%)	9 (6%)	117 (78%)	150

Figure 1: Mosaic plot (frequency and historical development)



(3) Examples

a. MK H (low freq.): pjə̋~pjə́~pjə̌ ‘rice plant’, hwɛ̋~hwɛ́~hwɛ̌ ‘torch’, ka̋s~kás ‘hat’

b. MK H (high freq.): ta̋l ‘moon’, hı̋m ‘strength’, ső ‘cow’

c. MK L (low freq.): jő~jó ‘mattress’, őc^{h}~óc^{h}~ǒc^{h} ‘lacquer’, tők~tók ‘jar’

d. MK L (high freq.): cíp ‘house’, k*óc^{h} ‘flower’, kúk ‘soup’

e. MK R (low freq.): ı̋m~ím ‘one’s beloved’, mɛ̋~mɛ́ ‘hawk’, mé̋~mé~mě ‘mountain’

f. MK R (high freq.): ǐl ‘work’, tǒn ‘money’, pjə̌l ‘star’

4.4 Coda effects

Table 12 shows the observed over expected distributions of the three monosyllabic accent classes with respect to the coda (obstruent or sonorant) summed over our five speakers. As seen in Table 12, the

Rise class prefers words with sonorant codas (1.31), while the H(L) class prefers words with obstruent codas (1.50). According to the internal reconstruction of Korean accent proposed by Whitman (1994), the monosyllabic H(L) and R classes both derive from disyllabic CVCV stems with a LH pitch pattern; the different tonal outcomes reflect the character of the intervocalic consonant, with the H(L) deriving primarily from voiceless consonants and the R from sonorant (voiced) consonants.

Table 12: Monosyllabic accent class with respect to the coda (obstruent or sonorant). Left = Observed numbers, right = Observed/Expected numbers. Statistically significant correlations (α-level at .05) are in bold. These are cells whose Chi-square exceeds the .05 alpha level for the table as a whole and thus contribute most to the overall Chi-square statistic [χ^2 = 171.1472, df = 2, p < 2.2e-16].

coda \ accent	H(H)	H(L)	R	Totals
Obstruent	342	235	25	602
Sonorant	791	320	425	1,536
Totals	1,133	555	450	2,138

H(H)	H(L)	R
1.07	**1.50**	**0.20**
0.97	**0.80**	**1.31**

(4) Examples

a. Obstruent coda with H(H): mɨ̋k ‘jelly’, ɨ̋p ‘mouth’, pɨ̋c^h ‘light’
b. Obstruent coda with H(L): tə́c^h ‘trap’, móks ‘share’, cíp ‘house’
c. Obstruent coda with R: [see Table 13]
d. Sonorant coda with H(H): c^{h}ə̋l ‘discretion’, cwɨ̋ ‘cramp’, mə̋ŋ ‘bruise’
e. Sonorant coda with H(L): mál ~ mǎl “horse”, k*wáŋ ‘blank’, tɨ̋ŋ ~ tɨ́ŋ ‘back’
f. Sonorant coda with R: kǐm ‘laver’, t^{h}ǎl ‘mask’, sǒm ‘cotton’

Most words in our corpus drawn from the MK Rise class with obstruent codas have been reassigned to the Kyengsang H(H) or H(L) classes. We cite these items in Table 13, along with the cognate Seoul form from Martin (1987), where the long vowel is consistent with the Rising tone source.[5] The preference for sonorant codas of course makes phonological sense, since they can carry F0 pitch distinctions, which are difficult if not impossible to detect in obstruents. See Zhang (2002) for in depth discussion of this general point.

Table 13: MK Rise class with an obstruent coda

Item		YD	CP	MA	KM	YJ	Seoul	MK
pəs	‘friend’	H(L)	H(H)/H(L)	H(L)	H(L)	H(H)	pə:s	R
cas	‘pine nuts’	H(H)	H(H)/R	R	H(H)/R	H(H)	ca:s	R
cis	‘gesture’	R	H(H)	H(H)	H(H)	H(H)	ci:s	R
nat^h	‘piece’	H(L)	H(L)		H(H)	H(L)	na:t^h	R
sok	‘interior’	H(L)	H(L)	R	H(H)/H(L)/R	H(L)	so:k	R

More interesting is the fact that a number of words from the MK H and L classes have been lured into the marked R class in South Kyengsang, running against the general trend towards the largest

[5] A long vowel in the initial syllable is a regular reflex of the MK rising tone (Martin 1992: 35). For younger speakers of the standard Seoul dialect, the vowel length distinction is disappearing, eliminating the last vestige of the MK accents.

accent class. All of these items end in a sonorant (Table 14). The connection between the Rise class and its signature sonorant coda thus forms an "Island of Reliability" (Albright 2002). Some of these items have a long vowel in Seoul, suggesting that this association between tone and syllable shape may have operated there as well.

Table 14: MK H and L class words with a sonorant coda

Item		YD	CP	MA	KM	YJ	Seoul	MK
um	'dugout'	R	R	R	H(H)		u:m	H
k*ɨl	'chisel'	R	H(H)	H(H)	H(H)	R	k*ɨl	H
kul	'oyster'	R	R	R	R	R	kul	H
hom	'groove'	R	H(H)/H(L)	R	R	H(H)/R	ho:m	H
com	'moth'	R	R	H(H)	H(H)/R	R	com	H
sol	'pine'	R	R	H(L)	R	R	sol	H
pɛ	'pear'	H(L)	H(H)	H(L)	H(H)	R	pɛ	L
p*oŋ	'mulberry leaf'	H(L)	R	H(L)	H(H)	H(L)	p*oŋ	L
k*wəŋ	'pheasant'	H(L)	H(L)/R	H(L)	H(H)	R	k*wəŋ	L
hjuŋ	'scar'	H(L)	R	H(L)	H(H)	R	hjuŋ	L
t^həl	'fur'	H(L)	H(L)	H(L)	R	H(L)/R	t^həl	L

4.5 Onset effects

Aspirated and tense stops tend to raise F0 in the following vowel, both in standard Korean (Cho et. al 2002) as well as in the Kyengsang dialect (Kenstowicz and Park 2006). We might wonder whether this property affects the distribution of the three accent classes in the South Kyengsang lexicon. Table 15 shows the observed over expected distributions of the three monosyllabic accent classes with respect to the onset type (sonorant, lax, aspirated/tense). /s/ and /h/ are excluded in the calculation.[6]

Table 15: Monosyllabic accent classes with respect to onset type [χ^2 = 93.2756, df = 4, p < 2.2e-16].

onset \ accent	H(H)	H(L)	R	Totals
Sonorant	290	147	120	557
Lax	335	180	195	710
Aspirated/Tense	382	131	42	555
Totals	1,007	458	357	1,822

H(H)	H(L)	R
0.94	1.05	1.10
0.85	1.01	**1.40**
1.25	0.94	**0.39**

(5) Examples

a. Sonorant with H(H): mə̋s 'smartness', ȉ 'louse', nőn 'rice field'

b. Sonorant with H(L): jə́p^h 'side', mít^h 'bottom', wí 'upside'

c. Sonorant with R: nɛ̌ 'stream', ǒm 'itch', nǔn 'snow'

[6] Fricatives behave like an aspirate/tense consonant in some cases (in disyllabic words, the HH class is overrepresented), whereas they behave like a lax/sonorant one in other cases (the R/LH classes are overrepresented and the H(H) class is underrepresented). Given this, it is difficult to group fricatives with only aspirate/tense or lax/sonorant, and hence we exclude them from the analysis for simplicity.

d. Lax with H(H): cə̋ 'ruler', pɛ̋ 'boat', kə̋ps 'value'
e. Lax with H(L): cíp 'house', kúk 'soup', kə́t^h 'surface'
f. Lax with R: kǒŋ 'ball', pɛ̌m 'snake', pǎm 'chestnut'
g. Aspirated/Tense with H(H): c^hə̋l 'season', p^hə̋n 'place', t*ɛ̋ 'unreasonable demand'
h. Aspirated/Tense with H(L): k*óc^h 'flower', c*ák 'pair', t^hə́l ~ t^hə̌l 'fur'
i. Aspirated/Tense with R: c^hə̌n 'cloth', t^hǎl 'mask'

We see that the Rise class is significantly underpopulated by aspirated and tense onsets (0.39), while it has a relative surplus of lax onsets (1.40). On the other hand, the H(H) class has the opposite profile with a relatively higher proportion of aspirated and tense (1.25) consonants and a lower proportion of lax consonants (0.85). The H(L) class displays no significant imbalances with respect to the laryngeal category of the onset.

This result leads us to expect that when items with aspirated/tense onsets vacate the R class, they will be attracted to the H(H) class. It turns out that in the original pool of 63 MK Rise-class words, there are only five items with an aspirate/tense onset in the contemporary language. The majority of these have shifted out of the R class into the H(H) class. Table 16 summarizes the correspondences for each of our speakers. These data suggest that while a sonorant coda biases a stem to the R class, the dispreference for an L tone following an aspirated/tense stop is stronger (assuming that the surface Rise tone is phonologically /LH/).

Table 16: MK Rise-class words with an aspirated or tense onset

Item		YD	CP	MA	KM	YJ	Seoul
c^hɛ	'shredded vegetables'	H(H)	H(H)	H(H)	H(H)	R	c^hɛ:
p*al	'bamboo blind'	R					pa:l
p*jəm	'span of hand'	H(H)	H(H)/H(L)	H(H)	H(H)	H(H)	p*jə:m
t*e	'crowd'	H(H)	H(H)	H(H)	H(H)	H(H)	t*e
k*i	'crab'	H(H)	R	R	H(H)		ke:

In order to assess the statistical significance of the various factors reviewed above, we ran a mixed-effects log-linear regression model using the lmer function from the lme4 package (Bates and Maechler 2013) in R (R Development Core Team 2011). The dependent variable was membership in one of the three accent classes (H(H), H(L), R) and the predictor variables were MK accent class (H, L, R), frequency (high, low), onset type (sonorant, aspirated, tense), and coda type (vowel, sonorant, obstruent). A random intercept was set for item and subject. The baseline was set to H(H), high frequency, sonorant onset and vowel coda. The results appear in Table 17. For example, the 12th row (H(L) class: MK L) indicates the effect of the MK L class for the distribution of the Kyengsang H(L) class compared to the baseline Kyengsang H(H) class. All of the factors discussed above that bias a word for or against a particular accent class are significant. As expected, the regular correspondences have much higher z values than irregular ones: e.g. H(L)–MK L (18.21) > H(L)–MK R (5.16). Both aspirated and tense onsets are negatively associated with the Rise class. Coda obstruents are positively associated with the H(L) class and negatively associated with the Rise class. Coda sonorants are positively associated with the Rise class. Finally, low frequency words are biased against the H(L) class, reflecting the tendency for H(H) words to attract items from the H(L) class rather than vice versa.

Table 17: Results of a mixed-effects log-linear regression model (monosyllabic lexicon)

		Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)		-3.21493	0.08053	-39.92	< 2e-16	***
H(L) class		-2.18484	0.18350	-11.91	< 2e-16	***
R class		-2.30665	0.20326	-11.35	< 2e-16	***
MK L		-2.17222	0.11728	-18.52	< 2e-16	***
MK R		-2.39229	0.12960	-18.46	< 2e-16	***
Onset-Asp		-1.53534	0.09823	-15.63	< 2e-16	***
Onset-Tns		-1.38801	0.09253	-15.00	< 2e-16	***
Coda-Obs		0.06037	0.08976	0.67	0.501249	
Coda-Son		0.40130	0.08324	4.82	1.43e-06	***
Frequency-Low		0.17548	0.06859	2.56	0.010518	*
H(L) class:	MK L	3.13567	0.17216	18.21	< 2e-16	***
	MK R	1.29370	0.25057	5.16	2.43e-07	***
	Onset-Asp	-0.48524	0.20498	-2.37	0.017918	*
	Onset-Tns	-0.38125	0.18595	-2.05	0.040335	*
	Coda-Obs	1.01187	0.17040	5.94	2.88e-09	***
	Coda-Son	-0.02756	0.18252	-0.15	0.879981	
	Frequency-Low	-0.43270	0.12857	-3.37	0.000764	***
R class:	MK L	1.13076	0.29825	3.79	0.000150	***
	MK R	3.65342	0.20487	17.83	< 2e-16	***
	Onset-Asp	-2.11961	0.42509	-4.99	6.16e-07	***
	Onset-Tns	-1.66080	0.32221	-5.15	2.55e-07	***
	Coda-Obs	-1.79496	0.32576	-5.51	3.59e-08	***
	Coda-Son	0.50880	0.16601	3.06	0.002178	**
	Frequency-Low	-0.08707	0.14425	-0.60	0.546101	

Signif. codes: ‘***’ 0.001, ‘**’ 0.01, ‘*’ 0.05, ‘.’ 0.1

To briefly summarize our results for monosyllables, we have seen that the three South Kyengsang monosyllabic accent classes reflect their MK cognates with a regularity that approximates their overall size in the lexicon: H(H) > H(L) > R. The H(H) class attracts items from the H(L) and R classes. The words that correspond regularly with their MK sources are biased to higher token frequency items, while the words that transfer their affiliation to the H(H) class tend to be drawn from the lower frequency lexical items. The Rise class is marked by a sonorant coda and has expelled most stems with an obstruent coda. It has also attracted words from the H(H) and H(L) classes that have a sonorant coda (an Island of Reliability effect). Finally, the Rise class is biased against stems that begin with an aspirated or tense onset.

5. Disyllables

The South Kyengsang disyllabic tonal classes are repeated in Table 18. We also add examples of stems ending in a vowel.

Table 18: South Kyengsang disyllabic tonal class

Accent class	Citation	Nominative		Citation	Nominative	
HH	kúrɨ́m	kúrɨ́m-i	‘cloud’	táncí	táncí-ka	‘jar’
HL	mánɨl	mánɨl-i	‘garlic’	nápi	nápi-ka	‘butterfly’
LH(L)	parám	parám-i	‘wind’	namú	namú-ka	‘tree’
LH(H)	sarám	sarám-í	‘person’	papó	popó-ká	‘fool’

Table 19 shows the distribution of disyllabic lexical items across the four tonal classes for each of our five speakers. Each data point represents an individual word for one of our speakers. The total is 4,043 items.

Table 19: Accentual distribution of South Kyengsang disyllabic words

accent \ speaker	YD	CP	MA	KM	YJ	Totals
HH	173	154	163	113	113	716 (18%)
HL	452	326	329	302	277	1,686 (42%)
LH(L)	263	186	222	261	183	1,115 (28%)
LH(H)	134	106	90	65	131	526 (13%)
Totals	1,022	772	804	741	704	4,043

5.1 Middle Korean–South Kyengsang correspondences: disyllables

Table 20 shows the distribution of the individual words in our corpus that have cognates in MK, along with the MK tonal class taken from Ito (2013). We see that the MK final accent LH class accounts for over half of the total (54%). Its regular South Kyengsang reflex HL remains the largest class for all five speakers but has lost ground, accounting now for 42% of the total corpus. The remaining three accent classes have each grown by a small amount. This development is similar to that found in the monosyllables (Table 9), where the dominant MK H class remains the largest but has also lost some ground to the other two contrasting accent patterns.

Table 20: Accent distribution of the South Kyengsang disyllables across the four MK tonal classes

MK accent	MK attested	SK attested	SK accent	SK all
HX	97 (13%)	75 (14%)	HH	716 (18%)
LH	417 (54%)	297 (56%)	HL	1,686 (42%)
LL	206 (27%)	117 (22%)	LH(L)	1,115 (28%)
RX	55 (7%)	38 (7%)	LH(H)	526 (13%)
Totals	775	527		4,043

5.2 Type Frequency

We now examine the words in our corpus with MK cognates so that we can track the shift of lexical items from their expected outcome to another accent class. Again we are interested in whether particular accent classes are attractors or deflectors. Table 21 shows the magnitudes of the correspondences summed over all five speakers. The expected outcomes (shaded) are again the largest. They show the degree of regularity in the correspondences. The MK HX class has the lowest degree of regularity at 56%, while the others fall in the range of 66% ≈ 71%. The relative regularity of the correspondences largely aligns with the size of the class seen in Table 20. The one exception is the HX class.

Table 21: Correspondences between MK and South Kyengsang (disyllabic)

MK \ SK	HH	HL	LH(L)	LH(H)	Totals
HX	166 (56%)	47 (16%)	54 (18%)	27 (9%)	294
LH	89 (8%)	779 (71%)	156 (14%)	74 (7%)	1,098
LL	39 (8%)	76 (16%)	326 (67%)	42 (9%)	483
RX	11 (7%)	8 (5%)	37 (23%)	107 (66%)	163
Totals	305	910	573	250	2,038

(6) Examples of regular correspondences

a. MK HX: cánc^{h}í ‘party’, cáŋsá ‘trade’, p^{h}árí ‘fly’, síkól ‘countryside’, s*ɨ́lkɛ́ ‘gallbladder’

b. MK LH: ísɨl ‘dew’, páŋul ‘bell’, k*óri ‘tail’, hánɨl ‘sky’, cári ‘seat’, núe ‘silkworm’

c. MK LL: tarák ‘loft’, kaɨ́l ‘autumn’, kitúŋ ‘pillar’, carú ‘sack’, cokɛ́ ‘shellfish’, məncí ‘dust’

d. MK RX: kəcɨ̋s ‘lie’, sinɛ̋ ‘brook’, soŋkős ‘gimlet’, amű ‘anyone’, sokjə̋ŋ ‘blind person’

If individual words are attracted to the larger classes, then we predict that exceptions to the regular development should tend to shift to the Kyengsang HL class. And for the Kyengsang HL class itself, exceptions should predominate in the next largest LH(L) class. As shown in Table 21, these predictions seem to be borne out. First, for the MK LH (> Kyengsang HL) class, the majority of exceptions to the regular South Kyengsang HL correspondence are found in the next largest (LH(L)) accent class: 156 items appear in the LH(L) class versus 74 in LH(H) and 89 in HH. Second, for the MK LL > Kyengsang LH(L) class, the largest number of exceptions to the regular development are found in the HL class (76) and then LH(H) (42). For MK HX, the largest number of exceptions to the expected HH correspondence are found in the LH(L) (54) class and then the HL class (47), but this may be due to several words whose accent in the pre-Kyengsang stage was different from the accent attested in MK. See Ito (2013) for details.

Another exception to the generalization that deviations from the expected development track the magnitude of the other accent classes is Kyengsang LH(H), where more items are attracted to LH(L) (37) than to HL (8). But this makes sense because in South Kyengsang, LH(H) and LH(L) are identical in their citation form. If the speaker encounters a word in this form and correctly observes the final accent, he must still guess the tonal contour of the inflected form, and so the predominant LH(L) is a safer choice. This in turn predicts that there will be more items that shift from LH(H) to

LH(L) than from LH(L) to LH(H). This is also correct. The LH(H) > LH(L) change is 37/163 (23%), while LH(L) > LH(H) is 42/483 (9%).[7]

5.3 Token Frequency

Is the token frequency effect we noted for monosyllables in 4.3 observed in disyllabic words as well? In order to answer this question, we divided the disyllables into high and low frequency groups in the manner described in 4.3 (based on the median). Given the reconstruction in (1), each of the four disyllabic accent patterns is the product of a sound change from MK. We therefore expect the regular correspondences to be biased to the high frequency group. This prediction is confirmed (Table 22). For all four accent classes, the regular correspondences (shaded) are drawn more from the “high-frequency” group than from the “low-frequency” group. The overall rate of regular correspondence for the high-frequency group is 79%, while that for the low-frequency group is 54%. A Pearson’s Chi-square test with the Yates continuity correction pitting regular vs. irregular MK-South Kyengsang correspondence against high vs. low frequency group is highly significant: χ^2 = 135.4683, df = 1, p-value < 2.2e-16.

Table 22: Frequency and correspondence with MK (disyllabic)

MK \ SK	HH	HL	LH(H)	LH(L)	Totals
HX (Low)	68 (49%)	20 (14%)	15 (11%)	36 (26%)	139
HX (High)	89 (66%)	26 (19%)	8 (6%)	12 (9%)	135
LH (Low)	60 (11%)	297 (57%)	43 (8%)	122 (23%)	522
LH (High)	29 (6%)	413 (84%)	19 (4%)	32 (6%)	493
RX (Low)	7 (14%)	1 (2%)	25 (50%)	17 (34%)	50
RX (High)	3 (3%)	6 (6%)	72 (73%)	17 (17%)	98
LL (Low)	28 (15%)	40 (22%)	25 (14%)	90 (49%)	183
LL (High)	11 (4%)	36 (12%)	17 (6%)	231 (78%)	295

(7) Examples

a. MK HX (low freq.): uré~urể ‘thunder’, tʰúkú~ tʰukú ‘helmet’, kɨ́mɨ́m~kɨ́mɨm ‘the last day of the month’

b. MK HX (high freq.): ákí ‘baby’, ə́mí ‘mother’, táncí ‘jar’

c. MK LH (low freq.): paŋá~paŋấ ‘mill’, pə́cim~pəcím ‘scabies’, kámtʰu~kamtʰú~kamtʰű ‘horsehair cap’

d. MK LH (high freq.): ípul ‘bedding’, páta ‘sea’, kásɨm ‘breast’

e. MK LL (low freq.): puré~purể ‘air bladder’, kɛám~kɛấm ‘hazelnut’, kɨ́ru~kɨrú ‘stump’

f. MK LL (high freq.): puə́k ‘kitchen’, harú ‘day’, maɨ́m ‘heart’

[7] It is interesting that for the Sino-Korean stratum of the lexicon the LH(H) class is larger than LH(L), reflecting a disparity in the words that entered Korean from Middle Chinese. In tracing the development of these two classes into South Kyengsang, Ito (2014b) finds that more words shifted from the LH(L) class to LH(H) than vice versa. This is exactly the inverse of what we observe for the native vocabulary and implies that whether a word belongs to the native or Sino-Korean region of the lexicon plays a crucial role in determining its phonological behavior. See Ito (2014b) for details.

g. MK RX (low freq.): kálkí~kálki ‘mane’, púc^{h}ú~puc^{h}ú ‘leek’, kecı̋p~ kecíp ‘woman’
h. MK RX (high freq.): ankɛ̋ ‘fog’, ərɨ̋n ‘adult’, paŋkű ‘gas’

5.4 Onset effects

What is effect of the onset consonant? As we saw in 4.5, aspirated and tense consonants tend to raise F0 and are statistically incompatible with the Rise tone. We might therefore expect to find a comparable imbalance in the distribution of the disyllabic accent classes with respect to the initial segment. Table 23 shows the observed over expected values from our data pooled across our five speakers. /s/ and /h/ are excluded from the calculation. The data indicate that aspirated/tense onsets are significantly overrepresented in the tonal classes with an initial high tone (HH and HL) and are significantly underrepresented in the tonal classes with an initial low tone (LH(H) and LH(L)). On the other hand, a lax or sonorant onset is significantly overrepresented in the tonal classes with an initial low tone (LH(H) for sonorant, LH(L) for lax).

Table 23: Disyllabic accent classes with respect to the initial phonemes [χ^2 = 69.5015, df = 6, p = 5.174e-13].

onset \ accent	HH	HL	LH(H)	LH(L)	Totals
Sonorant	203	512	198	356	1,269
Lax	289	725	191	553	1,758
Aspirate/Tense	104	234	33	80	451
Totals	596	1471	422	989	3,478

HH	HL	LH(H)	LH(L)
0.93	0.95	**1.29**	0.99
0.96	0.98	0.90	**1.11**
1.35	**1.23**	**0.60**	**0.62**

5.5 Syllable weight effects

It is well known that in the adaptation of English (Western) loanwords into Kyengsang Korean, the original English accent location is ignored and a default accent is assigned. Various studies have shown that syllable weight has a significant effect on the location of accent in Kyengsang loanwords (N-J. Kim 1997, Kenstowicz and Sohn 2001, H-K. Jun 2006, D-M. Lee 2006, 2009). Heavy syllables attract an H tone, while a word with two light syllables regularly receives penultimate accent.[8] A few examples from H-K. Jun (2006) are cited below.

(8) Examples of loanwords
a. Heavy-Light: mémpə ‘member’, sént^{h}ə ‘center’
b. Light-Heavy: mɛcík ‘magic’, seíl ‘sale’
c. Light-Light: rópi ‘lobby’, t^{h}érə ‘terror’

The penultimate accent could be viewed as a reflex of the preponderance of the HL class in the native grammar. But to the best of our knowledge, no one has thought to test the connection between syllable

[8] For the North Kyengsang dialect, if the initial syllable has a long vowel then it is regularly assigned to the double-H class: *marmalade* > má:málleitɨ (Kenstowicz and Sohn 2001).

weight and an H tone in the native South Kyengsang lexicon.[9] Since long vowels are not found in South Kyengsang disyllables, the only possible effect of syllable weight is in the presence and nature of a coda consonant. Table 24 shows the distribution of the 4,043 data points from our five speakers with respect to syllable weight (Heavy = (C)VC, Light = (C)V). We have collapsed together the LH(L) and LH(H) classes.

Table 24: Disyllabic accent classes with respect to syllable weight [χ^2 = 121.1467, df = 6, p < 2.2e-16].

weight \ accent	HH	HL	LH	Totals
Heavy-Heavy	115 (23%)	151 (30%)	244 (48%)	510
Heavy-Light	217 (27%)	323 (40%)	273 (34%)	813
Light-Heavy	183 (13%)	589 (42%)	633 (45%)	1,405
Light-Light	201 (15%)	623 (47%)	491 (37%)	1,315
Totals	716 (18%)	1,686 (42%)	1,641 (41%)	4,043

HH	HL	LH
1.27	**0.71**	**1.18**
1.51	0.95	**0.83**
0.74	1.01	**1.11**
0.86	**1.14**	0.92

(9) Examples[10]

a. Heavy-Heavy: kə́kcə́ŋ ‘anxiety’, tóŋan ‘period’, talkjál ‘egg’, həŋ.kə̋p^{h} ‘rag’

b. Heavy-Light: ə́kcí ‘stubbornness’, tɛ́ŋki ‘ribbon’, tanc^{h}ú ‘button’, an.kɛ̋ ‘fog’

c. Light-Heavy: ə́mə́m ‘mother’, mósɨp ‘appearance’, kərɨ́m ‘manure’, kə.cɨ̋s ‘lie’

d. Light-Light: súc^{h}ɛ́ ‘sewer’, páci ‘trousers’, muní ‘pattern’, pa.pő ‘fool’

As expected, when syllable weight is not at play (Light-Light), the HL accent class is significantly overrepresented (1.14). For Light-Heavy words, final accent LH is significantly overrepresented (1.11). Another, more surprising effect is that the HH class is significantly overrepresented in Heavy-Light structures (1.51) and significantly underrepresented in Light-Heavy (0.74). This distribution may reflect biases in the MK lexicon from which the native South Kyengsang stems derive. We recall from section 3 that the HH accent corresponds to the MK HX tonal contour. Ramsey (1991) points out that HX was a minority pattern in MK, where the majority of words had final accent LH. A Heavy-Light syllable profile would constitute the most favorable circumstance for the pitch peak to be attracted away from the right edge of the word. If this is the primary origin of the MK HX class, then these biases showing up in the current lexicon make some sense. See Ito (2013) for detailed discussion of MK accent.

In order to assess the joint roles of syllable weight and the laryngeal character of the onset consonant in biasing the distribution of the three accent classes (HH, HL, LH (aggregated)), we ran a mixed-effects log-linear regression model using the lmer function in R. Weight (Heavy vs. Light) in the penult and final syllables as well as the laryngeal category of the word-initial consonant (lax+sonorant vs. aspirated + tense) were the predictor variables, and accent class (HH, HL, LH (aggregated)) was the dependent variable. A random intercept was set for subject. The baseline was set to HL with a lax/sonorant onset. Table 25 shows three noteworthy results. First, a heavy penult has a significant positive effect on the HH class, which coincides with the association of penult weight and an initial H tone seen in Table 24. Second, a heavy final syllable has a significant positive effect on the LH accent, mimicking the overrepresentation of the LH accent with a final heavy seen in Table 24.

[9] That is, until H-J. Kim (2012), whose research parallels our own and reaches similar conclusions for the North Kyengsang dialect. See section 8 below for discussion.

[10] The velar nasal occupies the syllable coda in Korean.

Finally, there is a negative effect of an aspirated/tense onset on the LH accent class, replicating the effect seen in Table 23. There is no significant effect of an aspirated/tense onset on the HH class, since both the HH and the baseline HL classes are overrepresented when the onset is aspirated/tense.

Table 25: Result of a mixed-effects log-linear regression model (disyllabic lexicon)

		Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)		-2.24282	0.07092	-31.623	< 2e-16	***
HH class		-1.11726	0.07226	-15.461	< 2e-16	***
LH class		-0.20728	0.05532	-3.747	0.000179	***
Penult-Heavy		-0.93883	0.05417	-17.33	< 2e-16	***
Final-Heavy		-0.24560	0.04908	-5.004	5.60e-07	***
Onset-Asp-Tns		-1.82538	0.07044	-25.913	< 2e-16	***
HH class:	Penult-Heavy	0.79333	0.09247	8.579	< 2e-16	***
	Final-Heavy	-0.09278	0.09031	-1.027	0.304283	
	Onset-Asp-Tns	0.05304	0.12733	0.417	0.676971	
LH class:	Penult-Heavy	0.15402	0.07647	2.014	0.043999	*
	Final-Heavy	0.38114	0.07019	5.430	5.63e-08	***
	Onset-Asp-Tns	-0.78897	0.12174	-6.481	9.12e-11	***

Signif. codes: ‘***’ 0.001, ‘**’ 0.01, ‘*’ 0.05, ‘.’ 0.1

6. Trisyllables

Trisyllables occur in just four accent patterns in South Kyengsang and parallel the disyllables in certain ways. There is an initial double-H class HHL that is the counterpart to disyllabic HH. The disyllabic HL contour is matched by trisyllabic HLL and LHL. Finally, trisyllabic LHH is the counterpart of disyllabic LH(H) as well as LH(L). The former (i.e. LHH) derives from MK trisyllables with a word-initial Rise (cf. North Kyengsang *má:núra* ‘wife’). South Kyengsang LHH is also cognate with North Kyengsang LLH (cf. *mintɨllé* ‘dandelion’). In contrast to North Kyengsang, in South Kyengsang the pitch peak spreads leftward up to the peninitial syllable of the word/phrase. Hence LLH is not a possible output in South Kyengsang (Lee and Davis 2009).[11] Table 26 shows the structural correspondences between the disyllables and trisyllables.

Table 26: South Kyengsang tonal classes (disyllable, trisyllable)

Disyllables			Trisyllables		
HH	kúrɨ́m	‘cloud’	HHL	múcíkɛ	‘rainbow’
HL	mánɨl	‘garlic’	HLL	ménɨri	‘daughter-in-law’
			LHL	apə́ci	‘father’
LH(L)	parám	‘wind’	LHH	mintɨ́llé	‘dandelion’
LH(H)	sara̋m	‘person’	LHH	manúrá	‘wife’

[11] In actuality, Co (2000) reports that the South Kyengsang Changnyeng dialect has the LLH class, which corresponds with the North Kyengsang LLH class. Probably the Changnyeng dialect shows a more conservative accent system than Pusan/Changwen.

Our corpus contains 370 trisyllabic forms. They occur with the distribution shown in Table 27 across our five South Kyengsang speakers. Some of the words we queried our consultants about were rare or obsolete. On average they knew about 70%. Recall from Table 8 that trisyllables had the largest agreement rate 74%. The LHL tonal class is by far the largest (78%), with LHH, HHL and HLL appearing in more or less equally smaller increments.

Table 27: Accentual distribution of South Kyengsang trisyllabic words

accent \ speaker	YD	CP	MA	KM	YJ	Totals
HHL	25	19	11	9	17	81 (5%)
HLL	14	8	14	6	8	50 (3%)
LHL	260	235	244	249	198	1,186 (78%)
LHH	43	40	56	35	35	209 (14%)
Totals	342	302	325	299	258	1,526

6.1 Middle Korean–South Kyengsang correspondences: trisyllables

These four South Kyengsang accent classes derive from five MK accent patterns. Assuming the regular sound changes discussed in section 3, the following correspondences are predicted.

Table 28: MK–South Kyengsang correspondence and sound change

MK			South Kyengsang
HXX	H doubling	→	HHL
LHL	H retraction	→	HLL
LLH	H retraction	→	LHL
LLL	H insertion + leftward spreading	→	LHH
RXX	Rise decomposition + H doubling	→	LHH

As mentioned above, the South Kyengsang LHH contour thus derives from two different sources: MK LLL, which became LLH to satisfy culminativity and then LHH by tone spread (recall South Kyengsang *mintɨllé* vs. North Kyengsang *mintɨllé* 'dandelion'), and MK RXX, which became LHX by Rise decomposition and then LHH by H doubling. Table 29 shows the correspondences between MK and South Kyengsang summed across the five speakers.

Table 29: Correspondences between MK and South Kyengsang (trisyllabic)

MK \ SK	HHL	HLL	LHL	LHH	Totals
HXX	27 (35%)	0 (0%)	48 (62%)	2 (3%)	77
LHX	7 (4%)	33 (20%)	116 (70%)	9 (5%)	165
LLH	2 (1%)	0 (0%)	189 (97%)	3 (2%)	194
LLL	0 (0%)	0 (0%)	4 (17%)	20 (83%)	24
RXX	2 (5%)	0 (0%)	20 (51%)	17 (44%)	39
Totals	38	33	377	51	499

(10) Examples of regular correspondences

a. MK HXX: hálmə́ni 'grandmother', ə́mə́ni 'mother', múcíkɛ~mucíkɛ 'rainbow'
b. MK LHX: ə́cək*e 'yesterday', ménɨri 'daughter-in-law', áhɨre~ahɨ́re 'nine days'
c. MK LLH: haŋári 'jar', pitúlki 'dove', ipáci 'contribution', k*amákwi 'raven'
d. MK LLL: poksúŋá 'peach', pusɨ́rə́m 'ulcer', k*ucírám 'scolding'
e. MK RXX: manúrá 'wife', samákwí 'mole', ərɨ́sín 'esteemed elder'

The largest MK class was LLH with a final H (just as in disyllables and monosyllables) and this accounts for part of the preponderance of LHL in South Kyengsang. But in addition the LHL class has attracted items from most of the other accent classes, outweighing the correspondences predicted by the regular sound changes. The biggest disparity is in the evolution of the MK LHL class, where the predicted HLL with retraction only accounts for 20% of the data. The vast majority of words in this class (70%) appear with a penultimate peak. This may indicate that the Kyengsang retraction of H was primarily from the final syllable (Nonfinality) rather than a general sound change.[12] If so, then MK LLH would merge with MK LHX to become the largest class. If the MK LHX (143) and LLH (176) classes are combined, then the disparity with the remaining accent classes becomes quite large, as shown in the Table 30. Penultimate H becomes over five times larger than the closest of the remaining classes (HXX, 14%). Earlier we observed that for the monosyllables and disyllables, the largest MK classes actually lost ground in their evolution into South Kyengsang, although irregular correspondences were more frequently observed in the lower type frequency class. The trisyllables exhibit the opposite trend, with the most frequent type attracting items from the other three classes to grow in size. We may conclude from these two different evolutionary paths that an accent class must enjoy an initial significant competitive advantage in order to grow and become the predominant default class via analogical change. See Yang (2005) for discussion along these lines.

Table 30: MK accent distribution (trisyllable)

MK	Number	Ratio
HXX	60	14%
LHX + LLH	319	73%
LLL	29	7%
RXX	28	6%

6.2 Onset effects

Is there an effect of the onset consonant in the trisyllabic lexicon as well? Table 31 shows the correlation between the onset of the initial syllable and accent class. As in disyllabic words, when the onset is aspirate or tense, the HHL class is overrepresented while the LHH class is underrepresented. However, an aspirate/tense onset is also significantly underrepresented in the HLL class, whereas a sonorant onset is overrepresented in this class. This is probably an accidental result reflecting a correlation in an earlier stage of Korean: recall that the South Kyengsang HLL class corresponds to

[12] Cross-linguistically tonal assimilation (spreading) is predominantly perseverative (Hyman 2002). Leftward H shift or spread is typically motivated by tonal crowding from a boundary tone or attraction to a heavy syllable.

MK LHX. Thus, we do not find a general association between an aspirate/tense onset and an H tone in the initial syllable, and only an association with the HHL class is observed.

Table 31: Trisyllabic accent classes with respect to initial onsets [χ^2 = 69.5027, df = 6, p = 5.171e-13].

onset \ accent	HHL	HLL	LHL	LHH	Totals
Sonorant	26	32	325	71	454
Lax	15	14	503	94	626
Aspirated/Tense	24	3	246	10	283
Totals	65	49	1,074	175	1,363

HHL	HLL	LHL	LHH
1.20	**1.96**	0.91	1.22
0.50	0.62	1.02	1.17
1.78	**0.29**	1.10	**0.28**

Concerning the onset of the peninitial syllable, as seen in Table 32, no significant effect of an aspirate/tense consonant is observed. There is a positive correlation between a lax onset and the LHH class, while a sonorant onset shows a negative correlation with the same class. It is difficult to give a phonetic rationale for this relation.

Table 32: Trisyllabic accent classes with respect to peninitial onsets [χ^2 = 23.0198, df = 6, p = 0.0007899].

onset \ accent	HHL	HLL	LHL	LHH	Totals
Sonorant	32	14	446	51	543
Lax	28	23	436	106	593
Aspirated/Tense	13	7	195	24	239
Totals	73	44	1,077	181	1,375

HHL	HLL	LHL	LHH
1.11	0.81	1.05	**0.71**
0.89	1.21	0.94	**1.36**
1.02	0.92	1.04	0.76

To summarize, in the trisyllabic lexicon, an association between an aspirate/tense onset in the initial syllable and the HHL class is observed. Onset types in the peninitial syllable do not show a phonetically natural association.

6.3 Weight effects

Table 33 shows the correlation between the syllable weight structure and accent class. The vast majority of words fall in the LHL and LHH accent classes. Many words in the HHL and HLL classes have variants in LHL. Focusing on the dominant patterns, we see that LHH predominates over LHL when the final syllable is heavy. Otherwise, the LHL contour prevails. The trisyllables thus present a profile that approximates the 'default' accent that appears in English loanwords (see 5.5).

Table 33: Correlation between syllable weight structure and accent class (trisyllable)

weight \ accent	HHL	HLL	LHL	LHH	Totals
Heavy-Heavy-Light	7 (4%)	3 (2%)	111 (66%)	48 (28%)	169
Heavy-Light-Heavy			1 (10%)	9 (90%)	10
Heavy-Light-Light	30 (8%)	1 (0%)	346 (89%)	12 (3%)	389
Light-Heavy-Heavy			1 (14%)	6 (86%)	7

Light-Heavy-Light	5 (3%)	1 (1%)	157 (92%)	7 (4%)	170
Light-Light-Heavy	4 (4%)	1 (1%)	15 (16%)	76 (79%)	96
Light-Light-Light	35 (5%)	44 (6%)	555 (81%)	51 (7%)	685
Totals	81	50	1,186	209	1,526

(11) Examples

a. Heavy-Heavy-Light: olc^{h}ɛ́ŋi ‘tadpole’, cintállɛ́ ‘azalea’

b. Heavy-Light-Heavy: malmícal~malmícál ‘sea anemone’, simpúrɨ́m ‘errand’

c. Heavy-Light-Light: olkámi ‘trap’, maktɛ́ki ‘stick’

d. Light-Heavy-Heavy: hotɨ́lkáp ‘making a great fuss’

e. Light-Heavy-Light: kaúnte ‘middle’, sikɨ́mc^{h}i ‘spinach’

f. Light-Light-Heavy: tosírák ‘lunch box’, put*úmák ‘cooking fireplace’

g. Light-Light-Light: p*ək*úki ‘cuckoo’, turúmi ‘crane’, sat^{h}úri~sat^{h}úrí ‘dialect’

In order to probe the syllable weight effect, Table 34 focuses on the weight in the final two-syllable window with respect to the two major accent classes (LHL and LHH). The small number of examples with the Heavy-Heavy syllable structure are excluded. As can be seen, for the Light-Heavy syllable profile, the LHH tonal contour is significantly overrepresented while LHL is underrepresented. The opposite correlation is observed in the Light-Light structure.

Table 34: Correlation between syllable weight structure and accent class (trisyllable, final two-syllable window) [$\chi^2 = 443.1057$, df = 2, $p < 2.2e\text{-}16$].

weigh \ accent	LHL	LHH	Totals
Heavy-Light	268	55	323
Light-Heavy	16	85	101
Light-Light	901	63	964
Totals	1,185	203	1,388

LHL	LHH
0.97	1.16
0.19	**5.75**
1.09	**0.45**

We ran a mixed-effects log-linear regression model using the lmer function in R to test the factors of syllable weight on the penult and final syllables and the laryngeal character of both the initial and penultimate onset consonant on the distribution of accent among the four trisyllabic classes in the data in Table 27. A random intercept was set for subject, and LHL was the baseline. The results appear in Table 35. The most noteworthy effects ($p < 0.001$) are that a final heavy syllable significantly biases a word to the LHH accent pattern and that an initial aspirated-tense onset is negatively associated with the LHH accent class. At a lower level of significance ($p < 0.05$) an initial tense/aspirated consonant was biased to the HHL class, but the same class of consonants was negatively associated with the LHH at the onset of the penultimate syllable. There thus does not appear to be any consistent relation between the double H accent and the onset of the syllable.

Table 35: Result of a mixed-effects log-linear regression model (trisyllabic lexicon)

		Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)		-1.15878	0.05853	-19.797	< 2e-16	***
HHL class		-2.86419	0.18695	-15.321	< 2e-16	***
HLL class		-2.73752	0.18327	-14.937	< 2e-16	***
LHH class		-2.24644	0.13351	-16.826	< 2e-16	***
Initial Ons-Asp-Tns		-1.27520	0.07734	-16.489	< 2e-16	***
Penult Ons-Asp-Tns		-1.56123	0.08438	-18.502	< 2e-16	***
Penult-Heavy		-1.13708	0.07451	-15.261	< 2e-16	***
Final-Heavy		-4.03685	0.24472	-16.496	< 2e-16	***
HHL class:	Initial Ons-Asp-Tns	0.66001	0.28817	2.290	0.0220	*
	Penult Ons-Asp-Tns	-0.25116	0.39063	-0.643	0.5202	
	Penult-Heavy	-0.67532	0.38862	-1.738	0.0823	.
	Final-Heavy	0.01154	1.03834	0.011	0.9911	
HLL class	Initial Ons-Asp-Tns	-1.31511	0.60373	-2.178	0.0294	*
	Penult Ons-Asp-Tns	-0.25798	0.44823	-0.576	0.5649	
	Penult-Heavy	-1.14021	0.53039	-2.150	0.0316	*
	Final-Heavy	0.29926	1.04119	0.287	0.7738	
LHH class	Initial Ons-Asp-Tns	-1.34949	0.33660	-4.009	6.09e-05	***
	Penult Ons-Asp-Tns	-0.77912	0.30250	-2.576	0.0100	*
	Penult-Heavy	0.34074	0.19265	1.769	0.0769	.
	Final-Heavy	3.73682	0.29588	12.629	< 2e-16	***

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

To summarize our survey of the South Kyengsang native lexicon for nouns, we found several several major trends. First, for both disyllables and trisyllables there is a significant bias in favor of penultimate accent when the final two syllables are light. Second, a heavy final syllable biases the word towards LH/LHH. Third, a heavy penultimate syllable is also positively associated with the double-H accent in disyllabic (HH) stems but not in trisyllabic (LHH) stems.[13] Finally, an initial aspirated or tense onset is biased against a low tone but does not show any general preference between HL and HH contours.

7. Generalizing from the Lexicon

The results from the previous three sections showed that syllable weight, aspiration, and sonorant vs. obstruent coda are significantly correlated with the statistical distribution of accent classes for various of the one, two, and three-syllable native nouns in the South Kyengsang lexicon. We may wonder if these factors are simply the by-products of earlier sound changes unbeknownst to the linguistically naïve native speaker or are they reflections of the native speaker's tacit knowledge about the structure of South Kyengsang accent? The classic answer to these questions is to demonstrate that such predictors are called into play when the speaker is set the task of assigning a linguistic property like accent to a new word. The adaptation of loanwords is one such test; another is the novel word (so-

[13] The three monosyllabic tonal patterns are distributed equally across CV (H(H) 54%, H(L) 26%, R 20%) and CVC (H(H) 53%, H(L) 26%, R 21%). Thus, there is no weight effect in monosyllables.

called Wug) test (Gleason 1958). In this section (and the next), we report the results of a novel word study. In section 9, we summarize what is known about the accentual adaptation of loanwords into South Kyengsang.

7.1 Novel Word Study

Following the methodology employed by Zuraw (2010), we constructed a list of novel word stimuli, which were presented to a group of South Kyengsang subjects living in Pusan who were asked to report what accent they would assign to these lexical items. We instructed the subjects that the stimuli were rare or obsolete words that they probably would not know and that they should answer the question based on their personal knowledge of South Kyengsang Korean. We constructed our stimuli by computing the bigram transitional probabilities of lexical items in our corpus. The algorithm was run over the words in our corpus to compute their bigram transition scores.[14] Based on the calculated log bigram transitional probabilities, the algorithm generated a list of probable words. We selected items from among the words that had the 500 highest scores. We thus have some confidence that the stimuli sound like normal native South Kyengsang nouns. For disyllables, 40 novel words were selected: 10 each for Heavy-Light and Light-Heavy syllable structures and 20 for Light-Light. The latter group was divided into 10 with an initial aspirated or tense onset and 10 with a lax or nasal onset. 40 trisyllabic words were also selected in which the final two syllables were similarly varied: 10 each for Light-Heavy-Light and Light-Light-Heavy and 20 for Light-Light-Light, broken down into 10 with an aspirated/tense and 10 with lax/nasal for the onset of the penultimate syllable. Finally, there were 12 CVC monosyllables split between 8 with an obstruent coda and 4 with a sonorant coda. We attempted to balance our stimuli to control for lexical neighbors (one phoneme change). The following is the list of our stimuli. NC denotes the neighbor count.

Table 36: List of novel words for the experiment (monosyllables)

Obstruent coda		NC	Sonorant coda		NC
눅	nuk	6	굼	kum	8
넙	nəp	7	훌	hul	7
걱	kək	9	뭄	mum	7
둣	tus	8	슴	sɨm	12
맙	map	10			
봋	poc^h	7			
뎃	tes	6			
줏	cus	9			
	Av.	7.75		Av.	8.5

[14] Some of the bigrams had a frequency of 0. As a result, some lexical items happen to contain unattested bigrams. Smoothing techniques avoided assigning these items a probability of zero by estimating probabilities for unseen sequences. We employed Witten-Bell discounting (see Jurafsky and Martin 2000 for the application) and made our calculation with the SRILM toolkit (Stolcke 2002) on our corpus.

Table 37: List of novel words for the experiment (disyllables)

CVC.CV		NC	CV.CVC		NC	C^{h}V.CV/C*V.CV		NC	CV.CV		NC
할모	halmo	13	이범	ipəm	6	까나	k*ana	3	보니	poni	8
알보	alpo	6	노달	notal	7	쪼마	c*oma	5	노메	nome	5
엄기	əmki	7	수겁	sukəp	6	써기	s*əki	5	기노	kino	5
장누	caŋnu	8	가녹	kanok	9	터구	t^{h}eku	6	대히	tɛhi	2
솔모	solmo	9	기줄	kicul	7	또지	t*oci	4	조바	copa	8
덜비	təlpi	5	미살	misal	5	쏘배	s*opɛ	7	바미	pami	4
볼래	pollɛ	3	배날	pɛnal	4	따보	t*apo	6	다개	takɛ	6
남배	nampɛ	4	사둑	satuk	12	포모	p^{h}omo	12	무너	munə	1
놀리	nolli	6	너벅	nəpək	5	푸무	p^{h}umu	5	나보	napo	6
뭉미	muŋmi	3	마문	mamun	7	뽀노	p*ono	7	호비	hopi	11
	Av.	6.4		Av.	6.8		Av.	6		Av.	5.6

Table 38: List of novel words for the experiment (trisyllables) [15]

CV.CVC.CV		NC	CV.CV.CVC		NC	CV.C^{h}VCV/ CV.C*V.CV		NC	CV.CV.CV		NC
이덕무	itəkmu	1	나부독	naputok	2	서태로	sət^{h}ɛro	0	가보도	kapoto	3
자삿노	casasno	0	부지덕	pucitək	3	지포시	cip^{h}osi	0	노구마	nokuma	1
고줄니	koculni	0	소마굴	somakul	4	기푸보	kip^{h}upo	1	무저니	mucəni	0
모좀미	mocommi	0	구저반	kucəpan	2	바꺼래	pak*ərɛ	0	노대보	notɛpo	3
서덤비	sətəmpi	0	미조길	micokil	4	도포리	top^{h}ori	3	부지구	puciku	2
나죽모	nacukmo	0	사미밤	samipam	1	이푸무	ip^{h}umu	0	다모저	tamocə	1
고돈부	kotonpu	3	노지벅	nocipək	0	저코니	cək^{h}oni	1	고노디	konoti	3
아잡미	acapmi	0	가주낫	kacunas	0	거푸지	kəp^{h}uci	2	사나비	sanapi	2
서린매	serinmɛ	0	버무삼	pəmusam	2	노태소	not^{h}ɛso	0	자비니	capini	0
소발리	sopalli	1	이조솔	icosol	1	마꾸미	mak*umi	2	아고노	akono	3
	Av.	0.5		Av.	1.9		Av.	0.9		Av.	1.8

The stimuli were presented in written form on a local network in the Hangul script to a group of 46 adult residents of Pusan who are employees in a major construction company (38 males and 8 females, average age 42). The nonce words appeared in their isolation form on computer screens at the subjects' work stations and then in a sentence frame as direct objects of the transitive verbs 'I/he like' or 'I/he bought', where they took the accusative case inflection *-ɨl*. This inflected form allowed us to disambiguate the H(H) vs. H(L) and LH(L) vs. LH(H) accent classes. Subjects were asked to choose

[15] Our stimuli were presented in orthographic form. In CV.CVC.CV, five stimuli (itəkmu, casasno, nacukmo, acapmi, koculni) contain a sequence of segments where a phonological alternation is applied: nasalization (obstruent coda + nasal onset → nasal coda + nasal onset) and lateralization (liquid coda + nasal onset → liquid coda + liquid onset). The spelling of these test items may imply that these words are compound words, since the written forms of simplex nouns do not contain these sequences as a rule. However, the test results were more or less the same regardless of whether the test words contain a disallowed sequence or not: for the words that contain a disallowed sequence, HHL = 6, HLL = 3, LHH = 10, LHL = 185; for the words that do not contain a disallowed sequence, HHL = 6, HLL = 3, LHH = 16, LHL = 182. A two-sample t-test showed no significant difference between the two ($t = 0.012$, $df = 6$, p-value = 0.9908). Thus, it is safe to use all the data from the CV.CVC.CV structure for the analysis.

among the different possible accent patterns made available by South Kyengsang grammar for the accusative form by clicking on the appropriate pattern. For example, for the disyllabic novel word *notal-ɨl* there were five options: HHL, HLL, LHL, LHH, as well as a fifth category ‘other’, where subjects could type in an alternative accent pattern if they disagreed with the choices offered. Subjects were familiarized with these accentual contour labels in an instruction session before beginning the test by being shown some common words whose syllables were explicitly labeled as to tonal pattern. In the test, once the subjects made their selection for a given nonce word, they were passed to the next word. There was no limitation on the amount of response time. Most subjects completed the test of 94 words within 25 minutes (mean = 23, min = 14, max = 34). Our test also contained 10 common real words such as *parám* ‘wind’—words for which there was no disagreement in accent class among our five South Kyengsang consultants. These real words acted as controls allowing us to gauge whether the subjects understood the experimental task. All but three of the subjects gave correct answers to all of these real words. The responses of the three subjects who failed to respond correctly to the real words were excluded from the analysis. The order of presentation of the 104 items was randomized for each subject.

7.2 Results: monosyllables

The results for monosyllables are shown in Table 39 below.

Table 39: Wug test results (monosyllable)

coda \ accent	H(H)	H(L)	R	Other	Totals
Obstruent coda	202 (58%)	72 (21%)	50 (15%)	20 (6%)	344
Sonorant coda	69 (40%)	9 (5%)	84 (49%)	10 (6%)	172
Totals	271 (53%)	81 (16%)	134 (26%)	30 (6%)	516

Subjects had four options to assign to each novel word: H(H) (like *múl*, *múl-ɨ́l*), H(L) (like *súl*, *súl-ɨl*), R (like *tǒn, ton-ɨ́l*), and finally ‘other’. The vast majority (94%) chose among the three classes offered by native grammar. Recall that our survey of the lexicon revealed several generalizations about the distribution of the three accent classes in monosyllables. First, H(H) predominated over H(L) both in magnitude and as an attractor from the other classes. If subjects were assigning accent based on their implicit understanding of the statistical distribution of nominal accent classes in their lexicons, we would expect the H(H) pattern to predominate. This expectation is matched by the results. Overall, 53% of the total number of responses fall into the H(H) class. We also see evidence for the affinity between sonorant codas and the R class: novel words with sonorant codas are predominantly assigned to the Rise class (49%). Moreover, this class is disprefered by obstruents codas, having just 15% of the 344 total. To see the statistical significance of the results, we ran a mixed-effects log-linear regression model using the lmer function in R. Coda consonant (obstruent vs. sonorant) was sum coded and entered as a predictor. A random intercept was set for items and random intercepts and slopes were set for subjects. With the baseline set to the H(H) class and an obstruent coda, the Rise class was significantly biased to a sonorant coda, while the H(L) class showed a significant trend towards an obstruent coda. The subjects’ responses thus seem to demonstrate a cognizance of the complementarity between the coda consonant and the H(L) vs. R classes mentioned in 4.4.

Table 40: Results of a mixed-effects log-linear regression model (wug monosyllable)

		Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)		-0.93787	0.07036	-13.329	< 2e-16	***
H(L) class		-1.05217	0.14160	-7.431	1.08e-13	***
R class		-1.40311	0.15942	-8.801	< 2e-16	***
Coda-Son		-1.07528	0.13968	-7.698	1.38e-14	***
H(L) class:	Coda-Son	-1.01506	0.38434	-2.641	0.00826	**
R class:	Coda-Son	1.59852	0.22700	7.042	1.90e-12	***

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

7.3 Results: disyllables

The results for disyllables are shown in Table 41, which indicates how our subjects distributed the four accent classes with respect to the syllable-weight profiles we tested. CVC.CV indicates the Heavy-Light structure and CV.CVC the Light-Heavy structure. CV.CV shows the results for Light-Light stimuli in which the first consonant is a lax consonant or nasal, whereas C^hV.CV indicates an aspirated or tense onset.

Table 41: Wug test results (disyllable)

accent / weight	HH	HL	LH(L)	LH(H)	Other	Totals
CVC.CV	90 (21%)	261 (61%)	57 (13%)	13 (3%)	9 (2%)	430
CV.CVC	73 (17%)	51 (12%)	290 (67%)	14 (3%)	2 (0%)	430
CV.CV	29 (7%)	349 (81%)	41 (10%)	11 (3%)	0 (0%)	430
C^hV.CV	280 (65%)	114 (27%)	31 (7%)	2 (0%)	3 (1%)	430
Totals	472 (27%)	775 (45%)	419 (24%)	40 (2%)	14 (1%)	1,720

We recall from the lexical survey that in South Kyengsang the HL class predominated for disyllables. Also the LH(L) class predominated over LH(H) both numerically as well as in shifts between the two classes. Both of these lexical trends are seen in the novel words results: HL is the largest class overall (45%), while LH(H) is the smallest (2%).[16] Our survey of the lexicon also revealed two phonological properties that biased a word to one of the accent classes. First, aspirated/tense onsets were significantly underrepresented in the LH classes, while they were overrepresented in the HH and HL classes. This bias shows up in the novel words. For both LH classes, their share of the total drops when the onset is changed from lax-sonorant to aspirated-tense. But the more dramatic effect is found with HL vs. HH. The HL class predominates with lax-sonorant onsets (349 vs. 29 or 81% vs. 7%) but reverses position with an aspirated-tense onset (114 vs. 280 or 27% vs. 65%). This association between an aspirated/tense onset and the double-H accent profile was observed in the lexical survey, but it appears more strongly in the novel words, in that the ratio of the HH class is higher than that of the HL class: in the lexicon, the HL class is larger than the HH class

[16] It is interesting that the double H arising from the addition of an inflectional suffix is preferred in the monosyllables (Table 39) but highly dispreferred in disyllables (Table 41). This difference has no plausible explanation in markedness terms and reflects the correlation with the size of these accent classes in mono- vs. di-syllables. It also shows that whether a word is one or two syllables is a significant factor in deciding how to assign an accentual structure to a novel word.

even when the onset was an aspirate/tense. Second, as far as syllable weight is concerned, we recall from Table 24 that a Light-Heavy syllable profile in the lexicon attracts words to the LH classes (45%), while the inverse Heavy-Light profile is biased against LH (34%) and favors HL (40%); a heavy syllable in the penult is also positively associated to the HH class. The syllable weight effect also appears in the novel word experiment. In Light-Heavy structures, the LH(L) class predominates over HL (290 vs. 51 or 67% vs. 12%), while in Heavy-Light structures, the combined LH(L) and LH(H) account for only 16% of the total (57+13), the rest being shared between default HL at 261 (61%) and HH at 90 (21%).

In order to test the effects of syllable weight and onset consonant on the choice of accent class, we ran two separate mixed-effects log-linear regressions using the lmer function in R with subject and item as random factors.[17] The minority LH(H) accent class was excluded from the analysis. The first test was run over the data shown in the first three columns and rows in Table 41. The predictor variables were syllable weight in the penultimate and final syllables and the dependent variable was accent class: HH, HL, LH(L). Random intercepts were set for items and subjects. The base line was set to HL. The test revealed that a final heavy significantly boosted both the LH(L) and HH responses compared to the baseline HL, while a penult heavy decreased the LH(L) response compared to the baseline HL. A penultimate heavy did not significantly affect the HH response relative to the baseline HL. This differs from the lexical trend noted in Table 25, where an initial heavy in a disyllable was positively associated with the double-H accent. Thus, our subjects did not project this lexical trend into their novel words (perhaps because it has no phonetically natural basis). On the other hand, the association between a final heavy and the HH class is an emergent property in the novel words and was not found in the lexicon.

Table 42: Results of a mixed-effects log-linear regression model (wug disyllable, weight)

		Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)		-1.25123	0.05125	-24.415	< 2e-16	***
HH class		-1.76456	0.12500	-14.116	< 2e-16	***
LH(L) class		-1.48508	0.11521	-12.890	< 2e-16	***
Penult-Heavy		-0.42694	0.07957	-5.366	8.07e-08	***
Final-Heavy		-2.48164	0.14576	-17.025	< 2e-16	***
HH class:	Penult-Heavy	0.30177	0.16506	1.828	0.0675	.
	Final-Heavy	1.99297	0.20820	9.572	< 2e-16	***
LH(L) class:	Penult-Heavy	-1.33213	0.16401	-8.122	4.57e-16	***
	Final-Heavy	3.56646	0.18682	19.091	< 2e-16	***

Signif. codes: ‘***’ 0.001, ‘**’ 0.01, ‘*’ 0.05, ‘.’ 0.1

The second test was run over the data shown in the last two rows in Table 41 to assess the effects of the onset consonant. Following the suggestion of a reviewer, we tested whether tense and aspirated consonants would have a parallel effect on the choice of accent, as assumed to this point. The laryngeal character of the initial consonant was the predictor variable (tense vs. aspirated vs. lax-sonorant) and the dependent variable was the same three accent classes. A random intercept was set for subjects. The baseline was HL with an aspirated onset. The test results seen in Table 43 indicate that changing the onset to lax-sonorant is negatively correlated with the HH response, while changing

[17] It was not possible to combine these into a single test because the wug word stimuli unfortunately did not include CVC.CV words with an initial tense or aspirated consonant.

the onset to tense had no significant effect, indicating that tense and aspirated consonants had a similar effect on the subjects' responses.

Table 43: Results of a mixed-effects log-linear regression model (wug disyllable, onset)

		Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)		-3.3896	0.1857	-18.254	< 2e-16	***
HH class		1.1545	0.213	5.421	5.92e-08	***
LH(L) class		-1.4214	0.4211	-3.375	0.000737	***
Onset-Lax/Son		2.4878	0.1933	12.873	< 2e-16	***
Onset-Tns		1.0754	0.2151	5.000	5.72e-07	***
HH class:	Onset-Lax/Son	-3.6423	0.2876	-12.665	< 2e-16	***
	Onset-Tns	-0.3607	0.2499	-1.444	0.148865	
LH(L) class:	Onset-Lax/Son	-0.7201	0.4523	-1.592	0.111375	
	Onset-Tns	0.1563	0.4804	0.325	0.744849	

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

7.4 Results: trisyllables

We recall that the LHL accent pattern predominates in the South Kyengsang lexicon with 78% of the trisyllabic items in our survey overall. It was also the major class in X-Light-Light (84%) and X-Heavy-Light (79%) syllable structures. The only sector of the lexicon where LHL does not predominate is in X-Light-Heavy structures where LHH (80%) overshadows LHL (15%). There is thus a syllable weight effect in trisyllables. Our novel word experiment was designed to detect this syllable weight effect. We also varied the laryngeal category of the onset of the penult between lax/sonorant vs. aspirated/tense. The results for trisyllabic novel words are shown in Table 44.

Table 44: Wug test results (trisyllable)

weight \ accent	HHL	HLL	LHL	LHH	Others	Totals
CV.CVC.CV	12 (3%)	6 (1%)	367 (85%)	26 (6%)	19 (4%)	430
CV.CV.CVC	5 (1%)	20 (5%)	201 (47%)	197 (46%)	7 (2%)	430
CV.CV.CV	34 (8%)	6 (1%)	351 (82%)	21 (5%)	18 (4%)	430
CV.C^{h}V.CV	8 (2%)	8 (2%)	210 (49%)	198 (46%)	6 (1%)	430
Totals	59 (3%)	40 (2%)	1,129 (66%)	442 (26%)	50 (3%)	1,720

The LHL accent predominates across all four syllable types (66%). It has just two significant competitors. First, when the penultimate syllable has an aspirated or tense onset, then the double accent LHH pattern is significantly boosted (46% vs. 5%). This result suggests that speakers are associating the aspirated/tense onsets with a double-H, even though this association is attested only in word-initial position in the lexicon. Second, when the final segment of the stem is a consonant and hence would be heavy in the isolation form (*naputok* ≈ *naputok-ɨl*), the expected LHH accent that would be associated with a final heavy syllable is also boosted (46% vs. 6%).

As in the case of disyllables, two separate mixed-effects logistic regressions using the lmer function in R were run to test these effects. A random intercept was set for items and subjects. The minority HHL and HLL responses were excluded and the test focused on the majority LHL and LHH accent choices. The first test was run over the data summarized in the first three rows of Table 44. Syllable weight in the penultimate and final syllables were the predictor variables and accent class (LHL, LHH) was the dependent variable. The baseline was set to CV.CV.CV structures with the LHL accent. The test indicates that changing the weight of the final syllable is strongly biased towards the LHH response, while changing the penultimate syllable to heavy did not deflect the accent choice relative to the baseline. This result mimics the lexical trend noted earlier (Table 35), where a heavy final was biased towards the LHH accent class.

Table 45: Results of a mixed effects logistic regression model (wug trisyllable, weight)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-2.8393	0.2357	-12.047	< 2e-16	***
Penult-Heavy	0.1699	0.3179	0.535	0.593	
Final-Heavy	2.8194	0.2632	10.714	< 2e-16	***

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

As for the effect of the laryngeal character of the penult onset, the baseline was set to aspiration with the LHL accent to test if subjects distinguished tense vs. aspirated onsets in their choice of accent class. The results show that a lax-sonorant onset is significantly biased against the double H accent contour, just as in the disyllables. Once again, there is no difference between tense vs. aspirated onsets.

Table 46: Results of a mixed effects logistic regression model (wug trisyllable, onset)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.05470	0.12715	-0.430	0.667	
Onset-Lax/Son	-2.82733	0.25555	-11.064	< 2e-16	***
Onset-Tns	-0.01691	0.24772	-0.068	0.946	

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

7.5 Comparison of the lexical trends and the wug test results

To what extent can we say that the accent patterns assigned to the novel words by our South Kyengsang subjects coincide with trends in the lexicon, and thus might be explained by saying that South Kyengsang speakers are aware of these trends and call on them to decide how to accent a novel word? Some of the experimental results are consistent with this hypothesis, but others are not. Let us review them, starting with syllable weight. First, when the final two syllables are light and contain a laryngeally unmarked onset, then there is a strong bias towards penultimate accent. This holds for both disyllabic (81%) and trisyllabic (82%) nonce words. These frequencies match well with the predominance of penultimate LHL accent in the lexical trisyllables (81%). But for lexical disyllables with a light-light syllable profile, just 47% have penultimate HL accent, as against 37% with LH. Second, a final heavy syllable significantly biases the word towards a final H in both the lexicon and in novel words. The effect was significant ($p < 0.001$) in both disyllables and trisyllables. However, its magnitude was much smaller in the lexical disyllables (β = 0.38) compared to trisyllables (β = 3.73) and the wug words (disyllable β = 3.56, trisyllable β = 2.81).

But there are also a couple of notable mismatches between the trends in the lexicon and the accents assigned to the nonce words. First, the nonce words evidenced a significant bias in favor of the double-H accent following a tense or aspirated consonant. This held for both disyllables and trisyllables (penultimate aspirated/tense ≈ LHH) and was robust both in terms of reliability ($p < 0.001$) and magnitude (disyllables β = –3.64, trisyllables = –2.82). We detected this effect at the word-initial onset of lexical trisyllables, but it was much weaker in reliability and magnitude (p = 0.02, β = 0.66). However, the penultimate aspirated/tense onset is negatively associated to the LHH class in the lexicon. In lexical disyllables there was no effect (p = 0.67), reflecting no significant difference from the baseline HL class. It thus appears that the association between the double H accent and a laryngeally marked onset is an effect emerging in the experiment rather than one based on matching the statistics of the lexicon. See Becker et al. (2012) for discussion of another example of emergent effects in a novel word experiment.

It is worth taking a slight detour here to point out another set of data showing the connection between the double-H accent and the aspirated-tense onset. It is furnished by a sound change that has affected certain lexical items in the Kyengsang dialect. Some words with an initial lax stop or /s/ have changed the initial consonant to tense in this dialect. This segmental change has produced variants for a number of words. The following list, taken from data cited in G-Y. Lee (1996), shows the tonal patterns assigned by the first author to these words. While many retain the original tone pattern, the majority change to HH when the initial consonant is tensed.

(12) Onset tensification and accent

kĕ	k*í-ká	'crab'	kuptá	k*úptá	'to bake'
kálc^{h}i	k*álc^{h}í	'hairtail'	tutə́rita	t*útírita	'to knock'
kwə́ŋi	k*éŋi	'scythe'	tə́ncita	t*ə́ncita	'to pass'
kacámi	k*acɛ́mi	'flatfish'	pjəŋári	p*íkári	'chick'
kompó	k*ompó	'person with skin disease'	pollák	p*óllákɨ	'chin'
keám	k*ékóm	'hazelnut'	caktɛ́ki	c*áktɛ́ki	'stick'
kúlpam	k*úlpam	'punch'	cúkci	c*úkcí	'shoulder blade'
kámt*á	k*ámt*á	'to shampoo'	siráki	s*írɛ́ki	'dried vegetable'
kosóháta	k*osíháta	'be savory'	sísuk	s*ísúk	'brother-in-law'

The motivation for this segmental-tonal association is not clear. It might be based at least in part on our South Kyengsang subjects' knowledge of the Standard Seoul Korean dialect. It is well known that at the beginning of an accentual phrase, the LH boundary tones characterizing the initial rise are replaced by HH when the phrase begins with a laryngeally marked onset (S-A. Jun 1993). It could be that our subjects treated the novel words as loanwords from the standard dialect and assigned an accent based on their presumed knowledge of the distribution of the phrasal tones in that dialect.

A final mismatch between the lexicon and the novel word results concerns the association between a heavy syllable and the double-high accent. In the lexicon a heavy penult is positively associated with the double accent (HH) in disyllables (β = 0.79, $p < 0.001$). But in the novel word experiment, this association was not nearly as reliable nor strong (β = 0.30, p = 0.0675). As we observed in section 5.5, the association between the heavy syllable and the double H can be traced back to a much earlier stage in Korean when final accent was predominant. Words with an initial heavy syllable may have attracted the word-final accent before the change doubling an initial H tone occurred and thus became inputs to the latter sound change. The heavy syllable – double H association lacks a natural phonetic basis and so it is perhaps not surprising that it was not applied to the novel words in our experiment.

In sum, the attraction of a H to a final heavy syllable and otherwise the penult shows up quite robustly in the novel word results but is attenuated in the lexical disyllables. Since disyllabic native stems are over twice as frequent as trisyllables in the lexicon, the overall proportions of the penult vs. final accents do not align especially well between the lexicon and the nonce words. In addition, our subjects failed to show the lexically attested association between a heavy syllable and the double-H accent on the one hand, but on the other hand, they did manifest an association of the double-H accent with a laryngeally marked onset that has only a weak lexical precedent. We return to this general issue in section 10 where we model the accent distributions in terms of familiar metrical constraints, which one assumes provide a more meaningful comparison of the experimental – lexical correspondences in comparison to the more descriptive and atomized predictors used in our statistical tests: HL, LH, HH accent class, penult heavy, final heavy, etc.

8. North Kyengsang

In this section we briefly compare our results on the distribution of the South Kyengsang tonal classes in the lexicon and novel word experiment with the findings of H-J. Kim (2012) on the North Kyengsang dialect, centered in and around the city of Taykwu (Taegu). Kim's study was undertaken parallel to and independent of our investigation. Both employ the same basic methodology and reach similar results.

In our review of Kim's study, we note two principal differences between the accent in the South and North Kyengsang dialects that bear on the results. First, the MK Rise class is reflected in a long vowel and a double-H tone in North Kyengsang.

Table 47: Reflex of MK Rise class in South and North Kyengsang

MK	South Kyengsang	North Kyengsang	
nŭn	nŭn, nun-í	nú:n, nú:n-í	'snow'
sărʌm	sarám, sarám-í	sá:rám, sá:rám-i	'person'

Second, in South Kyengsang a high tone spreads leftward up to the peninitial syllable, while North Kyengsang lacks this phonetic process. As a result, no South Kyengsang word/phrase begins with two successive low tones, while in North Kyengsang the F0 rise from the initial syllable to a downstream H peak is more gradual.

Table 48: Existence/absence of leftward spreading in South and North Kyengsang

South Kyengsang	North Kyengsang	
satárí	satarí	'ladder'
k^hɛllíp^{h}ónía	k^hɛllip^{h}onía	'California'

H-J. Kim (2012) based her study on a corpus of c. 1,200 common words taken from a frequency dictionary. When adjustments are made for the different reflexes of the MK Rise, the frequency distribution of the HH, HL, and LH tonal patterns for disyllables and HHL, LHL, and LLH for trisyllables align well between the lexicons of North and South Kyengsang dialects. A logistic regression model using the glm function run on the lexical data tabulated in H-J. Kim (2012) revealed a strong bias for the HH accent for di- and tri-syllables with an initial heavy syllable and LH/LLH for trisyllables with a final heavy syllable. Kim also conducted a novel word study to explore the connection between accent and syllable weight. Similar to our South Kyengsang subjects, a final heavy syllable biased Kim's subjects to the LH accent response for disyllables and to LLH for trisyllables. One striking difference between the two dialects concerns the behavior of initial CVC syllables in di- and tri-syllabic stems. North Kyengsang subjects showed a bias towards the HH accent (32%) and the HL accent (66%) for such novel words, while South Kyengsang subjects showed a relatively lower response for HH (21%), along with HL (61%) and LH(L) (13%). We believe that this difference arises from the fact that North Kyengsang words with an initial long vowel (the reflex of the MK Rise) are regularly associated with the HH accent (section 3). Kim's North Kyengsang subjects appear to have generalized this connection to CVC syllables. South Kyengsang lacks this association (long vowels ≈ the HH class) and hence the weight effect of the initial syllable did not emerge clearly.

H-J. Kim (2012) also tested the effect of aspiration in a separate novel word experiment. Stimuli varying the location of an aspirated onset for trisyllabic stems were constructed: $C^hV.CV.CV$, $C^hV.C^hV.CV$, $CV.CV.C^hV$, CV.CV.CV. Relative to the baseline CV.CV.CV, which took penultimate LHL in 76% of the responses vs. HHL 20%, the proportions reversed in $C^hV.CV.CV$ with LHL at 46%, while HHL had 52%. Parallel to our South Kyengsang results discussed in section 7, Kim found the effect of aspiration on the accent to be an emergent one not motivated by the statistics of the North Kyengsang lexicon.

9. Loanwords

In this section, we briefly summarize the other common source of evidence for the extension of the phonological grammar to new lexical items: loanwords. N-J. Kim (1997) and Kenstowicz and Sohn (2001) observe that in assigning an accent to an English loanword in North Kyengsang, the place of accent in the source word is ignored and an accent is assigned based on the weight of the initial syllable, as well as in a disyllabic window at the right edge of the word. In a study of over 2,000 English loans, D-M. Lee (2009) finds that a similar system operates in the South Kyengsang dialect (H-K. Jun 2006 reaches essentially identical results based on a corpus of c. 600 loanwords). In Table 49, we show the predominant accentuation assigned to disyllabic and trisyllabic English loanwords found in D-M. Lee's (2009) study.

Table 49: English loanwords in South Kyengsang

Weight	Tonal contour	Rate	Example	Gloss
Heavy-Heavy	HH	80%	rə́ntə́n	'London'
Heavy-Light	HL	100%	nə́mpə	'number'
Light-Heavy	LH	91%	remón	'lemon'
Light-Light	HL	97%	múpi	'movie'
Heavy-Light-Light	HHL	96%	áksént^{h}ɨ	'accent'
Light-Heavy-Light	LHL	97%	mesíncə	'messenger'

Light-Light-Heavy	LHH	91%	rimúcín	'limousine'
Light-Light-Light	LHL	92%	tomíno	'domino'

It is clear that two factors determine the accent contour: a heavy syllable (CVC) must be high; if the final two syllables are light, then a (default) high is assigned to the penultimate syllable. In addition, a Light-Light-Heavy takes a final H which spreads back to the peninitial syllable. These simple principles account for over 90% of the data in all syllable profiles except for Heavy-Heavy.

10. Grammatical Modeling

To what extent is it legitimate to say that the accent patterns assigned to novel words and loanwords are based on the statistics of the native lexicon? Zuraw (2000, 2010) inaugurated this line of inquiry with the finding that variations in the rate of NC cluster reduction in Tagalog nonce words and loanwords mirrored the distribution of these factors in the lexicon. This result was explained by the hypothesis that the strength of the markedness constraints controlling the reduction was based on the statistical distribution of the competing data structures in the lexicon, but that faithfulness constraints to a lexically listed structure prevented any particular word from straying from its actual output. Since a novel word by definition lacks a lexical input to be faithful to, the normally hidden markedness constraints step in to assign a structure.

In this section, we summarize our exploration of this question for our South Kyengsang data using the Maxent Grammar Tool (Hayes 2009), which assigns weights to constraints based on the frequency of each output and the number of violations in the input training data. We assumed the familiar constraints governing accent distribution listed in (13). For simplicity, we did not postulate foot structure; the words with an onset /s/ and /h/ were excluded.

(13)

a. Weight-to-High (W-to-H): penalize a heavy syllable not associated to a high tone (cf. De Lacy 2002)
b. Nonfinality (NF): penalize a word-final syllable associated with a high tone (Prince and Smolensky 2004)
c. Align-High-Right (Align-R): assign a penalty for each low tone syllable that separates a high tone from the end of the word
d. *Lapse: penalize two successive low-tone syllables (Zoll 2003)
e. *Clash: penalize two successive syllables associated with a high tone (Zoll 2003)
f. *Asp-L: penalize a syllable with low tone whose onset is aspirated or tense

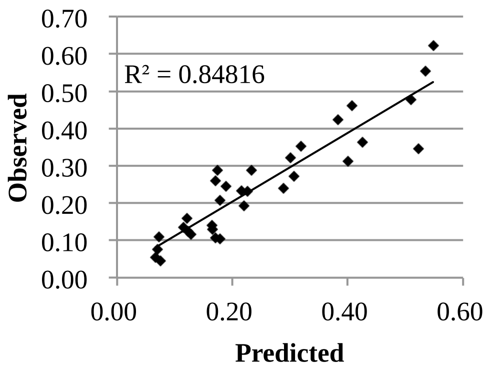
Table 50 shows the weights assigned to these constraints by the Maxent algorithm with its default mu and sigma settings based on the 5,569 disyllabic and trisyllabic words from Tables 19 and 27. In disyllabic words, *Clash and *Asp-L are weighted high, followed by NF, while W-to-H and Align-R receive the lowest weights and do not work to position prominence on the heavy syllable in the disyllabic lexicon. On the other hand, in trisyllabic words, W-to-H, *Lapse, *Clash are weighted high, and Align-R, *Asp-L and NF obtained relatively lower weights. For both disyllabic and trisyllabic models, the R^2 scores between observed and predicted frequencies are quite high (0.8482 and 0.9422, respectively), suggesting that the models explain the existing distribution well. For the trisyllabic model, Spearman's rank correlation rho also showed a significant result: S = 3945.254, p = 3.68e-11.

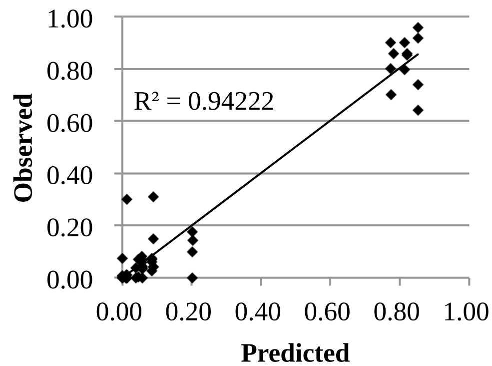
Table 50: Obtained weights (lexicon). Left = disyllable, right = trisyllable.

*Clash	0.91
*Asp-L	0.81
NF	0.29
W-to-H	0.10
Align-R	0.00

W-to-H	3.60
*Lapse	2.87
*Clash	2.68
Align-R	0.43
*Asp-L	0.19
NF	0.00

Figure 2: Correlation between predicted and observed distribution (lexicon). Left = disyllable, right = trisyllable.



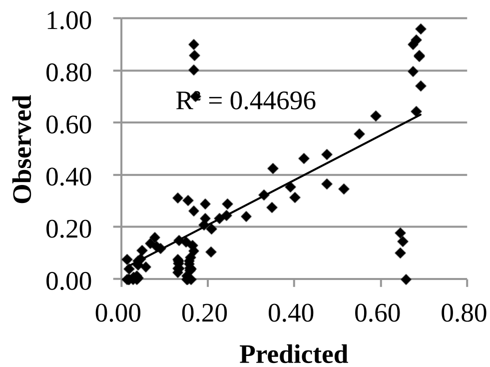


When we combine disyllabic and trisyllabic words and run the same simulation, the weight hierarchy changed dramatically. Now W-to-H is weighted lowest, and the relative weights of Align-R and NF are higher than in the weights obtained from only disyllables. More importantly, the R^2 declined significantly (0.447), resulting in many cases that do not reflect the actual distribution accurately.

Table 51: Obtained weights (lexicon, disyllable and trisyllable aggregated)

*Lapse	2.87
*Clash	1.46
*Asp-L	0.78
NF	0.61
Align-R	0.42
W-to-H	0.30

Figure 3: Correlation between predicted and observed distribution (lexicon, disyllable and trisyllable aggregated)



This suggests that the grammars of disyllabic and trisyllabic lexicons are different, a point that becomes clear when we compare the distribution of the final HL vs. LH/HH accents tabulated in Table 52.

Table 52: Weight and accent (disyllabic and trisyllabic lexicon)

	Disyllable		Trisyllable	
Weight	HL	LH	XHL	XHH
Light-Heavy	42%	45%	15%	80%
Light-Light	47%	37%	84%	6%

In trisyllabic words, there is a robust default (the LHL class) that attracts many words. However, when the final syllable is Heavy, most words appear with the LHH class, leading to the essentially binary distinction (LHL vs. LHH) based on the final syllable's weight. On the other hand, in disyllabic words, the syllable weight effect is not nearly as strong as in trisyllabic words, as shown by the accentual distribution in Light-Light structures (HL = 47%, LH = 37%). Thus, combining these two cases resulted in the miscalculation of the relative importance of the W-to-H constraint and reinforces our decision in sections 4, 5, and 6 to treat the accent of one, two, and three syllable stems separately.

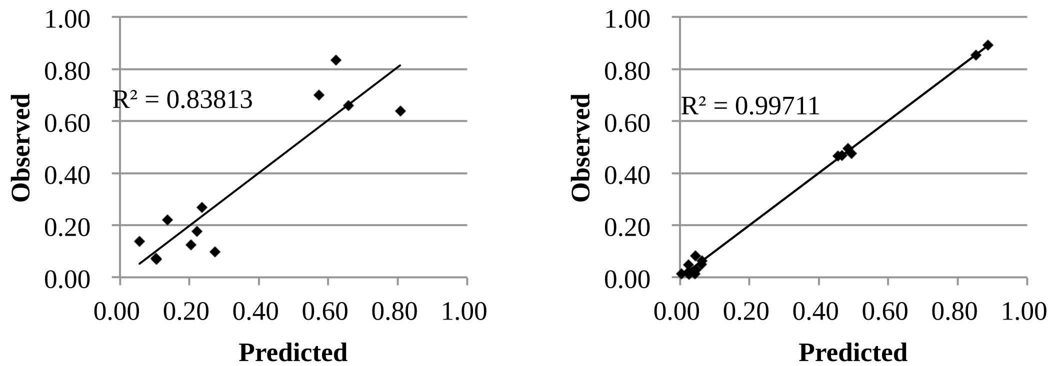
Table 53 shows the results of the simulation run for the wug test items. In these cases, we assumed that *Asp-L bans a low tone in both the penultimate and final syllables, penalizing both LH and HL accents. In both cases, the model predictions fit the observed distribution well. For the trisyllabic model, Spearman's rank correlation rho also showed a significant result: S = 75.5554, p = 4.16e-06.

Table 53: Obtained weights (wug). Left = disyllabic, right = trisyllabic.

*Asp-L	2.79
W-to-H	1.85
*Clash	0.95
NF	0.82
Align-R	0.00

*Lapse	2.99
*Clash	2.95
*Asp-L	2.59
W-to-H	2.55
Align-R	0.32
NF	0.00

Figure 4: Correlation between predicted and observed distribution (wug). Left = disyllable, right = trisyllable.



Moreover, in the case of wug test items, aggregating the results of disyllabic and trisyllabic words does not significantly lower the model's predictability (Table 54 and Figure 5). This suggests that in wug test items, the same grammar is utilized for both disyllabic and trisyllabic words, which is a different state of affairs compared to the lexicon.

Table 54: Obtained weights (wug, disyllable and trisyllable aggregated)

*Asp-L	3.00
*Lapse	2.86
W-to-H	2.21
*Clash	1.79
NF	0.64
Align-R	0.00

Figure 5: Correlation between predicted and observed distribution (wug, disyllable and trisyllable aggregated)

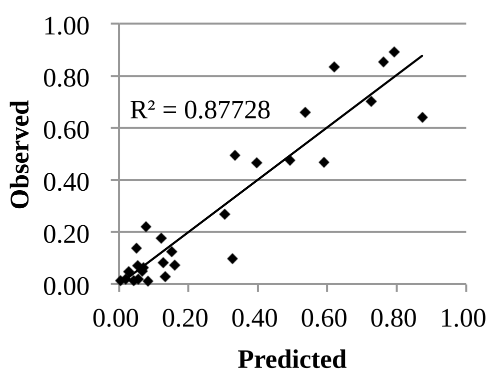
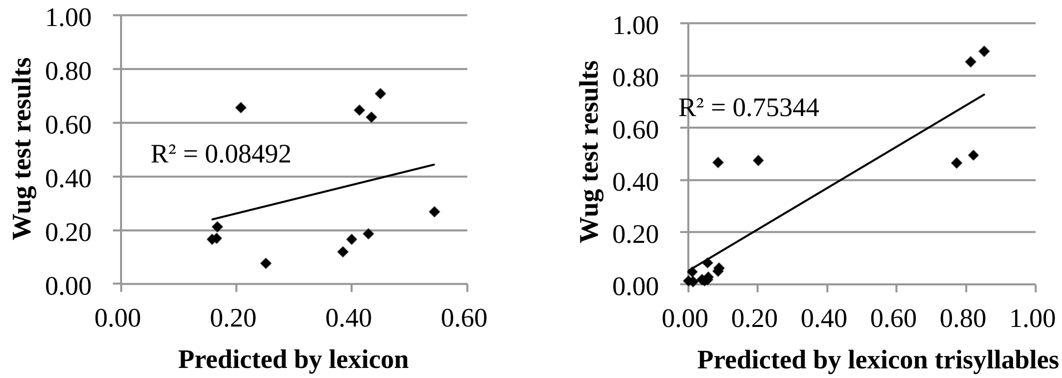


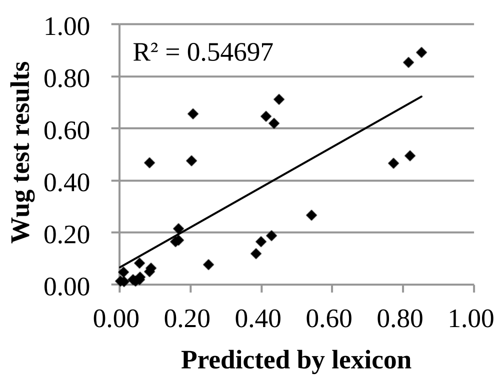
Figure 6 compares predicted values based on the lexicon with wug test results for both disyllabic and trisyllabic words. For disyllables, LH(L) and LH(H) are aggregated. The predicted distribution based on the disyllabic lexicon does not correlate with the wug test results at all (R^2 = 0.0849), whereas the predicted distribution based on the trisyllabic lexicon fits relatively well. For the trisyllabic model, Spearman's rank correlation rho also showed a significant result: S = 84.562, p = 8.822e-06.

Figure 6: Comparison between predicted values based on the lexicon and wug test results. Left = disyllable, right = trisyllable.



Finally, as seen in Figure 7, when the disyllabic and trisyllabic lexicons are aggregated, the fit with wug test results deteriorates considerably (R^2 = 0.547).

Figure 7: Comparison between predicted values based on the lexicon and wug test results (disyllables and trisyllables aggregated)



One possible interpretation of these results runs as follows. For the trisyllables, accenting a final heavy and otherwise the penult accounts for over 80% of the data and thus is a relatively reliable grammar. For this sector of the lexicon, the accent for the majority of words can be generated with a relatively small number of exceptions (many of which have variants that follow the regular cases). The latter will be lexically listed as exceptions. On the other hand, for disyllables no single accent pattern stands out as dominant, at least when compared to trisyllables, although there are many tendencies with high statistical regularity. Thus, for Kyengsang disyllables the accent is contrastive and memorized, even distinguishing a few minimal pairs or triples (there are no minimal pairs distinguished by accent in the trisyllables). In other words, on this proposal there is a default grammar (Table 50, right) that applies productively to trisyllables (longer words) and it is this grammar that the Kyengsang speaker consults when accenting a novel word.

In addition to word length there are two other factors that distinguish lexical trisyllables from disyllables that can help explain why they are the origin of the default accent. First, they are on the whole less frequent. We calculated the token frequency of items in our corpus that have a Middle Korean cognate (the only items for which we have frequency data). The disyllables have a mean token frequency of 68 in the King Sejong corpus while trisyllables are 56. Trisyllabic nouns are thus less firmly in the grasp of lexical faithfulness and conversely more susceptible to the effects of markedness instantiated in the Weight to High, Nonfinality, and Lapse constraints that define the default accent.

Another factor is that c. 80% of native trisyllabic nouns end in the vowel -i, many of which correspond to an accented suffix (diminutive) in Middle Korean. Although the vast majority of such words can no longer be parsed out with an independently occurring stem and the -i does not have a consistent meaning like the -*berry* morphemes in English *raspberry*, *strawberry*, etc., this word class nevertheless has a uniform accentual structure with an identifiable cue: of the 1,190 such words collected from our five South Kyengsang consultants, 1,041 (87%) have penultimate accent.

We might also conjecture that this default grammar has been hardened into a convention for accenting loanwords, where it applies with close to 100% accuracy. In the novel word experiment, the effect of the default accent is depressed somewhat due to noise and other competing factors like the laryngeal character of the onset, a factor that has only a marginal effect in loans.[18] Recall that we designed the novel word stimuli to sound like native words. Loanwords, especially on first exposure, are distinctly foreign and so may not trigger such segmental effects. Nevertheless, a loanword must be assigned an accent and so the default grammar productively assigning accent in the lexical trisyllables could step in to assign one to the loan. In effect, this explanation coincides with that suggested by Kenstowicz and Sohn (2001), who also note that in longer words the distinctiveness of Kyengsang accent is diminished, since most of the contrastive burden is distributed among the segmental and syllable structure profiles.

So at this point, we cannot conclude that the accentual preferences that emerged in our novel word experiment as well as the default accent found in western loanwords derive from matching the statistics of the overall lexicon, as claimed by H-J. Kim (2012) for the North Kyengsang dialect. Rather, the statistical effects that would be encoded in the submerged markedness constraints appear to be overshadowed by the default grammar that assigns accent to the longer words.

We close the discussion with a challenge to the hypothesis that the Kyengsang loanword accentuation derives from the native default grammar. In an analysis of the Yanbian dialect of Korean (spoken in northeastern China), Ito (2014a) suggests an entirely different approach to the loanword accent. Yanbian more faithfully reflects the accent of MK, since it did not undergo the sound change retracting the H from the final syllable: cf. MK, Yanbian *mərí* vs. Kyengsang *mə́ri* ‘head’. As a result, the majority of native words have final accent, reflecting the bias in MK: final LH accent 76% vs. penult HL 24%. Nevertheless, in Yanbian Korean English loanwords still show a preference for penultimate accent, similar to what is found in Kyengsang: final LH 20% vs. HL penult 80%. Ito (2014a) attributes this finding to the hypothesis that after Yanbian speakers were exposed to a sufficient number of English loans, they internalized and generalized the quantity-sensitive trochaic accent pattern at the right edge of the word that characterizes most English words. This principle has now become a conventionalized (phonologized) rule of loanword adaptation in Yanbian. On this view, the loanword accent reflects regularities of the English data source rather than the native Korean lexicon. See Shinohara et al. (2011), Kang (2013), and de Jong and Cho (2012) for discussion of other

[18] H-K. Jun (2006) observes some loans with a light-heavy syllable profile that take HH in place of the expected LH; all have an initial aspirated consonant: t^{h}ót^{h}ál ‘total’, c^{h}ɛ́nə́l ‘channel’ vs. t^{h}érə ‘terror’, k^{h}ónə ‘corner’.

examples of phonologization in Korean loanwords. Exploring this alternative hypothesis with respect to the loanword accent found in North and South Kyengsang is a task for future research.

11. Summary and Conclusions

In this paper, we have documented and analyzed changes in the affiliation of c. 1,900 noun stems with respect to the nominal accent classes in the South Kyengsang dialect of Korean. Our major results are summarized as follows.

The regularity of the accentual correspondences between Middle Korean and contemporary South Kyengsang mirror the size of the accent classes: larger classes are more regular. Deviations from the regular development by and large track the size of the classes: larger classes attract items from smaller classes. Besides this well established type-frequency effect, there is also a token-frequency effect: words of higher token frequency are less likely to change their accent. We also documented several segmental and syllable-weight factors that statistically bias a word for or against a particular tonal profile. The monosyllabic Rise class prefers sonorant codas, aspirated or tense syllable onsets disprefer low tone on the following vowel, and a final heavy syllable tends to attract the pitch peak. We then presented the results of an experiment with 43 South Kyengsang speakers designed to test whether these factors play a role in assigning accent to a novel word. Both the syllable weight and onset laryngeal properties biased the responses of the subjects in the expected direction. But the experiment also revealed two findings that did not accord with the statistics of the lexicon. First, a penult heavy syllable did not bias a word towards the double-H (HH) accent (contradicting a lexical trend) and second, a laryngeally marked onset in the penultimate syllable boosted the LHH response, even though there is no significant lexical support for this segmental-accentual connection, at least in trisyllables. The former finding supports the strong UG stance taken by Becker et al. (2011), with the default HL accent taken into account. We then reviewed the results of H-J. Kim (2012) who found experimental support for the syllable weight and aspiration effects in a novel word experiment with North Kyengsang subjects; and the findings of H-K. Jun (2006) and D-M. Lee (2009) that heavy syllables in western loanwords are regularly assigned a high tone and, in the absence of a heavy syllable, a high appears on the penult. H-J. Kim (2012) attributed both the novel word results and the loanword accent to the action of UG constraints trained on the statistics of the lexicon, employing the USE LISTED model of Zuraw (2000, 2010). In the final section of the paper, we questioned the suitability of this model for our South Kyengsang data. Modeling with the Maxent Grammar Tool revealed that the statistics of the disyllabic lexicon gave a rather weak fit with the grammar that emerges from the novel word experiment. A much better match was obtained when the Maxent Grammar Tool was run over the much smaller class of lexical trisyllables. We attributed this result to the fact that in the lexical trisyllables, accent is much less contrastive: 83% of the data can be covered by principles that place a high over a heavy syllable in a disyllabic window at the right edge of the word and otherwise on the penult. This is precisely the grammar that applies with over 90% accuracy in D-M. Lee's (2009) loanword data. This result in turn suggests an accentual grammar for South Kyengsang, in which the accent is productively assigned by the rules/constraints operating over the trisyllables with a relatively small number of exceptions, while the accent is contrastive and lexically listed for disyllables. Thus, it is the default rules that apply productively in the trisyllables that emerge in the wug test and possibly the loanwords as well. More precise matching with the statistics of the lexicon does not seem to play a role. Yang (2005) reaches a similar conclusion concerning the much discussed past tense morphology of English verbs.

There are a couple of more general issues raised by our study. First, we may wonder why the inflectional accent is relatively unstable in South Kyengsang. We recall from Table 8 that the overall agreement rate among our five consultants was 64%. While we lack any precise figures, we have the impression that the alignment of lexical items with respect to other phonological contrasts such as the three-way lax-tense-aspirated laryngeal contrast in stops is more stable both in its evolution from Middle Korean and its use in contemporary speech. A speaker's confidence in the phonological properties of a particular lexical item is undoubtedly influenced by frequency of exposure and usage. A significant portion of our five SK speaker's experience with Korean vocabulary is via the standard language, which lacks the accentual contrasts but retains the laryngeal contrasts. Also, for educated speakers their experience with particular words often takes place in a written context. Since accent is not marked orthographically, speakers are deprived of the reinforcement of contrasts provided by the writing system for laryngeal contrasts, which are marked orthographically. Indeed, the words for which our speakers agree occupy the high frequency group while those for which there is disagreement are drawn primarily from the low frequency group. We may conjecture that with the passage of time more and more Kyengsang tonal contrasts will be lost and words will consolidate into the default accent type.

Another finding of significance was that in the native vocabulary the LH(L) class outnumbered the LH(H) class, resulting in the drift of words from the latter class to the former and the almost complete absence of the latter in the novel word experiment. This behavior contrasts sharply with that noted by Ito (2014b) for the Sino-Korean vocabulary, where the LH(H) class predominates over LH(L). This suggests that the vocabulary is stratified with different and sometimes contradictory accentuation. If so, then what are the cues that the learner relies on in order to classify a particular word with respect to the lexical strata? Can this behavior be tested with novel word experiments? These are tasks for future research.

Acknowledgments

Portions of this paper were presented at the Linguistic Society of Japan and the University of Illinois as well as the MIT Phonology Circle and the USC Linguistics Colloquium. We thank the audiences for their comments. We also thank our four Lingua reviewers for feedback and criticisms that improved our presentation and understanding of the data. We are grateful to Adam Albright and Edward Flemming for advice about statistics and acknowledge the kind assistance of our Korean language consultants.

References

Albright, Adam. 2002. Islands of reliability for regular morphology: evidence from Italian. *Language* 78, 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.

Albright, Adam and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: a computational study. *Cognition* 90, 119–61.

Bates, Douglas and Martin Maechler. 2013. Lme4: linear mixed effects models using S4 classes. R package version 0.999999-2. Online: http://cran.r-project.org/web/packages/lme4/index.html.

Becker, Michael, Nihan Kertez, and Andrew Nevins. 2011. The surfeit of the stimulus: analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87, 84–125.

Becker, Michael, Andrew Nevins, and Jonathan Levine. 2012. Asymmetries in generalizing alternations to and from initial syllables. *Language* 88, 231–68.

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

Ceng, Yen-chan. 1971. Cwungsey sengco.wa kyengsangto pangen sengco.uy pikyo [A comparative study on the tonemes of Middle Korean and that of the Kyengsang dialects]. *Hankul hakhoy 50 tol kinyem nonmuncip* [Papers commemorating the 50th anniversary of the Hangul Association].

Ceng, Yen-chan. 1972. Cwungsey kwuke sengco.uy pyentong.kwa kiponhyeng [Tonal perturbation and its basic pattern in Middle Korean]. *Hankul* 150, 19–116. Reprinted in Yen-chan Ceng (1976), *Kwuke sengco.ey kwanhan yenkwu* [Studies on the Tone in Korean], 32-79, Seoul: Il cho kak.

Cho, Thaehon, Peter Ladefoged, and Sun-Ah Jun. 2002. Acoustic and aerodynamic properties of Korean stops and fricatives. *Journal of Phonetics* 30, 193–228.

Co, Ceng-hwan. 2000. Kankokugo shōnei hōgen-no akusento taikei: taigen-o chūshin-ni [Accent system of Changnyeng dialect of Korean: focus on nouns]. In Rei Fukui (ed.), *Kankokugo akusento ronsō* [Studies in Korean Accentology], 21–49, Tokyo: ICHEL, The University of Tokyo.

Davis, Stuart, Natsuko Tsujimura, and Jung-yueh Tu. 2012. Toward a taxonomy of loanword prosody. *Catalan Journal of Linguistics* 11, 13–39.

de Jong, Kenneth and Mi-Hui. Cho. 2012. Loanword phonology and perceptual mapping: comparing two corpora of Korean contact with English. *Language* 88, 341–68.

De Lacy, Paul. 2002. The interaction of tone and stress in Optimality Theory. *Phonology* 19, 1-32.

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

Fukui, Rei. 1985. Chūki chōsengo 19-no akusento taikei-ni tsuite. [On Middle Korean accentual system]. *Tokyo University Linguistic Papers '85*, 61–71.

Fukui, Rei. 2013. *Kankokugo oninshi-no tankyū*. [Explorations in Korean historical phonology]. Tokyo: Sanseidō.

Gleason, Jean Berko. 1958. The child's learning of English morphology. *Word* 14, 150–77.

Hayata, Teruhiro. 1974. Accent in Korean: synchronic and diachronic studies. *Gengo Kenkyu* 66, 73–116. Reprinted as Japanese translation in Hayata (1999), 127–170.

Hayata, Teruhiro. 1999. *Onchō-no taiporojī* [Typology of Tone]. Tokyo: Taishukan.

Hayes, Bruce. 2009. Maxent Grammar Tool. http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/.

Hayes, Bruce, Siptár, Péter, Zuraw, Kie, and Londe, Zsuzsa. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85, 822–63.

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

Hyman, Larry. 2002. Is there a right-to-left bias in vowel harmony? Paper presented at the 9th International Phonology Meeting, Vienna.

Ito, Chiyuki. 2013. Korean accent: internal reconstruction and historical development. *Korean Linguistics* 15:2, 129–198.

Ito, Chiyuki. 2014a. Loanword accentuation in Yanbian Korean: a weighted constraint analysis. *Natural Language and Linguistic Theory* 32, 537–592.

Ito, Chiyuki. 2014b. The accent of Sino-Korean words in South Kyengsang Korean. *Gengo Kenkyu* 145, 61–96.

Jun, Ho Kyung. 2006. Factors affecting accentual patterns of loanwords in Pusan Korean. *Harvard Studies in Korean Linguistics* 11, 158–170.

Jun, Jongho. 2010. Stem-final obstruent variation in Korean. *Journal of East Asian Linguistics* 19, 137–179.

Jun, Sun-Ah. 1993. The Phonetics and Phonology of Korean Prosody. Ph.D. dissertation, Ohio State University.

Jurawsky, Daniel and James Martin. 2000. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Englewood Cliffs, New Jersey: Prentice-Hall.

Kadowaki, Seiichi. 1976. Chs, New Jerse-ni okeru seich Chs, N-ni tsuite [Tonal alternation in Middle Korean]. *Chōsen Gakuhō* 79, 17–54.

Kadowaki, Seiichi, Isamu Matsuo, Yoshirō Takashima and Yukitoshi Yutani. 1993. *Chōsengo Jiten* [Korean-Japanese Dictionary]. Tokyo: Shōgakukan.

Kang, Yoonjung. 2003. Sound changes affecting noun-final coronal obstruents in Korean. In W. McClure (ed.), *Japanese/Korean Linguistics* 12, 128–139. Stanford: CSLI.

Kang, Yoonjung. 2013. L1 phonotactic restrictions and perceptual adaptation: English affricates in Contemporary Korean. *Journal of East Asian Linguistics* 22, 39–63.

Kenstowicz, Michael and Hyangsook Sohn. 2001. Accentual adaptation in North Kyungsang Korean. In Michael Kenstowicz (ed.), *Ken Hale: A Life in Language*, 239–70. Cambridge: 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, 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, Hyun-ju. 2012. Syllable Structure, Frequency, Analogy, and Phonetics: Factors in North Kyengsang Korean Accentuation of Novel Words. Ph.D. dissertation, Stony Brook University.

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

Kim, Wan-cin. 1973. *Cwungsey kwuke sengco.ui yenkwu* [A study on Middle Korean tones]. Seoul: Tower Press.

Kubozono, Haruo. 2006. Where does loanword prosody come from? A case study of Japanese loanword accent. *Lingua* 116, 1140-1170.

Kwuklip kwuke yenkwuwen [National Institute of the Korean Language]. (ed.) 1999. *Phyocwun Kwuke Taysacen* [Standard Korean Dictionary]. Seoul: Tusangtong'a.

Lee, Dongmyung. 2006. Weight-sensitive tone patterns in loanwords of South Kyungsang Korean. *Harvard Studies in Korean Linguistics* 11, 197–210.

Lee, Dongmyung. 2009. The Loanword Tonology of South Kyungsang Korean. Ph.D. dissertation, Indiana University.

Lee, Dongmyung and Stuart Davis. 2009. On the pitch accent system of South Kyungsang Korean: a phonological perspective. *Language Research* 45, 3–22.

Lee, Dongmyung and Stuart Davis. 2010. Evidence for accentless words in South Kyungsang Korean. *Studies in Phonetics, Phonology and Morphology* 16, 245-258.

Lee, Geun-Yeol. 1996. Kyengnam pangen.ui malmeli toinsolitoiki [Word-initial tensification in South Kyengsang dialect]. *Kwuke Kwukmunhak* 33, 71–88.

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

Lee, Sang Oak. 1978. Middle Korean Tonology. Ph.D. dissertation, University of Illinois, Urbana. Reprinted in 2007, Seoul: Hankook Publishing Co.

Martin, Samuel. 1987. *Korean-English Dictionary*. Yale University Press.

Martin, Samuel. 1992. *A Reference Grammar of Korean*. Tokyo: Tuttle Publishing.

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

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

Prince, Alan and Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA: Blackwell Publishing.

R Development Core Team. 2011. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online: http://www.R-project.org.

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

Ramsey, Robert S. 1991. Proto-Korean and the origin of Korean accent. In William Boltz and Michael Shapiro (eds.), *Studies in the Historical Phonology of Asian Languages*, 215–238. Amsterdam: John Benjamins.

Shinohara, Shigeko, Seong-Rim Ji, Tomohiko Ooigawa, and Takahito Shinya. 2011. The limited role of perception in Korean loanword adaptation: the Korean three-way laryngeal categorization of Japanese, English, French, and Chinese plosives. *Lingua* 121, 1461–1484.

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.

Sondereggar, Morgan and Partha Niyogi. 2013. Variation and change in English noun/verb pair stress: data and dynamical systems models. In Alan Yu, (ed.), *Origins of Sound Change: Approaches to Phonologization*, 262–284. Oxford: Oxford University Press.

Stolke, Andreas. 2002. SRILM–an extensible language modeling toolkit. *Proceedings of International Conference on Spoken Language Processing* 2, 901–904.

Whitman, John. 1994. The accentuation of nominal stems in Proto-Korean. In Young-Key Kim-Renaud (ed.), *Theoretical Issues in Korean Linguistics*, 425–439, Stanford: CSLI publications.

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

Zhang, Jie. 2002. *The Effects of Duration and Sonority on Contour Tone Distribution–A Typological Survey and Formal Analysis*. New York: Routledge.

Zhang, Jie, Yuwen Lai, and Craig Sailor. 2011. Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. *Lingua* 121, 181–206.

Zoll, Cheryl. 2003. Optimal tone mapping. *Linguistic Inquiry* 34, 225–68.

Zuraw, Kie. 2000. Patterned Exceptions in Phonology. Ph.D. dissertation, UCLA.

Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28, 417–472.