

Stem-final obstruent variation in Korean

Jongho Jun

(Seoul National University)

Abstract

It has been observed in the literature that stem-final coronal obstruents of nouns in Korean are generally in variation with [s] in prevocalic position: for example, /pat^h-il/ [pat^hil] ~ [pasil] ‘field, accusative’. In addition, nouns with final noncoronal stops have variants ending in lenis stops: for example, /ip^h-e/ [ip^h-e] ~ [ipe] ‘leaf, locative’. Recent survey and experimental studies reveal that a wide set of coronal obstruents [s, c^h, t^h, c, t] may occur as variants. Moreover, there is an order of preference among them: in general, s >> c^h, t^h >> c, t.

This paper first shows that the observed relative preference among variants is matched by the distribution of lexical final obstruents in noun stems. Building on Albright’s Paradigm Learning Model, I provide a unified account for the occurrence of most variants and their relative preference by proposing stochastic rules deriving the paradigmatically-related forms of noun stems. In addition, a wug-test is carried out to investigate the productivity of such rules. Results suggest that these rules are productive, providing the evidence for their cognitive presence.

1. Introduction

The typical data patterns employed in traditional analysis within generative phonology are systematic with no variation. Even when exceptional or variant output forms are attested, they have often been ignored in formal analysis. However, the variable, gradient, patterns have recently attracted much attention with the understanding that the exceptions and variants occur in a systematic way in many cases. The results of some recent research on the variation data (Bybee 2001; Boersma and Hayes 2001; Zuraw 2000, 2002, 2007; Albright 2002 a,b, 2005, 2008; Albright and Hayes 2003; Pierrehumbert 2003, 2006; Hayes and Londe 2006) show that the variation pattern observed in the speakers’ behavior is matched by a statistical pattern in the lexicon, supporting the hypothesis that speakers may internalize the variable lexical pattern and use that knowledge in their behavior. Crucial evidence in favor of this hypothesis comes from the results of productivity tests such as loanword adaptation and wug-tests (Berko 1958). The present study provides additional supporting evidence for the hypothesis,

discussing stem-final obstruent variation patterns observed in a dialect of Korean, spoken in the Seoul-Gyeonggi area, and the results of an acceptability judgment survey with native speakers of the same dialect. Specifically, this study provides an analysis of the Korean data within the framework of Albright's (2002a,b, 2005, 2008) Paradigm Learning Model in which the acquisition of morphology is based on the distributional pattern of the learning data, not the conventional lexicon, consisting of inflected surface forms available to the learner at the time of learning. Accordingly, the results of this study have implications about how and what speakers do to internalize statistical lexical patterns.

It has been observed in the literature on Korean phonology and morphology that stem-final coronal obstruents of nouns are generally in variation with [s] in prevocalic position (Ko 1989; Martin 1992; Hayes 1998 and many others): for example, /pat^h-il/ [pat^hil] ~ [pasil] 'field, accusative'. Recent survey and experimental studies (S. Kim 2003; Choi 2004; E. Kang et al. 2004; Oh 2006; Oh & Shin 2007; Jun & Lee 2007) reveal that a wide set of obstruents [s, c^h, t^h, c, t] may occur as variants. Moreover, there is an order of usage frequency or preference among them: in general, s >> c^h, t^h >> c, t.

Following the general idea pursued by Ko (1989), Hayes (1998), Albright (2002a, 2005, 2008), and Y. Kang (2003a,b), this paper first shows that the observed relative frequency and preference among the innovative variants is matched by the distribution of lexical final obstruents in noun stems. Such frequency matching can be verified with a corpus, but not a dictionary. It will be shown that inflected, not bare, forms should be considered for the frequency counting.

To explain the observed occurrence of multiple variants and their frequency matching with the corpus counts, I provide multiple stochastic rules of the sort proposed by Albright (2008) and Albright & Hayes (2003) based on the type frequency of relevant inflected forms of nouns. In addition, a wug-test is carried out to investigate productivity of these rules. Results suggest that those rules are productive, providing the evidence for their cognitive presence. Notice that the lexical frequency matching and the occurrence of multiple variants cannot be directly incorporated into the previous approaches in which only a single rule or constraint, morphological or phonological, is posited mainly to explain the occurrence of the [s]-final variants, possibly along with an additional separate mechanism for the occurrence of a few other variants. I will argue that the proposed analysis exceeds all previous analyses on an explanatory level.

The remainder of this paper is organized as follows. In the next section, I first

discuss the general phonological processes of Korean that cause basic stem-final alternations in the noun paradigm. This is followed by a discussion of some recent studies on the variation involved in noun alternations. In section 2, I show that all the observed variation patterns can be understood given the frequency facts of Korean nouns. In section 3, I first introduce Albright's Paradigm Learning Model. Extending Albright's account of the Korean data, I then provide an analysis of the variation patterns discussed in section 2 by proposing multiple stochastic rules. In section 4, I describe a judgment survey employing a certain set of bound stems. The results show that topicalized forms of the bound stems are subject to the pattern of acceptability similar to the frequency and preference of stem-final obstruents of topicalized nouns. The implications of these results are discussed. Section 5 discusses previous alternative accounts. The final section concludes the present study.

1.1 Stem-final alternations in nouns

Korean has a three-way laryngeal contrast between lenis, aspirated and tense (or glottalized) obstruents, as shown below.

(1) Three-way laryngeal contrast among obstruents in Korean

	<i>Labial</i>	<i>Coronal</i>			<i>Velar</i>
	Stop	Stop	Fricative	Affricate	Stop
Lenis	p	t	s	c	k
Aspirated	p ^h	t ^h		c ^h	k ^h
Tense	p'	t'	s'	c'	k'

These obstruents may be targeted in various phonological changes, two of which are directly relevant to this paper. First, all obstruents neutralize to their homorganic lenis stop counterparts in coda position: coronal stops, affricates and fricatives neutralize to [t], labial stops to [p] and velar stops to [k]. This coda neutralization applies with no exception. In addition, coronal obstruents are palatalized before high front vocoids [i, j]. Specifically, alveolar stops [t, t^h] become palato-alveolar affricates [c, c^h], and alveolar fricatives [s, s'] become palato-alveolar counterparts [ʃ, ʃ'] in the palatalization context. The former palatalization process applies only when a morpheme boundary intervenes whereas the latter is an automatic process. Among the two palatalization processes, only the cross-morphemic palatalization will be reflected in the phonetic transcription of the data in this paper since the fricative palatalization is completely predictable and it is not

relevant to any important part of this paper.¹

These phonological processes cause alternations in the noun paradigm. As shown in (2), Korean noun stems may appear with a variety of final obstruents when they combine with certain vowel-initial suffixes such as a locative case marker *-e*. However, unmarked isolation forms may end only in (unreleased) lenis stops, due to coda neutralization. Also, before a nominative case marker *-i*, stem-final coronal obstruents are realized only as palato-alveolars such as [c, c^h], due to coronal palatalization. As a result, noun stems ending in obstruents show stem-final alternations within the paradigm. For instance, the final consonant of the noun stem meaning ‘field’ appears as [t^h] before the locative case marker *-e*, [t] in the isolation form, and [c^h] before the nominative case marker *-i*.

(2) Alternations of stem-final obstruents in nouns.

stem-final C	(i) locative (-e)	(ii) unmarked isolation form	(iii) nominative (-i)	gloss
<u>Coronal</u>	os-e pat ^h -e nac-e pic ^h -e	ot ^ʔ pat ^ʔ nat ^ʔ pit ^ʔ	os-i pac ^h -i nac-i pic ^h -i	‘clothes’ ‘field’ ‘day’ ‘light’
<u>Labial</u>	pap-e ip ^h -e	pap ^ʔ ip ^ʔ	pap-i ip ^h -i	‘rice’ ‘leaf’
<u>Dorsal</u>	kuk-e pak ^ʔ -e puək ^h -e	kuk ^ʔ pak ^ʔ puək ^ʔ	kuk-i pak ^ʔ -i puək ^h -i	‘soup’ ‘outside’ ‘kitchen’

In a standard rule-based analysis of these stem-final alternations, the output form occurring before vowel-initial suffixes (except for [i]-initial ones) is posited as the underlying form, (which is reflected in standard Korean orthography). The above mentioned phonological rules apply to the underlying form to produce unreleased stops and palatalized consonants in the isolation and nominative forms respectively. This standard rule-based analysis can be illustrated by the following derivation of some allomorphic forms of the stem /pat^h/.

¹ Another automatic process in Korean, not reflected in the phonetic transcription in the paper, is inter-sonorant voicing in which lenis stops become voiced between sonorants.

(3) Standard rule-based analysis of the alternations.

	<i>locative</i>	<i>isolation form</i>	<i>nominative</i>
Underlying form:	/pat ^h +e/	/pat ^h /	/pat ^h +i/
Coda neutralization		t	
Palatalization			c ^h
Surface form:	pat ^h e	pat	pac ^h i

1.2 Variation in nouns

It has been observed in the literature (Ko 1989; Martin 1992; Hayes 1998; Albright 2005, 2008; Han 2002; Y. Kang 2003a,b; J-H Kim 2005; Park 2006) that pre-vocalic allomorphs of the noun stems, such as [pat^h-e] and [pac^h-i] in (3), show variation in final obstruents. The variants differ depending on the place of articulation of the obstruents. Stem-final coronal obstruents are mostly in variation with [s]: for example, /pat^h-il/ [pat^hil] ~ [pasil] ‘field, accusative’. Frequent occurrence of the final variant [s] has been extended to the pronunciation of the inflected forms of English loanwords ending in coronal stops (Ko 1989; H-S Sohn 2001; Choi 2002; Y. Kang 2003a; Davis and Kang 2006; Jun & Lee 2007). Word-final /t, d/ in English are normally realized as [s] in Korean when they are combined with vowel-initial suffixes: for instance, [kus-in] ‘English loanword: good, topicalized’ (Ko 1989). Noncoronal aspirated/tense stops, /p^h, k^h, k’/, are in variation with their homorganic lenis counterparts, [p, k]: for example, /ip^h-e/ [ip^h-e] ~ [ipe] ‘leaf, locative’ and /puək^h-e/ [puək^h-e] ~ [puəke] ‘kitchen, locative’.

Some recent survey studies (S. Kim 2003; Choi 2004; and E. Kang et al. 2004) reveal that the variation patterns involved are much more complicated than previously thought. In all these studies, a sizeable pool of native Seoul(-Gyeonggi) Korean speakers were employed: 1,174 subjects in Choi, 156 subjects in E. Kang et al., and 350 subjects in S. Kim. Subjects were consulted about their actual and preferred pronunciations in the surveys by Choi and E. Kang et al. respectively. S. Kim’s study, which reports somewhat limited data on noun stem-final variations, is based on the phonetic transcriptions of the speech of subjects who read experimental sentences.

Most notably, the range of the coronal variants is quite wide, as can be seen in (4). Five coronal obstruents [s, c^h, t^h, c, t] are involved as variants. For instance, in (4bv), the locative forms of the stem /toc^h/ may have all the five coronal obstruents [s, c^h, t^h, c, t]

in the stem-final position. Notice that most other noun stems have variants besides the one ending in [s] and the “standard” phonetic output of the underlying form which is either identical to the underlying form or, before [i]-initial suffixes, the result of the cross-morphemic coronal palatalization. The occurrence of “non-standard” variants is confirmed in other studies including Oh (2006), Oh & Shin (2007) and Jun & Lee (2007).

(4) Noun variations: Stem-final obstruent = coronal (data selected from S. Kim (2003), H. Choi (2004), and E. Kang et al. (2004))

<u>Stem-final C</u>	<u>UR</u>	<u>Standard</u>	<u>Variants</u>	<u>gloss</u>	<u>source</u>
a. /t ^h /	(i) /pat ^h -i/	pac ^h -i	pas-i [paʃi]	‘field’	Choi
	(ii) /pat ^h -ilo/	pat ^h -ilo	pas-ilo ~ pac ^h -ilo	‘field’	Choi
	(iii) /sot ^h -e/	sot ^h -e	sos-e ~ soc ^h -e ~ sot-e	‘pot’	Kang
	(iv) /mit ^h -il/	mit ^h -il	mis-il ~ mic ^h -il ~ mit-il	‘bottom’	Kang
b. /c ^h /	(i) /k’oc ^h -ilo/	k’oc ^h -ilo	k’os-ilo ~ k’ot ^h -ilo	‘flower’	Choi
	(ii) /k’oc ^h -a/	k’oc ^h -a	k’os-a ~ k’ot ^h -a ~ k’oc-a ~ k’ot-a	‘flower’	Kim, Choi
	(iii) /nac ^h -ilo/	nac ^h -ilo	nas-ilo ~ nat ^h -ilo	‘face’	Choi
	(iv) /tac ^h -il/	tac ^h -il	tas-il ~ tat ^h -il ~ tac-il	‘anchor’	Choi
	(v) /toc ^h -e/	toc ^h -e	tos-e ~ tot ^h -e ~ toc-e ~ tot-e	‘sail’	Choi
	(vi) /pic ^h -il/	pic ^h -il	pis-il ~ pit ^h -il	‘light’	Choi
c. /c/	(i) /nac-ilo/	nac-ilo	nas-ilo ~ nac ^h -ilo	‘daytime’	Kang
	(ii) /pic-in/	pic-in	pis-in ~ pic ^h -in	‘debt’	Kang
	(iii) /møkɕɛc-e/	møkɕɛc-e	møkɕɛs-e ~ møkɕɛc ^h -e	‘uvula’	Kang

(Suffixes: -i = nominative, -il = accusative, -ilo = directive/instrumental, -in = topic, -e = locative/dative, -a = vocative)

Another interesting observation is that these coronal variants are used with different frequency and preference. In general, [s] is the most frequent stem-final variant. In contrast, [c] and [t] are rarely adopted as a variant. The intermediate variants, [c^h] and [t^h], show different frequency of occurrence (and different degrees of acceptability) depending on the quality of the initial vowel of the suffix following the noun stem, as discussed and analyzed mainly with etymologically /t^h/ and /c^h/-final nouns by Y. Kang (2005, 2007): more precisely, [c^h] and [t^h] are likely to occur before [i] and [e],

respectively. Let us consider observations indicating this suffix-initial vowel effect favoring [c^h-i] and [t^h-e] sequences.

It has been relatively well-known in the literature that [c^h] may be attested, in addition to [s], as a stem-final variant especially when etymologically /t^h/-final stems combine with [i]-initial suffixes: e.g., /pat^h-il/ [pac^hil] ‘field, accusative’ (Martin 1992). For this reason, most recent studies on the occurrence of [c^h]-final variants focus on the pre-[i] context. H. Choi’s (2004) survey results for the pronunciation of etymologically /t^h/-final nouns, summarized in (5), show that [c^h]-final forms (cells marked in gray) are adopted almost as frequently as [s]-final ones before [i]-initial suffixes. In contrast, under the assumption that [c^h] can never (or at best very rarely) be adopted as a stem-final variant before [e]-initial suffixes, its frequency was not even directly examined in the survey. For instance, only the following four options are given in the survey form for the pronunciation of /pjət^h-e/ ‘sunshine, locative’: (i) [pjət^he], (ii) [pjøse], (iii) don’t know, (iv) other options () (test words and response options are all written in Korean orthography, which is alphabetic). There was no main response option for the [c^h]-final form, [pjəc^h-e] here, although a consultant may still volunteer it in the space provided for “other options”. H. Choi (2004) in fact mentions the occurrence of such [c^h] responses, not reporting their exact frequency or proportions. The proportions of [s] and [t^h] responses in (5) indicate that [c^h] responses should form very small part of the relevant results, less than 5~6% (boxed cells).²

(5) Proportions of [s], [c^h] and [t^h] responses for etymologically /t^h/-final stems (from the survey results reported in H. Choi 2004)³

<i>final sounds</i>	__[i]		__[e]		__[i]
	-il	-ilo	-e	-esə	
[s]	28.53	21.78	21.65	19.63	27.1
[c ^h]	33.94	19.42	less than 5	less than 6	70.3
[t ^h]	36.36	57.96	74.35	74.45	

Frequent occurrence or preference of [c^h]-final forms before [i] have also been reported

² In fact, Y. Kang (2007: p. 5, #9) reports, citing Choi (2004), that the proportion of [c^h] responses under consideration is 3.35%.

³ A total of fifteen /t^h/-final nouns were used in the survey, but different noun stems may be combined with different suffixes. Thus, there are 5 stems marked with -il, 10 with -ilo, 10 with -e, 5 with -esə, and 12 nouns with -i. The proportions under each suffix in (5) are calculated over responses of nouns marked with the same suffix.

by E. Kang et al. (2004: 12) and Oh (2006: 80). Further, the occurrence of [c^h]-final variants before [i] is not limited to etymologically /t^h-final nouns. [c^h] occurs, and can be preferred, as a variant of the stem-final /c/ (e.g., /pamnac-ilo/ [pamnac^hilo] ‘night and day, directive’ (E. Kang et al 2004: 8); /nolimpic-il/ [norimpic^hil] ‘gambling debt, accusative’ (Oh & Shin 2007: 221)), and adopted as a pronunciation of the stem-final /t/ in English loanword adaptation ([mɛc^hil] ‘Matt, accusative’ (Jun & Lee 2007: 177)).

In contrast, [t^h]-final variants show a somewhat reverse pattern of occurrence. H. Choi (2004: 67) reports that some subjects adopted [t^h]-final forms as their actual pronunciation for etymologically /c^h-final stems, especially when they combine with locative suffixes, i.e., [e]-initial suffixes. As for etymologically /c^h-final nouns, Choi’s study is mainly concerned with the relative frequency between [s] vs. [c^h]-final forms probably under the assumption that [t^h]-final variants are never, or very rarely, possible for etymologically /c^h-final nouns. Thus, in her survey forms asking for the pronunciation of /c^h-final nouns with [e]- and [i]-initial suffixes, there was no option for [t^h]-final variants, and thus all [t^h] responses of these test words are from “other options”. As shown below, the frequency of [t^h] responses reported in Choi is higher with *-e*, compared to *-ilo*. This relative frequency difference is crucially true when we consider overlapping cases (marked in gray) in which the same stems are used in combination with both *-e* and *-ilo* suffixes.⁴

⁴ Choi additionally reports the occurrence of [t^h] responses for a noun /tac^h/ ‘anchor’ that is marked with *-il* and *-e* in the survey form. There was only one [t^h] response with *-il*, and four [t^h] responses with *-e*, thus confirming the suffix-initial vowel effect under consideration.

(6) Frequency of [t^h] responses of etymologically /c^h/-final stems (from the survey results reported in H. Choi 2004)

	-ilo	-e
/k'oc ^h /	27	
/nac ^h /	2	
/tac ^h /		4
/toc ^h /	2	5
/mjəc ^h /	26	53
/təc ^h /	4	19
/pic ^h /	3	
/salkac ^h /		55
/suc ^h /	19	46
/oc ^h /		12
/juc ^h /		27
total	83	221
sum of overlapping cases (gray)	51	123

It has also been reported in the literature (H. Choi (2004: 20) and E. Kang et al. (2004: 12)) that Korean speakers are likely to retain the stem-final [t^h] sound in the pronunciation of etymologically /t^h/-final nouns with [e]-initial suffixes, not adopting the dominant variant [s]. Thus, it seems that the observed frequent occurrence of [t^h]-final variants before [e]-initial suffixes may be understood as a case of Korean speakers' general tendency favoring [t^h-e] sequences when the stem-final obstruent is a coronal.

In summary, previous studies suggest that [c^h]-final variants are more frequent and preferred before [i]-initial suffixes, compared to [e]-initial ones. In contrast, [t^h]-final variants are more frequent and preferred before [e]-initial suffixes, compared to [i]-initial suffixes. Y. Kang (2003a,b, 2005, 2007) shows that there is a comparable suffix-initial vowel effect in Korean speakers' acceptability judgments of variant forms with final [c^h] and [t^h] although there is speaker-dependent and word-dependent variation. Y. Kang's (2003b) well-formedness judgment data show that the mean ratings of [c^h] pronunciation of eight Korean speakers for (eight) etymologically /t^h/-final nouns are higher with nouns marked with [i]-initial suffixes, compared to those with [e]-initial ones. In contrast, the mean ratings of [t^h] pronunciation for (five) etymologically /c^h/-final nouns show the reverse pattern, i.e., higher with nouns with [e]-initial suffixes and lower with those with [i]-initial ones, though the [t^h] pronunciation is generally rated lower than the [c^h] pronunciation.

Consequently, it seems that [c^h]-final variants are more frequent, preferred and

acceptable before [i] whereas [t^h]-final variants are more so before [e]. Some previous studies show an even finer distinction among [i]-initial suffixes: that is, the directive suffix *-ilo* is medial between other [i]-initial suffixes and [e]-initial ones.⁵ Y. Kang's well-formedness judgment data show that [t^h]-final variants are more acceptable before *-ilo* than they are before *-il* (and, as mentioned above, they are most acceptable before *-e*). In addition, H. Choi's (2004) survey data, discussed in Y. Kang (2007), show that [c^h]-final variants are adopted less frequently before *-ilo* than they are before *-il*, as can be indicated by comparing proportions of [c^h] responses before *-il* (33.94%) and before *-ilo* (19.42%) in (5). Unfortunately, the relatively higher acceptability of [t^h]-final variants before *-ilo* shown in Y. Kang is hard to verify in the survey studies like H. Choi (2004) where the use of *-ilo* and other [i]-initial suffixes in the survey form is not balanced.⁶ Also, the relatively lower frequency of [c^h]-final variants before *-ilo*, shown in Choi's survey results, is not confirmed in Kang's well-formedness ratings data in which three out of eight subjects rated [c^h]-final variants marked with *-ilo* lower than those with *-il*, but two subjects rated the other way around. Thus, although there is some indication about the medial pattern involving the directive suffix *-ilo*, its exact status needs further examination since the relevant results of the previous studies are either incomparable or incompatible with each other.

Finally, variation patterns of nouns with final noncoronal stops have also been confirmed in recent surveys. As shown below, it is reported that noncoronal aspirated/tense obstruents are in variation with their lenis counterparts when the stems are combined with vowel-initial suffixes.

(7) Stem-final obstruents = Noncoronal (S. Kim 2003; H. Choi 2004; E. Kang et al. 2004 and others)

<i>Stem-final C</i>	<i>UR (= standard)</i>	<i>Variants</i>	<i>gloss</i>
/p ^h /	ip ^h -e	ip-e	'leaf'
	sup ^h -i	sup-i	'woods'
/k ^h /	puək ^h -e	puək-e	'kitchen'
	sepjəkɲək ^h -e	sepjəkɲək-e	'dawn'

All the observations about the stem-final obstruent variation of nouns, discussed thus far, are summarized as follows:

⁵ Thanks to Yoonjung Kang for pointing this out to me.

⁶ In Choi's survey, nine etymologically /c^h/-final nouns are marked with *-ilo*, compared to only two with *-il*.

- (8) A summary of observations about stem-final obstruent variation in nouns
- a. Five coronal obstruents [s, t^h, c^h, c, t] are involved in variation of stem-final coronal obstruents.
(Tense consonants are never chosen as variants.)
 - b. [s] is, in general, the most frequent/preferred variant.
 - c. [c] and [t] are least frequent/preferred variants.
 - d. [c^h] and [t^h] are intermediate variants.
 - e. [c^h] is a frequent/preferred variant before [i]-initial suffixes.
 - f. [t^h] is a frequent/preferred variant before [e]-initial suffixes.
 - g. Noncoronal aspirated/tense obstruents are in variation with their lenis counterparts.

Before discussing how to explain these observations, two things need to be clarified with respect to the interpretation of the data in recent studies on Korean noun variation, discussed above. First, most previous works are mainly concerned with inter-speaker variation. In Choi's survey, subjects were asked about what sounds they actually *produce* in stem-final position of each target word whereas in E. Kang et al, subjects were asked about what sounds they *prefer* in the same condition. In both surveys, the subjects were instructed to choose only one sound for each target word. Other recent experimental studies such as Oh (2006), Oh & Shin (2007) and Jun & Lee (2007), which employ a task designed to elicit production of stem-final obstruents and report the occurrence of non-standard variants, are not different in this respect. Thus, the data in (4) show a list of speaker-dependent variants. Comparable intra-speaker variation probably exists, as mentioned by K-A Choi (2002), but I do not know its details, such as the relative frequency and range of the variants. But, it seems at least true that variation patterns summarized in (8) may also indicate an average speaker's relative preference or acceptability of variants, as suggested by well-formedness ratings data reported in Kang (2003a,b) which are in general consistent with the frequency data of speaker-dependent variants such as those in H. Choi (2004).

Second, most previous works have some limitations with respect to the scope of their investigations. For instance, as far as I know, no previous studies have explored the relative frequency or preference of all FIVE coronal variants with equal importance or the variation patterns of nouns ending in the FIVE lexical coronal obstruents. For instance, as mentioned above, as for etymologically /c^h-final nouns, H. Choi's (2004) study is mainly concerned with the relative frequency of [s] vs. [c^h]-final variants and

thus does not even adopt other potential variants as main response choices. Also, to my knowledge, no previous studies provide a systematic investigation of the variation of etymologically /s/-final nouns. This must be based on the assumption that /s/-final nouns are not subject to variation (e.g., E. Kang et al. 2004, p. 7). But, there are small indications that /s/-final nouns may also have variants.⁷ Some Korean speakers make spelling mistakes for etymologically /s/-final nouns by replacing the //s// letter with a letter for some other coronal obstruent: e.g., //kuc^hil// (for /kus-il/ ‘exorcism, accusative’), //kac^hil// (for /kas-il/ ‘traditional Korean top hat, accusative’) and //mæc^hil// (for /mæs-il/ ‘taste, accusative’) (May 2009 at <http://google.co.kr>). These wrong spellings might reflect the usual pronunciation at least of the bad spellers and probably that of some others who always spell correctly, ignoring their usual pronunciation. This kind of variant forms of etymologically /s/-final nouns can also be found among non-standard dialectal forms⁸ (e.g., [pic^hi] & [pic^hilo] /pis/ ‘comb, nominative & directive’ in some dialects spoken in South Kyung-sang and South Ceu-la provinces) and in one subject’s production of [oc^hil] for /os-il/ ‘clothes, accusative’ reported in Jun & Lee (2007).⁹ Thus, it seems that it is not totally impossible to have variants of etymologically /s/-final nouns. Consequently, given that previous studies adopt quite limited and imbalanced sets of test nouns as well as potential variant forms, there is a possibility that the variation patterns found in the previous studies, summarized in (8), may be true more generally. For instance, although the suffix-initial vowel effect, i.e., the preference of [c^h] before [i] and [t^h] before [e], is established mainly based on the variation pattern of the etymologically /c^h/ and /t^h/-final nouns, it is still possible that the effect may be true for nouns ending in coronal obstruents in general. It will be shown that the survey results of the present study can be better understood under this broad interpretation of the data reported in the previous studies. For its final verification, a wide-scale and systematic investigation, experimental or survey, is necessary.

Let us briefly consider how some previous analyses of the stem-final variation of Korean nouns can deal with the variation patterns summarized in (8). Among the observations and patterns summarized in (8), I think the occurrence of a wide set of variants is the most important since it requires at least a drastic revision of most previous analyses of Korean noun variation which propose a single rule or constraint deriving [s]-final variants. Some relevant rules and constraints are shown below:

⁷ Thanks to Stuart Davis (personal communication) and an anonymous reviewer for bringing this issue to my attention.

⁸ Relevant examples can be found in a series of *Korean Dialectal Data* (written in Korean) published, in 1987-1995, by The Academy of Korean Studies.

⁹ Only one out of ten subjects produced a [c^h]-final variant for etymologically /s/-final nouns.

(9) Rules or constraints deriving the stem-final variant [s] of nouns

- a. Intervocalic (optional) assibilation: /t, c^h, c/ → [s] / V__+V (H. Kim 2001; cf. K-A Choi 2002)
- b. Morphological rule: /t/ → [s] / __+V (noun paradigm) (trivially adapted from Ko 1989)
- c. Anticorrespondence constraint: t/___]noun# → s/___]nounV
 “If a noun ends in [t] in isolation form, change it to [s] before a vowel initial suffix.” (Y. Kang 2003b; cf. Davis & Kang 2006)

(9a) is proposed as a phonetically-natural rule which changes intervocalic coronal obstruents to [s]. (See K-A Choi 2002 for the proposal of a constraint comparable to the rule in (9a).) (9b) and (9c) are a morphological rule and a constraint, respectively, which require that prevocalic allomorphs of nouns should end in [s] if their corresponding isolation forms end in [t]. It is obvious that these rules and constraints say nothing about the occurrence of other coronal variants. More specifically, it is very difficult to explain the occurrence of those variants which cannot be derived through the application of the rules in (9) or the cross-morphemic palatalization. (10) lists such problematic forms chosen from (4).

(10) Problematic variants

<u>Stem-final C</u>	<u>UR</u>	<u>Variants</u>	<u>gloss</u>
a. /t ^h /	(i) /pat ^h -i/		‘field’
	(ii) /pat ^h -ilo/	pac ^h -ilo	‘field’
	(iii) /sot ^h -e/	soc ^h -e ~ sot-e	‘pot’
	(iv) /mit ^h -il/	mic ^h -il ~ mit-il	‘bottom’
b. /c ^h /	(i) /k’oc ^h -ilo/	k’ot ^h -ilo	‘flower’
	(ii) /k’oc ^h -a/	k’ot ^h -a ~ k’oc-a ~ k’ot-a	‘flower’
	(iii) /nac ^h -ilo/	nat ^h -ilo	‘face’
	(iv) /tac ^h -il/	tat ^h -il ~ tac-il	‘anchor’
	(v) /toc ^h -e/	tot ^h -e ~ toc-e ~ tot-e	‘sail’
	(vi) /pic ^h -il/	pit ^h -il	‘light’
c. /c/	(i) /nac-ilo/	nac ^h -ilo	‘daytime’
	(ii) /pic-in/	pic ^h -in	‘debt’
	(iii) /møkəc-e/	møkəc ^h -e	‘uvula’

Notice that these variant forms are neither identical to their underlying forms nor the result of the rules or constraints mentioned above. (See Oh (2006), Oh & Shin (2007), and Jun & Lee (2007) for additional examples of the problematic variants.)

One might consider the possibility that different mechanisms derive different variants. For instance, those ending in [s] are derived by one of those in (9), and other variants shown in (10) are derived through independently motivated mechanisms. It will be shown in section 5 that this mixed approach is subject to some serious problems.

2. Frequency facts

In this section, I will discuss frequency facts of Korean nouns, showing that the observations regarding noun stem-final variations, summarized in (8), can be understood given the distribution of lexical final obstruents of suffixed nouns. Some previous studies (Ko 1989; Hayes 1998; Albright 2005, 2008; Kang 2003a,b) attribute the occurrence of the [s]-final variants to the “analogical” extension of the dominant alternation pattern in the noun paradigms. As discussed above, unmarked isolation forms of nouns and their corresponding forms marked with vowel-initial suffixes show stem-final alternations due to coda neutralization such as t-s (e.g., [maʔ] vs. [masil] ‘taste, isolation vs. accusative’) and t-c^h (e.g., [k’oʔ] vs. [k’oc^hil] ‘flower, isolation vs. accusative’). Since, as will be shown below, /s/-final nouns are most frequent in the lexicon among those ending in coronal obstruents, the t-s alternation would be the most frequent one among them, thus being analogically extended to nouns with less frequent alternations such as t-c^h, t-t^h, t-c and t-t alternations. As a result, less frequent stem-final coronal obstruents of some nouns are replaced with the most frequent [s].

Albright (2005, 2008) further argues that the occurrence of [c^h]-final variants is also due to high lexical frequency of /c^h-final nouns.¹⁰ In other words, not only t-s but also t-c^h alternations are extended, replacing less frequent ones. Albright’s argument is specifically based on corpus counts (type frequency) shown in (11a) which are calculated from 43,932 nouns in the Sejong text corpus of 1.5 million words (Kim & Kang 2000). Almost 80% of the stems end in /s/ or /c^h/ among coronal-final nouns. However, the Sejong corpus includes non-standard dialectal forms and loanwords. They should not be counted as nouns of the dialect of Seoul-Gyeonggi area, which may roughly be considered the standard dialect of Korean. Thus, in (11b), I revised and updated the corpus counts by excluding non-standard dialectal forms and loanwords

¹⁰ Details of his analysis are discussed in section 3.1.

from Kang & Kim's (2004) lists of 75,953 nouns attested in the expanded Sejong corpus of 5.5 million words.¹¹ Specifically, looking up the Standard Korean Dictionary (http://stdweb2.korean.go.kr/search/List_dic.jsp), I have excluded nouns which are “not listed in the dictionary”, “listed but classified as non-standard dialectal forms”, or “listed as loanwords”. This correction is made only for coronal-final nouns, not those ending in non-coronal obstruents which show a large difference between lenis vs. aspirated/fortis categories. Notice in (11b) that /s/ and /c^h/ still form a majority of lexical final coronal obstruents, 78%, which supports Albright's argument. In order to obtain a more reliable estimate of the distribution of lexical stem-final obstruents, I add dictionary counts, shown in (11c), from the reverse dictionary of Korean (You 1985).

(11) Distribution of final obstruents of Korean nouns

a. Corpus counts (from Albright (2008))

<u>Labial</u>		<u>Coronal</u>		<u>Velar</u>	
p	1,360	t	1	k	5,994
p ^h	64	t ^h	113	k ^h	18
p'	0	t'	0	k'	6
		c	17		
		c ^h	160		
		c'	0		
		s	375		
		s'	0		

b. Corpus counts (from Kang and Kim (2004); coronal-final loanwords and non-standard dialectal forms are excluded)

<u>Labial</u>		<u>Coronal</u>		<u>Velar</u>	
p	2,515	t	2	k	7,537
p ^h	84	t ^h	116	k ^h	24
p'	0	t'	0	k'	9
		c	18		
		c ^h	164		
		c'	0		
		s	308		
		s'	0		

¹¹ Kang & Kim (2004) provide token and type frequencies of words attested in the Sejong text corpus of 5.5 million words which were constructed from 1999 until 2001. The corpus consists of words chosen from news papers (22.6%), magazines (11.1%), books such as novels/essays (27.5%), books of other types (34.2%), etc (4.8%). (For detailed information about the Sejong project, see <http://www.sejong.or.kr/>.) Kang and Kim's lists of nouns include not only mono-morphemic nouns but also compound nouns which are listed as separate entries in Standard Korean Dictionary (written in Korean [phyocwunkwuketaesaceon], published in 1999, Dusan Dong-A).

c. Dictionary counts (from You 1985)

<u>Labial</u>		<u>Coronal</u>		<u>Velar</u>	
p	935	t	4	k	1,234
p ^h	96	t ^h	184	k ^h	17
p'	0	t'	0	k'	10
		c	38		
		c ^h	171		
		c'	0		
		s	473		
		s'	0		

Both corpus and dictionary counts show similar relative frequencies of nouns with final obstruents except that /c^h/ is more frequent than /t^h/ in corpus counts (11a,b) whereas /t^h/ is more frequent than /c^h/ in dictionary counts (11c) although the difference is small. Thus, the following hierarchies among lexical final obstruents may arise:

(12) Frequency hierarchies among lexical final obstruents

a. Coronal: s >> c^h, t^h >> c, t, (t', c', s')

b. Noncoronal: k >> k^h, k'

p >> p^h

Two points need attention. First, Albright's argument works only with corpus counts, not dictionary counts, since /c^h/-final nouns are not more frequent in the dictionary than /t^h/-final nouns as shown in (11c). It seems that corpus counts are more relevant to the current discussion of variation facts than dictionary counts. The reason for this will be discussed in the next section.

Second, more importantly, the frequency hierarchies, shown in (12), are overall consistent with observations on noun stem-final obstruent variations, discussed above and summarized in (8). As first pointed out by Ko (1989) and followed by Hayes (1998), Albright (2005, 2008) and Y. Kang (2003a,b), a majority of coronal-final noun stems in the lexicon end in /s/ which is frequently adopted as the final sound of the variant forms of the coronal-final nouns (8b). Similarly, only a few coronal-final noun stems end in /c/ and /t/ which are least frequently adopted as the final variant sound (8c). A similar system of frequency matching can be conceived for the variants ending in noncoronal lenis stops. The occurrence of lenis-final variant forms for stems ending in underlying noncoronal aspirated/tense stops may be related to the higher lexical frequency of the noun stems ending in noncoronal lenis stops /p, k/, compared to those ending in

aspirated/tense stops, /p^h, k^h, k'/. Thus, these frequency matchings suggest that most observations on the variation summarized in (8) can be explained by hypothesizing analogical extension of frequent alternations. More specifically, it may be hypothesized that stem-final variations in Korean, discussed above, occur while frequent alternations—such as t-s of coronal final nouns and p-p and k-k of non-coronal final nouns—are analogically extended to nouns with less frequent alternations. (Formal details of this explanation will be elaborated in the next section.)

What about the suffix-initial vowel effect favoring [c^h-i] and [t^h-e] sequences, summarized in (8e,f)? Expecting that the type frequency of the corpus also plays a role in accounting for the suffix-initial vowel effect, I calculated type frequencies of lexical final obstruents of suffixed nouns from the 5.5 million word text corpus of the Sejong Project. Specifically, using a concordance program, Hanmaru (downloadable from <http://www.sejong.or.kr>), I first collected all the occurrences of nouns marked with five different suffixes (three [i]-initial and two [e]-initial) from the Sejong text corpus of 5.5 million words. From the collected occurrences, I have then removed duplicated entries so that type frequency of suffixed nouns can be calculated. Finally, as I did above with the distribution in (11b), by looking up Standard Korean Dictionary (http://stdweb2.korean.go.kr/search/List_dic.jsp), I excluded nouns which are “not listed in the dictionary”, “listed but classified as non-standard dialectal forms”, or “listed as loanwords”. Based on the resulting lists, type frequencies of suffixed nouns with each stem-final obstruent are calculated.¹²

(13) shows the distribution of lexical final coronal obstruents of noun stems when they are combined with five different suffixes, three [i]-initial and two [e]-initial. The numbers in the table in (13a) indicate how many noun stems ending in the corresponding obstruent are conjugated with the corresponding suffix. For instance, the corpus contains 157 different nouns marked with the suffix *-il* whose stems end lexically in /s/ as in /nas-il/ ‘sickle, accusative’. It is suggested here that different nouns may tend to be conjugated with different suffixes, possibly due to their lexical meanings. All words in the Sejong text corpus are written in Korean orthography, and thus they usually represent their underlying or etymological forms, i.e., an older stage of Korean. (13b) shows proportions of nouns with each lexical final obstruent among nouns marked with the same suffix.

As indicated by their mean proportions, etymologically /s/-final nouns are generally

¹² As in (11b), the database for frequency counting includes compound nouns, which are listed as separate entries in Standard Korean Dictionary.

most frequent (45%): their proportions are the highest before all [i]-initial suffixes and second highest before *-e* and *-esə*. The proportions of /c/ and /t/ are lowest, regardless of suffixes. This (relative) frequency of /s/, /c/, and /t/-final nouns is somewhat predictable from the corpus counts as well as dictionary counts shown in (11). Our main interest is in whether /t^h/ and /c^h/-final nouns, occurring in general with medium frequency in the corpus and dictionary, show different frequency patterns depending on the quality of the suffix-initial vowel. As summarized in (13c-i), etymologically /t^h/-final nouns are very frequent before [e]-initial suffixes, out-numbering etymologically /c^h/-final nouns by 38% on average and even /s/-final nouns especially when they occur before *-esə*. In contrast, as summarized in (13c-ii), /c^h/-final nouns are more frequent before [i]-initial suffixes than before [e]-initial suffixes, and their proportions are the second highest before *-il* and *-in*, following that of [s]-final nouns. This reversal in corpus counts is mostly matched with the frequency/preference reversal of the two intermediate variants depending on the suffix-initial vowel, discussed in the previous section and summarized in (8e,f).

Furthermore, a closer examination of the distribution in (13) shows that etymologically /t^h/-final nouns occur more frequently before the directive suffix *-ilo* (30%) than they do before other [i]-initial suffixes (19~20%), but less so than they do before [e]-initial suffixes (40~58%). This medial frequency of /t^h/-final nouns before *-ilo* can be matched with the medial acceptability of [t^h]-final variant forms before *-ilo* shown in Y. Kang (2007), as discussed in the previous section. Recall that [t^h]-final variants are more acceptable before *-ilo* than they are before *-il*.¹³

(13) Distribution of lexical final coronal obstruents of **suffixed** nouns

a. Raw frequency

Final C (↓) vs. suffix (→)	il	in	ilo	e	esə
s	157	79	67	67	23
t ^h	57	32	49	69	46
c ^h	68	49	43	30	10
c	9	8	7	7	1
t	0	0	0	0	0
total	291	168	166	173	80

¹³ The distribution in (13) shows no indication of the medial frequency of [c^h] responses shown in Choi's (2004) survey results, i.e., the relatively lower frequency of [c^h] responses before *-ilo*, compared to those before *-il*. But, recall that this medial frequency of [c^h] responses before *-ilo* is not confirmed in Y. Kang's well-formedness ratings data.

b. Proportions

Final C (↓) vs. suffix (→)	il	in	ilo	e	esə	mean
s	0.54	0.47	0.40	0.39	0.29	0.45
t ^h	0.20	0.19	0.30	0.40	0.58	0.29
c ^h	0.23	0.29	0.26	0.17	0.13	0.23
c	0.03	0.05	0.04	0.04	0.01	0.04
t	0.00	0.00	0.00	0.00	0.00	0.00

c. relative frequencies of [t^h] v. [c^h] (gray cells)

(i) /t^h/ >> /s/, /c^h/ before __e, __esə

(ii) /c^h/ >> /t^h/ before __il, __in

The fact that the distribution of suffixed nouns, presented above, is consistent with the suffix-initial vowel effect favoring [c^h-i] and [t^h-e] sequences, strongly suggests that the observed patterns of variation should be explained by referring to the (relative) frequency of inflected forms, not bare stems. Based on the above mentioned proposal of analogical extension of the dominant alternation, we may hypothesize that the t-c^h alternation, which is frequent especially before [i], is more likely to be analogically extended to nouns with less frequent alternations before [i] whereas the t-t^h alternation, which is frequent before [e], likewise tends to be extended to those with less frequent alternations before [e].

The table in (14) shows the distribution of lexical final noncoronal obstruents of noun stems when they are combined with the same five suffixes, *-il*, *-in*, *-ilo*, *-e*, and *-esə*. The relative frequency here is not different from the one shown in (12b). Regardless of the suffixes attached, the number of stems ending in lenis stops far exceeds the number of those ending in corresponding aspirated (and tense) stops.

(14) Distribution of lexical final noncoronal obstruents of suffixed nouns

a. Raw frequency

Final C (↓) vs. suffix (→)	il	in	ilo	e	esə
p	711	395	370	504	213
p ^h	32	18	24	30	18
k	2471	1401	1336	1562	731
k ^h	3	3	5	7	6
k'	4	3	5	6	4

b. Proportions

Final C (↓) vs. suffix (→)	il	in	ilo	e	esə	mean
p	0.9569	0.9564	0.9391	0.9221	0.9391	0.9437
p ^h	0.0431	0.0436	0.0609	0.0779	0.0609	0.0563
k	0.9972	0.9957	0.9926	0.9917	0.9865	0.9927
k ^h	0.0012	0.0021	0.0037	0.0044	0.0081	0.0039
k'	0.0016	0.0021	0.0037	0.0038	0.0300	0.0083

The frequency facts of suffixed nouns, presented thus far, are summarized below:

(15) A summary of frequency facts: type frequency of suffixed noun forms in the Sejong corpus

a. /s/ is in general the most frequent coronal obstruent in the stem-final position.

b. /c/ and /t/ are the least frequent.

(Tense coronal obstruents /s', c', t'/ are never attested.)

c. /c^h/ and /t^h/ are in general intermediate.

d. /c^h/ is relatively frequent before [i]-initial suffixes.

e. /t^h/ is relatively frequent before [e]-initial suffixes.

f. Among noncoronal obstruents, lenis stops are a lot more frequent than their aspirated/tense counterparts.

As emphasized above, these frequency facts are matched well with patterns of stem-final obstruent variation, summarized in (8). In the following section, adopting the synchronic model of morphology learning proposed by Albright (2005, 2008), I will provide an explicit formal account of this frequency matching and variation patterns of Korean nouns, summarized in (8).

3 Analysis

This section first introduces Albright's (2002a,b, 2005, 2008) model of paradigm learning while discussing his account of certain innovative coronal-final variants of Korean nouns. Adopting Albright's model and building on his discussion of Korean data, I will then provide an analysis of the entire variation facts, discussed in the previous section.

3.1 Paradigm learning model

One important task for language learners is to learn to produce inflected forms of words accurately. For this purpose, language learners choose one part of the paradigm, which is called the base, and construct rules which can derive the rest of the paradigm. In order to discover the most accurate rules, they consider various different forms within the paradigm as a potential base while checking how reliably the entire paradigm can be derived from each of the candidate forms. The most reliable form is chosen as the real base. To illustrate how to choose the base and construct the rules, consider the following hypothetical language, adopted in Albright (2005), in which a single nominative case marker, -us, and two genitive markers, -i: and -oris, are used, and thus all words may be divided into two classes, one with -us ~ -i: and one with -us ~ -oris.

(16) Neutralization in the nominative (from Albright 2005 #8)

<i>nominative</i>		<i>genitive</i>
[gluptus]	~	[glupti:]
[nokus]	~	[noki:]
[reptus]	~	[reptoris]
[kortus]	~	[kortoris]

Since distinct case markers are adopted only in the genitive, not nominative, the genitive→nominative mapping is more predictable than the nominative→genitive mapping. Specifically, if the nominative form is taken as the base, then two morphological rules, [-us]→[-i:] and [-us]→[-oris], would be posited and derive the genitive form only with 50% accuracy. In contrast, when the genitive form is taken as the base, two rules, [-oris]→[-us] and [-i:]→[-us], would still be posited but may derive the nominative form with 100% accuracy. Thus, the genitive form, which can derive its corresponding nominative counterpart reliably, should be taken as the base.

Let us now consider how to discover the morphological rules in Albright's model. He adopts the *minimal generalization* algorithm (Albright and Hayes 2002). It takes as input ordered pairs of inflected forms, such as those shown in (16). In order to construct a rule consisting of two components, namely structural change (SC) and structural description (SD), the two members of each pair must be compared. The common part is taken for SD and the changed part for SC. Resulting word-specific rules for the partial paradigms in (16) would be the following:

(17) Word-specific morphological rules for the words in (16) (from Albright 2005 #9)

- a. [us] → [i:] / glupt __#
- b. [us] → [i:] / nok __#
- c. [us] → [oris] / rept __#
- d. [us] → [oris] / kort __#

These word-specific rules are in turn compared, and if a pair of rules shares the SC, a more generalized rule would result. For instance, in the above hypothetical language, rules in (17c,d) share the change from [us] to [oris], and they are compared to determine the common environment for a generalized rule, as illustrated below:

(18) Generalization over pairs of related rules (slightly adapted from Albright 2005

Figure 2.1)

<i>change</i>	<i>residue</i>	<i>shared features</i>	<i>shared segments</i>	<i>change location</i>	<i>shared segments</i>
us → oris	re	p	t	—	#
us → oris	ko	r	t	—	#
us → oris	X	<div style="border: 1px solid black; padding: 2px; display: inline-block;"> -syl +cons -nas -lat -dors etc. </div>	t	—	#

Common properties on the left and right sides of the “change location”, shown under “shared segments/features” above, form the SD part of a new generalized rule. The remaining material with no common properties, shown under “residue” above, is represented as a free variable, X. Notice here that the resulting rule is not much more general than the two input rules in (17c,d). In a real language situation with a realistically large lexicon, comparisons of this type are repeated over and over again, yielding multiple rules for the same change. Some resulting rules would be very general if the input rules did not share much in the rule context.

Since multiple rules may be hypothesized for the same change, to discover how reliably each of these rules applies is very important. The *reliability* of a rule is calculated as the ratio of the number of the forms taking its SC, i.e., hits, against the total number of the forms meeting the SD of the rule, i.e., scope, as shown below:

(19) Definition of a rule's reliability (Albright 2005 #10)

$$\frac{\text{\# of forms included in the rule's structural change (= hits)}}{\text{\# of forms included in the rule's structural description (= scope)}}$$

Consider, for example, the above hypothetical language shown in (16). The SD of the generalized rule in (18) is met in three forms, ([gluptus], [reptus], and [kortus]), but only two of them have [-oris] as the genitive case marker, taking the SC of the rule. As a result, the rule in (18) has a reliability of 2/3, i.e., 0.67. The rule reliabilities are then adjusted under the assumption that rules covering only a few input forms must be penalized whereas rules based on a large data set must be rewarded. Specifically, using lower confidence limit statistics (Mikheev 1997), Albright calculates adjusted values of the reliability ratios, i.e., *confidence* values. In the calculation of the confidence, among two rules with the same reliability ratio, the one based on a larger amount of data has a higher value than the one based on a smaller amount of data. For instance, a rule with 5/5 reliability must have a lower confidence value than the one with 1000/1000 reliability. According to Albright's (2005) calculation in which a confidence level of $\alpha = .95$ is adopted, the former has a confidence of 0.825 and the latter has a confidence of 0.999.

A couple of characteristic properties of Albright's model need attention with respect to the analysis of the Korean data under consideration. First, it is type, not token, frequency that directly contributes to the rule reliability/confidence. The more forms in the learning data that are covered by a rule, the higher values of the reliability/confidence this rule has. What is more important in the current project is that it is the learning data set, not the conventional lexicon, that determines the confidence of rules. There is no guarantee that the actual learning data include all inflected forms of the paradigm, in part because the base selection and the rule creation should occur early in the learning process. It is then possible that forms of different inflectional categories may occur at different rates in the learning data, causing differences in the confidence of relevant rules. To illustrate this point, if certain forms within the paradigm, say nominative and dative ones, occur more often in actual speech than some other form, say accusative, then both nominative and dative forms of a noun are more likely to occur in actual speech than both accusative and nominative or both accusative and dative forms of the same noun. More pairs of nominative & dative forms may then be available to the learner, compared to pairs of accusative & nominative forms or pairs of accusative & dative ones. Accordingly, the learning data set would include more

nominative//dative pairs than accusative//nominative or accusative//dative pairs, and thus, all other things being equal, rules would have higher adjusted confidence for the former mapping than the latter. In summary, token frequency plays no direct role in confidence calculation since the same pair cannot be duplicated in the learning data. But, since frequently occurring forms are more likely to be available to the learner than infrequently occurring forms, the former is more likely to be part of the learning data, and thus more likely to be considered in confidence calculations, compared to the latter. Only in such an indirect way, token frequency may contribute to the confidence. (Notice that Bybee (1995, 2001), Pierrehumbert (2001) and Albright & Hayes (2003) also assume or argue that extension of morphological patterns is based on type, not token, frequency.)

Second, more than one rule may apply to the same base form. Among such multiple rules, the one with the highest confidence is the one that applies in a case where a new form is introduced into a language or the speaker's memory of some inflected forms is not perfect. Thus, this model predicts that language change such as the morphological class change and the appearance of innovative forms occurs based on the dominant pattern of the lexicon or, more precisely, the learning data set. As mentioned above, Albright in fact attributes the occurrence of [s]/[c^h]-final variant forms in Korean to the fact that a majority of noun stems with final coronal obstruents end in /s/ or /c^h/ etymologically. If we take unmarked isolation forms as the base, then there would be multiple rules mapping the unmarked base form to each of the other forms of the paradigm. For instance, among the rules mapping the unmarked to the accusative, the [t]→[sil] rule ("if the unmarked form ends in [t], its corresponding accusative form must end in [sil]") and [t]→[c^hil] rule ("if the unmarked form ends in [t], its corresponding accusative form must end in [c^hil]") have higher reliabilities than other competing rules since in the Korean noun lexicon--more precisely, in its corpus counts--the number of /s/ and /c^h/-final stems exceeds the number of stems ending in other coronal obstruents /c, t/. Thus, the occurrence of [s]/[c^h]-final innovative variants is expected within Albright's model if we assume that the unmarked form is the base.

Albright (2008) provides an in-depth discussion of why the unmarked form should be the base even though coda neutralization occurs in the unmarked form and thus the unmarked form is not informative about the manner and phonation-type information of the stem-final obstruent. He argues that the unmarked form is in fact more informative about how to derive other paradigmatically-related forms than suffixed forms. This argument is based on two facts. First, a relatively small number of nouns are subject to

the coda neutralization.¹⁴ According to Albright (2008) based on Kim & Kang's (2000) report on the Sejong corpus, roughly 18% of Korean nouns are obstruent-final, and thus potentially ambiguous in the unmarked form due to the coda neutralization. In contrast, vowel-final nouns (39% of Korean nouns) and /l/-final ones (8%) are ambiguous in the accusative in which /-il/ and /-lil/ are used for C-final and V-final stems, respectively: for instance, /il-il/ [iril] 'work, accusative' vs. /i-lil/ [iril] 'tooth, accusative'.¹⁵ Thus, the unmarked form is at least more informative than the accusative form. This is confirmed by Albright's learning simulations in which the minimal generalization model is used to learn a set of partial paradigms of Korean nouns attested in the Sejong corpus. The following informativeness of unmarked, nominative and accusative forms is reported in Albright (2008).

(20) Relative informativeness of unmarked, nominative and accusative forms (from Albright 2008, #23-24)

	Accuracy of grammar	mean confidence
unmarked → accusative	97.5%	.971
nominative → accusative	98.6%	.986
accusative → unmarked	93.0%	.929
accusative → nominative	93.5%	.932

However, the above simulation results also suggest that the nominative base is more reliable than the unmarked although the difference is very small. As briefly mentioned above, token frequency may play a role, though an indirect one, in determining the rule confidence. The above simulation is carried out under the assumption that all inflected forms of each noun are equally frequent in actual speech. But, in Korean, unmarked forms are a lot more frequent than nominative and accusative forms. In child-directed speech, unmarked, nominative and accusative forms take 75%, 20%, and 5% of the occurrence, respectively (Albright 2008 citing I. Lee 1999). Thus, the same word is more likely to be produced in both the unmarked and accusative forms than in both the nominative and accusative forms. Accordingly, more unmarked-accusative pairs would

¹⁴ Silverman (2009) recently showed that the amount of homophony derived through coda neutralization in Korean is remarkably low.

¹⁵ In Korean, the underlying lateral appears as a tap [r] when it is a single onset.

be included in the learning data, compared to nominative-accusative pairs. Since rules based on a larger data set are rewarded in the confidence adjustment, the rules based on unmarked-accusative pairs may have higher confidence than those based on nominative-accusative pairs. This is confirmed in the results of Albright's revised learning simulation where token frequency, mentioned above, is taken into account in preparing the learning data: the mean confidence of the unmarked→accusative mapping is .795 whereas that of the nominative→accusative is .461 (Albright 2008 #27).

In summary, since there are more unmarked forms in the learning data simply due to the fact that they are more frequently available to the learner, rules based on input pairs with the unmarked forms have relatively higher confidence values, and thus the unmarked form can be a more reliable base. This conclusion correctly predicts that the location of the change in Korean nouns is in the suffixed form, not unmarked form. Also, the occurrence of innovative [s], [c^h]-final forms can be explained by another prediction of the model that language change proceeds in the direction favored by the grammar.

3.2 Analysis of Korean noun paradigms

Building on and extending Albright's account presented above, I will provide an analysis of the observed stem-final variation patterns of Korean nouns, primarily the relative preference of stem-final variants including the suffix-initial vowel effect favoring [c^h-i] and [t^h-e] sequences.

To construct morphological rules for the acquisition of Korean noun paradigms and determine their reliability/confidence, I will assume a learning data set based on the Sejong corpus. This is obviously inspired by the finding, discussed in section 2, that the relative frequency and preference of the stem-variants are relatively well matched with the lexical distribution of final obstruents of suffixed nouns attested in the corpus. This assumption can be considered somewhat unrealistic, since as mentioned in footnote 11, the Sejong corpus is a collection of written texts from magazines, novels, newspaper etc., and thus it would probably include words that learning children are unlikely to encounter. However, the corpus should be considered at least a more reliable and realistic source of the learning data than the dictionary which usually has thousands of outdated and rarely-used words. This can explain the observation, made in section 2, i.e., that corpus counts are more consistent with the relative frequency and preference of variants than dictionary counts. Also, given that the size of the corpus, i.e., 5.5 million words, is large, the set of words that are likely to be available to children is probably at

least a subset of the words contained in the corpus. It seems then that the relative frequency of alternation classes, the determinant of the rule reliability, may not be significantly different between the real learning data and the one based on the corpus. Finally, most importantly, the finding that Korean stem-final variation patterns and the related preference/acceptability of variants are mostly consistent with frequency facts based on the Sejong corpus, suggests that real learning data for Korean noun paradigms would not be significantly different from the one we can hypothesize based on the Sejong corpus. Thus, in a situation where no real learning data set is available to me, I will rely on the Sejong corpus and more specifically, frequency facts shown in (13-14) above.

Let us begin with the construction of rules mapping the unmarked to the accusative forms. Relevant input pairs of unmarked-accusative forms in the learning data would include pairs such as {mot, mosil}, {pat, pat^hil}, {k'ot, k'oc^hil} and {cət, cəcil}. According to the Minimal Generalization method discussed in the previous section, word-specific rules for these input pairs may be generalized as the following rules:

(21) Rules mapping the unmarked to the accusative

- a. $t \rightarrow \text{sil} / X_ \#$ (“if the unmarked form ends in [t], its corresponding accusative form must end in [sil]”)
- b. $t \rightarrow \text{t}^{\text{h}}\text{il} / X_ \#$ (“..., its corresponding accusative form must end in [t^hil]”)
- c. $t \rightarrow \text{c}^{\text{h}}\text{il} / X_ \#$ (“..., its corresponding accusative form must end in [c^hil]”)
- d. $t \rightarrow \text{cil} / X_ \#$ (“..., its corresponding accusative form must end in [cil]”)
- e. $t \rightarrow \text{til} / X_ \#$ (“..., its corresponding accusative form must end in [til]”)

To calculate the reliability and confidence values of each of these rules, I need to know the number of hits and scope of the rule, i.e., how many pairs in the learning data meet the SD of the rule and how many of those meeting the SD actually takes the SC of the rule. Based on the fact that, as mentioned in the previous section, unmarked isolation forms are a lot more frequent in natural speech than any of suffixed forms, I will take the frequency of the suffixed form in the corpus for the frequency of the input pair of unmarked-suffixed forms. Then, the number of hits and scope of the rules in (21) can be obtained from the distribution of final obstruents of suffixed nouns shown in (13), repeated below. For instance, the $t \rightarrow \text{sil}$ and $t \rightarrow \text{t}^{\text{h}}\text{il}$ rules in (21a,b) would have 157 and

57 hits, and both rules have 291 for the number of the scope.¹⁶

(22) Distribution of lexical final coronal obstruents of suffixed nouns

Final C (↓) vs. suffix (→)	il	in	ilo	e	esə
s	157	79	67	67	23
t ^h	57	32	49	69	46
c ^h	68	49	43	30	10
c	9	8	7	7	1
t	0	0	0	0	0
total	291	168	166	173	80

Using the definition of reliability in (19) and lower confidence limit statistics (Mikheev 1997), I calculate reliability and confidence values of the rules for the unmarked-to-accusative mapping as shown in (23a).¹⁷ Likewise, rules for other mappings are constructed, and their rule reliability/confidence values are calculated as shown in (23b-e). Notice that the numbers under hits/scope in (23) are from corresponding cells in (22).

¹⁶ Since there is no etymologically /t/-final noun in the corpus, regardless of the suffix, as shown in (13/22), rules for the alternation with pre-suffixal [t] cannot be constructed. Nonetheless, I will provide such rules in (21) and below for the sake of comparison with other rules constructed based on the attested words.

¹⁷ For the calculation of lower confidence limits, I follow Albright & Hayes (2002) and Mikheev (1997). (The following is based on footnote 4 of Albright & Hayes and Mikheev p. 413.) In the first step, reliability values are adjusted to avoid zeros in the numerator or denominator:

$$\hat{p}_i = \frac{(x_i + 0.5)}{(n_i + 1.0)}$$

Using this adjusted reliability value, the variance of the sample is estimated:

$$s_p = \sqrt{\frac{p(1-p)}{n}}$$

Using this variance, the lower confidence limit, at the confidence level α , is calculated:

$$\pi_L = \hat{p} - t_{(1-\alpha)/2}^{(n-1)} * s_p$$

$t_{(1-\alpha)/2}^{df}$ is a coefficient of the t-distribution: α is the level of confidence and df is the degree of freedom.

Here I adopt a confidence level of $\alpha = 0.75$. The t value for a given α can be looked up in the table for the t-distribution, which is usually provided in statistics textbooks. (Thanks to Adam Albright for helping me figure out all this.)

(23) Reliability and confidence of the rules for coronal-final nouns

	mapping to...	<u>Rule</u>	<u>Hits/scope</u>	<u>Reliability</u>	<u>Adjusted confidence</u>
i-initial	a. accusative	t → sil	157/291	0.540	0.520
		t → t ^h il	57/291	0.196	0.181
		t → c ^h il	68/291	0.234	0.218
		t → cil	9/291	0.031	0.026
		t → til	0/291	0.000	0.000
	b. topic	t → sin	79/168	0.470	0.444
		t → t ^h in	32/168	0.190	0.172
		t → c ^h in	49/168	0.292	0.269
		t → cin	8/168	0.048	0.039
		t → tin	0/168	0.000	0.000
	c. directive	t → silo	67/166	0.404	0.378
		t → t ^h ilo	49/166	0.295	0.272
		t → c ^h ilo	43/166	0.259	0.237
		t → cilo	7/166	0.042	0.034
		t → tilo	0/166	0.000	0.000
e-initial	d. locative/goal	t → se	67/173	0.387	0.363
		t → t ^h e	69/173	0.399	0.374
		t → c ^h e	30/173	0.173	0.156
		t → ce	7/173	0.040	0.033
		t → te	0/173	0.000	0.000
	e. locative/source	t → sesə	23/80	0.288	0.256
		t → t ^h esə	46/80	0.575	0.537
		t → c ^h esə	10/80	0.125	0.104
		t → cesə	1/80	0.013	0.008
		t → tesə	0/80	0.000	0.000

These multiple stochastic rules and their confidence values are responsible for the variation patterns and relative preference/acceptability of variants summarized in (8). First, [s]-final variants are in general most frequent and preferred since t→s rules generally have the highest confidence values (mean value = 0.392). Second, [c], [t]-final variants are rare and least preferred since t→c and t→t rules have lowest confidence

values across rules of different mappings. Third, [c^h]-final variants are frequent and preferred before [i]-initial suffixes since t→c^h rules have higher confidence values before *-il* (0.218), *-in* (0.269), and *-ilo* (0.237) than before *-e* (0.156) and *-esə* (0.104). Fourth, [t^h]-final variants are frequent and preferred before [e]-initial suffixes since t→t^h rules have higher confidence values before *-e* (0.374) and *-esə* (0.537) than before *-il* (0.181), *-in* (0.172) and *-ilo* (0.272). Also, the medial character of *-ilo*, discussed in section 1.2, can be captured in a similar way: *-ilo* is medial between other [i]-initial and [e]-initial suffixes with respect to Korean speakers' acceptability of [t^h]-final variant forms since the confidence of the t→t^hilo rule (0.272) is higher than those of t→t^hil (0.181) and t→t^hin (0.172) but lower than those of t→t^he (0.374) and t→t^hesə (0.537).

Let us now consider rules for noun stems ending in noncoronal obstruents. As above, my calculation of the reliability and confidence values of the rules for nouns with noncoronal obstruents is based on the lexical distribution of final obstruents of suffixed nouns shown in (14a), repeated below.

(24) Distribution of lexical final noncoronal obstruents of suffixed nouns

Final C (↓) vs. suffix (→)	il	in	ilo	e	esə
p	711	395	370	504	213
p ^h	32	18	24	30	18
k	2471	1401	1336	1562	731
k ^h	3	3	5	7	6
k'	4	3	5	6	4

Since the difference between lenis-final and aspirated/fortis-final nouns is so large across suffixes, it is quite predictable that all rules for lenis-final nouns have much higher reliability/confidence values than rules for aspirated/fortis-final nouns. To illustrate this point, I below provide rules involving the same five suffixes and their reliability/confidence values based on the frequency facts in (24).

(25) Reliability and confidence of the rules for non-coronal final nouns

	mapping to...	<u>Rule</u>	<u>Hits/scope</u>	<u>Reliability</u>	<u>Adjusted confidence</u>
i-initial	a. accusative	p → p ^h il	711/743	0.957	0.951
		p → p ^h il	32/743	0.043	0.039
		k → kil	2471/2478	0.997	0.996
		k → k ^h il	3/2478	0.001	0.001
		k → k' il	4/2478	0.002	0.001
	b. topic	p → pin	395/413	0.956	0.948
		p → p ^h in	18/413	0.044	0.038
		k → kin	1401/1407	0.996	0.994
		k → k ^h in	3/1407	0.002	0.002
		k → k' in	3/1407	0.002	0.002
	c. directive	p → pilo	370/394	0.939	0.930
		p → p ^h ilo	24/394	0.061	0.054
		k → kilo	1336/1346	0.993	0.991
		k → k ^h ilo	5/1346	0.004	0.003
		k → k' ilo	5/1346	0.004	0.003
e-initial	d. locative/goal	p → pe	504/534	0.944	0.936
		p → p ^h e	30/534	0.056	0.050
		k → ke	1562/1575	0.992	0.990
		k → k ^h e	7/1575	0.004	0.004
		k → k' e	6/1575	0.004	0.003
	e. locative/source	p → pesə	213/231	0.922	0.908
		p → p ^h esə	18/231	0.078	0.068
		k → kesə	731/741	0.987	0.983
		k → k ^h esə	6/741	0.008	0.006
		k → k' esə	4/741	0.005	0.004

Rules of this sort and their confidence values may be responsible for the occurrence of the innovative lenis-final variants of nouns with final non-coronal obstruents. All p → p and k → k rules have reliability/confidence values above 0.9 whereas all p → p^h and k → k^h/k' rules have values below 0.05. Accordingly, it is quite predictable that when the speakers' memory is not perfect, words with less frequent alternations such as p-p^h can be lexically reanalyzed as those with more frequent alternations such as p-p, being

subject to rules with high confidence such as the $p \rightarrow pil$ rule.

In summary, the occurrence of multiple stem-final variants of Korean nouns as well as their relative frequencies/preferences can be explained by multiple stochastic rules with the unmarked base and their rule reliability/confidence values, respectively. Notice that the pronunciation of the inflected form of a loanword such as [kus-in] ‘loanword: good, topic’ can be explained with the same mechanism. When a novel word is borrowed into Korean and it is conjugated with a suffix, general rules with highest confidence such as $s \rightarrow sin$ must apply since there is no lexical entry or rule specific to this word in Korean speakers’ memory.

Under Albright’s Paradigm Learning Model, I have assumed that reliability/confidence values are determined based on individual speakers’ learning input data. The learning input data, though similar, may not be identical across language learners (as well as dialects) since they usually have different language experience. Thus, speaker-dependent experience may plausibly explain the observed occurrence of speaker-dependent variation as well as speaker-dependent well-formedness judgment ratings. Since different dialects normally do not have exactly identical vocabularies, dialectal differences in the variation discussed by Han (2002) and Y. Kang (2005, 2007) may possibly be explained in an analogous way.

In order to verify that native Korean speakers have knowledge of the multiple stochastic rules provided in this section, I will test productivity of those rules by employing a certain set of Korean bound stems as experimental tokens.

4. Bound stems

In the previous section, to explain the fact that multiple variants exist and their relative preference is different, I have argued that multiple stochastic rules need to be posited, and they have different reliability/confidence according to the frequency of the relevant stem-final obstruent in the corpus. In this section, I will investigate productivity of these rules for the purpose of verifying that the rules and their rule reliability/confidence are part of individual Korean speakers’ mental grammar.

In Korean, certain noun stems are combined with the verb stem *ha* ‘do’ to form compound verbs, as shown in (26a). This compounding is productive, and it is a usual way to derive verbs from loanwords, as illustrated in (26b).

(26) Compound verbs

a. Noun stem + *ha* = verb stem (H-S Sohn 2001)

- (i) *il* 'work, noun'
- (ii) *il-ha-ta* 'work, verb, infinitive'
- (iii) *il-ha-ko* 'work, verb, connective'

b. loanword + *ha* = verb stem

- (i) *keim* 'English loanword: game, noun'
- (ii) *keim-ha-ta* 'play a game, infinitive'
- (iii) *keim-ha-ko* 'play a game, connective'

The same type of compounding is also adopted to derive verbs from a certain type of bound stems, as shown below:¹⁸

(27) Compound verbs: Bound stem + *ha* = verb stem¹⁹

- | | | | |
|------------------------------------|-----------------------|--------------------------------|-----------------------|
| (i) <i>p^hokin-ha-ta</i> | 'warm, infinitive' | <i>p^hokin-ha-ko</i> | 'warm, connective' |
| (ii) <i>pisish-ha-ta</i> | 'similar, infinitive' | <i>pisish-ha-ko</i> | 'similar, connective' |
| (iii) <i>taptap-ha-ta</i> | 'stuffy, infinitive' | <i>taptap-ha-ko</i> | 'stuffy, connective' |
| (vi) <i>t'okt'ok-ha-ta</i> | 'clever, infinitive' | <i>t'okt'ok-ha-ko</i> | 'clever, connective' |
| (v) <i>t'akt'ak-ha-ta</i> | 'hard, infinitive' | <i>t'akt'ak-ha-ko</i> | 'hard, connective' |

As the name indicates, these bound stems normally do not stand by themselves although some of them may combine with the adverbializing suffix *-i* in addition to the verb *ha*: for example, *p^hokin-i* 'warmly'. Mainly due to this limited distribution, it is very difficult to determine the exact lexical category of the bound stems, and they are thus sometimes classified as an independent category in Korean morphology (for instance, Kang & Kim 2004). Some bound stems of this type belong to the mimetic vocabulary. As shown in (27iii-v), many of them in fact have a form of reduplication which is typical of mimetic and onomatopoeic words in Korean. Such word forms and meanings suggest that the bound stems under consideration may belong to a type of adverb. But, as discussed by C-S Kim (2001, 2008), the bound stems are combined with *ha* just like real nouns, but unlike real adverbs, suggesting that they are a type of noun (H-M Sohn 2001). This nominal classification of the bound stems under consideration will be

¹⁸ The resulting compounds would be more correctly classified as adjectives. But, in Korean, the distinction between verbs and adjectives is not always clear mainly because they are inflected in a similar manner. Since the distinction is not relevant to any important aspects of this paper, for simplicity's sake I will consider the compounds as verbs in this paper.

¹⁹ In Korean, /h/ is phonetically realized differently depending on neighboring segments. Between sonorants, it is optionally deleted (27i). When /h/ is adjacent to a lenis stop, the h//stop sequence merges into an aspirated stop (27ii-v). See Kim-Renaud (1986) and Ahn (2001) for details.

supported by the results of the present survey.

The focus of the present study is on the topicalized forms of these bound stems in which a topic marker *-in* is inserted between the bound stem and *ha*, as shown in (28). When the bound stems ending in obstruents are topicalized, as in (28ii-v), the final obstruents appear in prevocalic position, which is a probable location for variation.

(28) Topic formation of bound stems: [stem + *in* + ha]

- (i) p^hokin-in-ha-ta ‘warm, topicalized, infinitive’
- (ii) pisis-in-ha-ta ‘similar, topicalized, infinitive’
- (iii) taptap-in-ha-ta ‘stuffy, topicalized, infinitive’
- (vi) t’okt’ok-in-ha-ta ‘clever, topicalized, infinitive’
- (v) t’akt’ak-in-ha-ta ‘hard, topicalized, infinitive’

To explore the phonetic realizations of the final obstruents, I consulted with native Korean speakers. The next section discusses the process and method of the judgment survey. For now, I want to emphasize that, although the topic formation of bound stems is completely acceptable to native Korean speakers, it is very rarely used in actual speech. So, the judgment task, described below, would be a type of wug-test (Berko 1958) for most native Korean speakers who participated in the experiment. (This assumption will be discussed in section 4.3.)

4.1 Well-formedness judgment survey

Bound stems employed in the task are listed in (29), where the stems are transcribed based on standard Korean orthography.

(29) Test stems

place of stem-final obstruent	<u>Stems</u>	<u>glosses</u>	<u>Stems</u>	<u>glosses</u>
coronal (11 stems)	pisis	similar	k’ek’is	clean
	t’at’is	warm	t’əst’əs	aboveboard
	himus	pleased	nikis	relaxed
	iicəs	dignified	et ^h is	pitiful
	p’utis	satisfied	c’alis	stimulating
	pantis	decent		
labial	pokcap	crowded	takip	imminent

(11 stems)	səŋkip	impatient	pikəp	cowardly
	cokip	hasty	mihip	insufficient
	taptap	stuffy	səpsəp	sorry
	k'amucapcap	darkish	kapkap	stuffy
	c'ipc'ip	unsatisfied		
velar (12 stems)	simkak	serious	palamcik	desirable
	solcik	honest	nəknək	enough
	t'akt'ak	hard	t'okt'ok	clever
	k'imc'ik	horrible	kipkjək	drastic
	makmak	desolate	t'ətɪls'ək	noisy
	katik	full	s'iks'ik	lively

These stems were combined with a topic marker *-in* followed by an inflected form of the verb *ha*. The resulting target words, together with their possible phonetic forms, were put in the context of sentences in the format shown in (30). The phonetic forms of each target word differ only in the stem-final obstruent. Five coronal obstruents [s, t^h, c^h, c, t] and two noncoronal ones, lenis and aspirated, were employed as the stem-final obstruent.

(30) Example experimental sentences

a. Final obstruent = coronal

wancəhi kat^hci-nin anh-təlato pisis-in he-jaci. (ha + əjaci → hejaci)
 [pisisin] ()
 [pisi^hin] ()
 [pisi^hin] ()
 [piscin] ()
 [pisi^hin] ()

‘completely’ ‘same-Top’ ‘not-although’ ‘similar-Top’ ‘do-should’
 ‘Even if (they) are not completely identical, they should be similar.’

b. Final obstruent = labial

ki salam səŋkjək-i səŋkip-in ha-nte, kiləto c^hakhe.
 [səŋkipin] ()
 [səŋkip^hin] ()

‘the’ ‘person’ ‘personality-Nom’ ‘impatient-Top’ ‘do-though’ ‘nonetheless’ ‘nice’
 ‘Although his personality is impatient, he is nice.’

The experimental sentences were randomly ordered and written on A4 size paper in Korean orthography (See Appendix for a complete list of the experimental sentences).

Thirty-one paid Seoul Korean speakers in their 20s through 40s were asked to evaluate the degree of acceptability of each form by assigning values between 1 to 4 where 4 means “good”, 3 “O.K.”, 2 “not impossible”, and 1 “impossible”. The subjects were allowed not to respond if the sentences sounded too awkward for them to determine their acceptability.

4.2 Results

The mean acceptability ratings of all 31 subjects for all 11 test bound stems with final coronal obstruents are shown in (31). The mean rating of [s]-final pronunciation, 3.78, is the highest, and the ratings of [c] and [t]-final pronunciations are lowest. Among two intermediate [t^h] and [c^h]-final pronunciations, [c^h]-final ones received a higher mean rating than [t^h]-final ones.

(31) Mean acceptability ratings (of all 31 subjects) for all 11 coronal-final bound stems

stem-final obstruent	s	t ^h	c ^h	c	t	total
Mean	3.78	2.03	2.39	1.92	1.90	2.4
standard deviation	0.35	0.57	0.59	0.72	0.61	0.91

Notice that this relative acceptability of potential variants (i.e., [s] >> [c^h] >> [t^h] >> [c], [t]) for bound stems suffixed with the topic marker *-in* is not qualitatively different from the observed preference and acceptability of variants of coronal-final nouns, discussed in section 1.2 and summarized in (8). Thus, not surprisingly, the acceptability hierarchy of the variants for the bound stems is well matched with the hierarchy of the reliability/confidence values of the corresponding morphological rules mapping the unmarked to topicalized forms, $t \rightarrow \text{sin} \gg t \rightarrow \text{c}^{\text{h}}\text{in} \gg t \rightarrow \text{t}^{\text{h}}\text{in} \gg t \rightarrow \text{cin} (\gg t \rightarrow \text{tin})$, shown in (23b), repeated below with the addition of the relevant acceptability ratings:

(32) Reliability and confidence of the unmarked → topic rules for coronal-final nouns

Rule	Hits/scope	Reliability	Adjusted confidence	cf. Acceptability ratings (from (31))
$t \rightarrow \text{sin}$	79/168	0.470	0.444	3.78
$t \rightarrow \text{t}^{\text{h}}\text{in}$	32/168	0.190	0.172	2.03
$t \rightarrow \text{c}^{\text{h}}\text{in}$	49/168	0.292	0.269	2.39
$t \rightarrow \text{cin}$	8/168	0.048	0.039	1.92

t → tin	0/168	0.000	0.000	1.90
---------	-------	-------	-------	------

The observed acceptability-confidence matching can be understood if Korean speakers do have knowledge of the proposed stochastic rules and give their ratings in the survey based on that knowledge. Then, the survey results would support the cognitive presence of these rules.

The mean acceptability ratings for the bound stems with final noncoronal obstruents are shown in (33). The mean ratings of lenis-final pronunciations are much higher than the mean ratings of aspirated-final pronunciations.

(33) Mean acceptability ratings (of all 31 subjects) for all noncoronal-final bound stems

stem-final obstruent	p	p ^h	k	k ^h
Mean	3.76	2.14	3.85	1.79
Standard Deviation	0.35	0.77	0.34	0.69

As above, it seems this relative acceptability of lenis-final pronunciations over aspirated-final ones for bound stems suffixed with the topic marker *-in* is not qualitatively different from the observed preference and acceptability of variants of noncoronal-final nouns, discussed in section 1.2 and summarized in (8). The relative acceptability of the variants for the bound stems is also matched well with the higher reliability/confidence values of the unmarked → topic rules for lenis-final nouns over those for aspirated-final nouns, as can be seen in (25b), repeated below with the addition of the relevant acceptability ratings.

(34) Reliability and confidence of the unmarked → topic rules for noncoronal final nouns

<u>Rule</u>	<u>Hits/scope</u>	<u>Reliability</u>	<u>Adjusted confidence</u>	cf. Acceptability ratings (from (33))
p → pin	395/413	0.956	0.948	3.76
p → p ^h in	18/413	0.044	0.038	2.14
k → kin	1401/1407	0.996	0.994	3.85
k → k ^h in	3/1407	0.002	0.002	1.79

As above, this acceptability-confidence matching may support the cognitive presence of

the rules shown above.

A somewhat more detailed acceptability-confidence matching can also be observed from the survey results for noncoronal-final stems. Notice in (33) that the acceptability rating difference between lenis and aspirated stops is larger for velar-final stems than for labial-final stems. The acceptability hierarchy is [k >> p >> p^h >> k^h]. This relative acceptability of potential pronunciations is also matched with the confidence hierarchy of the morphological rules for noncoronal final nouns [k → k^hin >> p → p^hin >> p → p^hin >> k → k^hin]. For now, it is not fully clear whether this matching across different places of articulation is meaningful or not.

In summary, the hierarchy involved in the mean acceptability ratings of potential pronunciations of the bound stems is not qualitatively different from the observed preference/acceptability of stem-final variants of Korean nouns, and it is well matched with the hierarchy of the reliability/confidence values of the morphological rules for nouns. This acceptability-confidence matching suggests that the morphological rules for nouns are responsible for the acceptability of potential variants of the bound stems, thus supporting the cognitive presence of such morphological rules.

4.3 Discussion

Let us consider some background information for the interpretation of the results of the judgment test, reported above, and potential problems which might be involved in the interpretation. I will first discuss arguments for the cognitive status of the proposed stochastic rules responsible for the stem-final variation.

In section 2, I have shown that Korean speakers' relative preference among stem-final obstruent variants of nouns is matched with the relative frequency of stems ending in an obstruent in the corpus. In section 3, I have shown that this frequency matching can be well understood within Albright's Paradigm Learning model in which multiple stochastic morphological rules are constructed based on a learning data consisting of pairs of inflected forms. Specifically, I have proposed rules and their reliability/confidence values for the learning of Korean noun paradigms based on the corpus counts of suffixed noun forms, as shown in (23) and (25). Results of the judgment survey exhibiting the same type of preference/acceptability as can be seen with nouns indicate that Korean speakers use the knowledge responsible for stem-final variation of nouns when they judge the acceptability of topicalized bound stems. Accordingly, the cognitive presence of the multiple stochastic rules and their reliability is supported.

Evidence in favor of the cognitive knowledge of the morphological correspondence generally comes from results of productivity tests such as wug-tests (Berko 1958) and loanword adaptation. Notice that my judgment survey must be a wug-test for most, if not all, participants. As mentioned above, the topic formation of bound stems is legitimate in Korean, and native Korean speakers have almost no difficulty understanding and producing it. Most participants in the present survey responded to most test items although they were allowed not to if they felt that the experimental sentences were unnatural.²⁰ Nevertheless, the adopted topic formation of bound stems is very rarely used in actual speech. I checked the 1.5 million word text corpus of Sejong Project (established in 1999) to obtain the token frequency of topicalized forms of thirty-four test stems employed in the judgment survey. I found only a single instance of a single word [mihɪp-in] ‘insufficient, topicalized’. The corpus includes no other occurrence of the topicalized forms of the test words despite the fact that all of them frequently occur without the topic marker, immediately followed by *ha*. Thus, the judgment survey of the present study is similar to a wug-test in that speakers are required to produce unattested inflected forms.²¹ Actually, the current survey provides more reliable information about Korean grammar than wug-tests since it employs real operations on real words in Korean unlike wug-tests which employ nonce words and thus might invite speakers to go outside their native system (thanks to Adam Albright for pointing this out to me). See also Jun & Lee (2007) for comparable evidence, from loanword adaptation, for the cognitive knowledge of the stem-final variation.

One potential problem with the present survey is that a letter //s// is used in the survey form for the stem-final *coronal* obstruents of the test bound stems. Also, letters

²⁰ One subject did not respond to four coronal-final and four velar-final items; three subjects did not respond to one labial-final item; and three subjects did not respond to one velar-final item.

²¹ There are several possible answers to a question of why the topic forms of the bound stems are rarely used despite the fact that they are completely acceptable to native Korean speakers. One simple answer might be that situations, in which bound stems need to be topicalized, occur very rarely in actual speech. Another answer, which I think is more plausible, would be the existence of an alternative way to topicalize the bound stems. Topic formation in Korean differs depending on whether the stem is a noun or a verb. Topic markers, -in and -nin, directly attach to the end of noun stems, and the choice depends on whether the stem ends in a consonant or a vowel: for instance, /salam-in/ ‘person, topic’ and /c^ha-nin/ ‘car, topic’. In contrast, in the topic formation of the verbs, topic markers attach to the nominal form of verb stems in which a nominalizing suffix -ki attaches to the end of the stem, and then the combined form is followed by the verb *ha* ‘do’ or re-appearance of the stem. For instance, the topic form of the verb stem /mæk/ meaning ‘eat’ is /mæk-ki-nin (ha-ta or mæk-ta)/. The bound stems under consideration may be topicalized according to the latter formation. The entire compound verb, consisting of the bound stem *plus ha*, becomes nominalized through the suffixation of -ki, and then the topic marker is attached to the resulting nominalized form. For instance, the topic form of /pisis-ha/ ‘similar’ is /pisis-ha-ki-nin (hata)/. This formation seems to be adopted as a default way of topicalizing the bound stems, leading to the rare use of the other topic formation which I adopt for the judgment survey.

of lenis stops, //p, k//, are used for the stem-final *noncoronal* obstruents. The use of //s, p, k// is in accordance with standard Korean orthography. Thus, it might be the case that high ratings of [s, p, k] pronunciations are not due to high confidence values of the relevant morphological rules, but simply due to the biased use of the letters in the survey forms.²² In addition, one reviewer pointed out that relatively high ratings of [t^h] can be expected from the fact that the bound stems normally occur with *ha* and the stem-final coronal stop *plus* the verb-initial [h] are always realized with [t^h], as mentioned in footnote 19. First of all, the remaining test results such as higher mean rating of [c^h] over [t^h] are not subject to these potential problems. In addition, high ratings of [s, p, k] pronunciations might already be predicted from the standard Korean orthography in which the bound stems under consideration are written with //s, p, k// for final sounds, regardless of how they are written in the survey form. The fact that the bound stems under consideration are written with //s, p, k// for final sounds in standard Korean orthography may suggest that the [s, p, k] pronunciations are at least considered dominant or default by Korean speakers.

Another problem concerns rules pertaining to the t-t alternation. Since there are no occurrences of etymologically /t/-final nouns in the corpus as can be seen in (13) and (22) (in fact, the number of underlyingly /t/-final nouns in Korean is virtually zero), rules for t-t alternations cannot be constructed according to the Paradigm Learning Model adopted in the present study. Therefore, the acceptability ratings of [t]-final forms of the bound stems, reported above, cannot be due to the reliability/confidence values of the rules for t-t alternation as such rules do not exist. Accordingly, my analysis incorrectly predicts that [t]-final variants of nouns would not occur. So, the questions are why the mean acceptability score of [t]-final forms of the bound stems is 1.90 (roughly meaning “not impossible”), not 1 (meaning “impossible”), and why [t]-final variants of nouns (e.g., [sot-e] /sot^h/ ‘pot, locative’) do occur although infrequently. One possible answer is Paradigm Uniformity or Base Identity (e.g., Kenstowicz 1996; Steriade 2000). If its effect in Korean noun paradigms is active but less so than the proposed multiple stochastic rules with relatively high confidence, the final [t] sound of isolation forms of nouns ending in coronal obstruents may sometimes be adopted in the suffixed forms, and Korean speakers’ acceptability of the [t]-final forms of the topicalized bound stems would not be zero.²³ So, my speculation is that the Paradigm

²² This problem has been pointed out to me by Kie Zuraw, Suyeon Yoon and one reviewer.

²³ Notice that this paradigm uniformity effect can be the effect of generalized identity rules such as [C_i → C_i-in] (“if the unmarked form ends in a consonant, its corresponding topicalized form must end in the same consonant.”). I do not know whether identity rules of this type are allowed within Albright’s Paradigm Learning Model. If this is the case, all variants of Korean nouns may be derived with no

Uniformity effect does exist in Korean noun paradigms, but the effect must be very limited. (See H. Ko (2006) for an in-depth discussion and an Optimality-Theoretic analysis of Paradigm Uniformity effects in Korean nominal inflection.)

5. Previous analyses

This section discusses some previous analyses of the variation patterns under consideration. The main contribution of this paper is to provide a unified account for the occurrence of most innovative stem-final variants of Korean nouns (except for [t]-final variants) and their relative frequency and preference. The multiple stochastic rules proposed in section 3.2 are responsible for the occurrence of four coronal-final variants [s, t^h, c^h, c] (and their relative preference), the occurrence of two noncoronal-final lenis stops [p, k], the well-formedness ratings of their corresponding pronunciations of bound stems as well as loanword pronunciations such as [kusin]. Recall that the range of coronal-final variants is quite wide, and their occurrence is not limited to certain underlying consonants. For instance, [c^h]-final variants have been reported to occur with underlyingly /t^h/, /c/, /s/-final nouns and even English loanwords ending in [t]. In addition, [c^h]-final pronunciation of the coronal-final bound stems, whose underlying status is not clear, received relatively high ratings of acceptability. As far as I know, no previous studies reach this level of explanation although a wide-scale descriptive investigation of variation patterns is needed for the final verification of the proposed analysis. This section focuses on how to deal with the suffix-initial vowel effect favoring [c^h-i] and [t^h-e] sequences. The suffix-initial vowel effects has in fact been observed and discussed mainly as exceptions to the dominance of [s]-final variant forms in previous research on Korean phonology and morphology.

Let us first consider previous accounts of the frequent occurrence of [c^h]-final variants before [i]-initial suffixes. Its standard explanation (for instance, Han 2002; H. Choi 2004; Park 2006) relies on Paradigm Uniformity effects. Underlyingly /t^h-final stems are realized as allomorphs ending with [c^h] in a palatalizing context, i.e., before [i]-initial suffixes, and these allomorphs influence the realization of allomorphs occurring in a non-palatalizing context, here before [i]-initial suffixes. For instance, under the influence of the standard output form [pac^hi] of /pat^h-i/ ‘field, nominative’,

exception by the same mechanism. But, according to my rough calculation, at least more than half of Korean nouns are non-alternating, and thus the reliability/confidence of such identity rules should be quite high. Then it is hard to understand why [t]-final variants are so rarely attested and why [t]-final forms of the topicalized bound stems are rated so low in the present survey.

/pat^h-il/ ‘field, accusative’ is realized as [pac^hil]. (Notice that this Paradigm Uniformity-based explanation is distinct from the possibility mentioned at the end of the previous section in that the base is the nominative, not unmarked isolation, form.) There are many problems with this account. First, recall that all five coronal obstruent variants may occur when the stems are combined with various different vowel-initial suffixes, including [i]-initial suffixes, although the relative preference among them may differ. If [c^h]-final forms occur due to Paradigm Uniformity, this explanation cannot be extended to the explanation of other coronal-final variants such as [s]-final and [t^h]-final variants, and thus separate mechanisms are necessary. Second, as can be seen in (9), most previous studies on Korean stem-final variations assume that unmarked isolation forms are the base for explaining the variations, mainly based on the fact that nominative and accusative case markers are often omitted in natural speech, and thus isolation forms have the highest token frequency. As discussed in section 3.1, Albright (2008) also argues that the unmarked isolation form is the base, based on its informativeness. It seems difficult to find comparable justification for claiming that the base of Paradigm Uniformity is the nominative form, inflected with the palatalizing suffix *-i*. Third, section 2 shows that [c^h]-final stems have a relatively high frequency in the corpus, only when they are combined with [i]-initial suffixes. The fact that [c^h]-final variants occur frequently in the same context, i.e., before [i]-initial suffixes, would be accidental within previous Paradigm Uniformity-based accounts. Finally and most importantly, as mentioned in section 2, [c^h]-final variants may occur even when the underlying obstruent is not /t^h/: for example, /nac-ilo/ [nac^h-ilo] ‘daytime, directive’. Notice that in its standard pronunciation, the underlying stem-final obstruent /c/ should surface as [c], not [c^h], in a palatalizing context, and thus the occurrence of [c^h] before a suffix *-ilo* cannot be due to the palatalized members of a paradigm. In conclusion, Paradigm Uniformity-based accounts cannot correctly explain the frequent occurrence of [c^h]-final variants before [i]-initial suffixes.

We are now in a position to consider the previous accounts of the frequent occurrence of [t^h]-final forms before [e]-initial suffixes. In the previous studies (H. Choi 2004: 20; cf. Park 2006: 196; Ko 1989: 19 and references therein), this suffix-specific fact may be attributed to the high token frequency of the locative suffix *-e* which is never omitted even in conversational speech unlike suffixes such as *-i* ‘nominative’ and *-il* ‘accusative’. It has been claimed that the high token frequency of noun forms suffixed with *-e* helps etymologically /t^h/-final stems to resist an analogical change to [s]-final forms or Paradigm Uniformity, and thus [t^h]-final stems are likely to surface as such before the suffix *-e*. This explanation is also subject to some difficult problems.

First, if the frequent occurrence of [t^h]-final forms is attributable to the frequent retention of the suffix *-e*, not only underlying /t^h/ but also other underlying obstruents such as /c^h/ should occur significantly more often in combination with *-e*. This is not true with /c^h/-final nouns, as shown below.

(35) Average ratio of standard pronunciation of 11 /c^h/-final nouns, classified by suffix types (H. Choi 2004: 67, Table 70)

<i>Suffix types</i>	<i>percentage of [c^h]</i>
i	61.92
e(sə)	48.45
ilo, il	55.66

In addition, as discussed above, the occurrence of stem-final [t^h] is not confined to etymologically /t^h/-final stems. For instance, [t^h]-final variants are attested with noun stems ending in /c^h/, especially when followed by a locative suffix *-e(sə)*: /suc^h-e/ [sut^he] ‘charcoal’ and /salkac^h-e/ [salkat^he] ‘the skin (surface)’. Therefore, the previous account relying on analogy-resistance cannot be considered as a general solution of the issue involving the frequent occurrence of [t^h]-final variants before [e]-initial suffixes.

Previous studies also propose a separate explanation for the occurrence of the lenis-final variants of the stem-final noncoronal obstruents. Most of them (Ko 1989; Davis & Kang 2006; cf. Park 2006; and others) rely on the effects of Paradigm Uniformity to the isolation form.²⁴ For instance, under the influence of the frequent isolation form [ip] of /ip^h/ ‘leaf’, /ip^h-i/ may be realized as [ipi]. Actually, as mentioned above, this Paradigm Uniformity effect might be active, but even if it does, it should not play a central role in explaining stem-final variation of nouns and relative acceptability or preference of the variants as suggested by the fact that [t]-final variants are least preferred and frequent among those ending in lexical coronal obstruents.

As discussed above, most previous approaches to Korean noun variation propose separate explanations for the occurrence of different variants, especially [s], [c^h] and [t^h].

²⁴ Although Davis & Kang (2006) adopt different types of constraints to explain the occurrence of [s]-final variants and the occurrence of noncoronal lenis-final variants (anticorrespondence and Paradigm Uniformity constraints, respectively), their analysis might be considered more unified than other alternatives since they propose at least a *single* constraint ranking. But, their analysis is not different from most other previous analyses in the limited level of explanation: it cannot be easily extended to explain the occurrence of many of “problematic” variants discussed in section 2.

Each of these previous explanations has its own problems, some of which are serious enough to merit rejecting their proposals. Worse is the fact that the proposed explanations are often incompatible with each other. Most previous approaches argue that the isolation form is the base in deriving the variant [s]. Specifically, the t-s alternation is extended to the stems ending in underlying coronal obstruents other than /s/ through the application of the rules and constraints in (9). However, the pre-[i] allomorph, not the isolation form, has been adopted as the base in the explanation of the variant [c^h] before [i]-initial suffixes. In addition, although the isolation form is adopted as the base in the explanation of the lenis variant of stem-final noncoronal obstruents, the effect involved is “paradigm uniformity”, not “alternation extension”. In summary, the assumptions that previous approaches make are conflicting with each other with respect to the base adopted as well as the effect involved.

Consequently, it is very difficult to maintain the previous approaches which propose independent accounts for the occurrence of variants. The proposed analysis in this paper is not totally unified, either, since a separate mechanism is needed to explain the occurrence of [t]-final variants. But, since the occurrence and (relative) preference of all other variants can be subject to the same mechanism (i.e., multiple stochastic rules whose confidence values are determined based on the type frequency of the relevant word pairs), the present analysis can be considered more unified than other alternative analyses.

6. Conclusions

The important points of the present study are summarized as follows.

First, based on survey and experimental evidence provided by some recent studies, I have shown that patterns involved in the variation of stem-final obstruents in Korean nouns are more complicated than previously thought. Crucially, several different coronal obstruents are involved in the variation, and their relative frequency and preference are different, often depending on the suffix.

Second, following the general idea of the analogy-based approaches (Ko 1989 and others), I have shown that the relative frequency and preference among variants of the noun stems are matched by the type frequency of suffixed nouns in the Sejong corpus.

Third, extending Albright’s (2005, 2008) analysis of certain variants of Korean nouns, I have proposed multiple stochastic rules for the learning of Korean noun paradigm and their rule reliabilities based on the corpus data. I have argued that the

proposed rules are responsible for the occurrence of most variants and their relative preference and acceptability.

Fourth, a similar preference/acceptability pattern has been observed in the results of a judgment survey employing a set of Korean bound stems. This observation has been taken as evidence that Korean speakers have the knowledge of the proposed stochastic rules.

This study provides not only empirical but also theoretical contributions to the literature. Among them, the contribution to Korean phonology and morphology is to provide a unified account for the occurrence of most coronal-final variants and their relative preference, the occurrence of all noncoronal-final variants as well as the relative acceptability of possible pronunciations of the bound stems. The proposed analysis can be considered more unified than any of the other alternatives although it is not completely unified since a separate mechanism is still needed to explain the sporadic occurrence and low, but not zero, acceptability of [t]-final variants. Another contribution of the present study is made to the literature on language acquisition. The present study provides support for Albright's learning hypothesis in which language learners internalize lexical statistical patterns by constructing stochastic rules based on the learning data to derive a set of paradigmatically-related forms.

Appendix

Experimental sentences (Bound stems)²⁵

1. Labial-final stems

pokcap 'crowded'	pusan-i com pokcap-in he poi-əs'-ciman k'ek'is-ha-n kos-i-əs'-ə Pusan-Nom a-little crowded-Top do look-Past-but clean-do-Con place-be-Past-SE 'Pusan looked a little crowded, but it was a clean area.'
takip 'imminent'	nalimtelo takip-in ha-nte, kileto amukəna sa-l su-nin in-one's-own-way imminent-Top do-but, though anything buy-Part thing-Top əps-ci nonexistent-SE 'It's urgent for me, but I shouldn't buy anything without consideration.'
səŋkip 'impatient'	ki salam səŋk'jək-i səŋkip-in ha-nte, kileto c ^h akhe the person personality-Nom impatient-Top do-but though nice 'Although his personality is impatient, he is still nice.'
pikəp 'cowardly'	ki salam pikəp-in he poi-ciman, kileto caki il-in cal the person cowardly-Top do look-but though self work-Top well al-asə he know-Con do 'He looked cowardly, but he can do his own work well.'
cokip 'rushed'	jənki-ja t ^h ant ^h an-ha-ni cokip-in ha-l p ^h iljo performance-Top 'sufficiently-good'-do-Con rushed-Top do-Part need əps-ə nonexistent-SE 'Since his performance is sufficiently good, he doesn't have to feel rushed.'
mihip 'insufficient'	lep ^h ot ^h i-ka acik mihip-in ha-ciman, cec ^h ulha-l su-nin report-Nom yet insufficient-Top do-but, submit-Part thing-Top is'-kes'-ne existent-Fut-SE 'The report is still insufficient, but you may submit it.'
taptap 'stuffy'	c ^h əim ipwən-ila pjəŋsil səŋhwal-to taptap-in ha-nte cam-in first hospitalization-thus room life-too stuffy-Top do-but sleep-in cal wa well come 'Since this is my first hospitalization, I feel stuffy, but I sleep well.'
səpsəp 'sorry'	siwən-səpsəp-ha-njako? cəkəto səpsəp-in ha-ci fresh-sorry-do-Inter at-least sorry-Top do-SE 'Are you asking whether I feel satisfied and, at the same time, sorry for it? At least I feel sorry.'
k'amucapcap 'a little dark'	əlkul-i k'amucapcap-in ha-ciman ip'-ə face-Nom 'a-little-dark'-Top do-but pretty-SE 'The face is a little dark but it's pretty.'
kapkap 'stuffy'	nepok ip-inik'a kapkap-in ha-nte an-ip-in kə-potan t'at'is underwear wear-because stuffy-Top do-but not-wear-Part thing-than warm ha-tələku do-SE

²⁵ (Top = Topic marker, Nom = Nominative case marker, Acc = Accusative case marker, Poss = Possessive case marker, Con = Converb, Part = Participle, SE = Sentence Ender, Past = Past tense suffix, Fut = Future tense suffix, Inter = Interrogative suffix)

	'Since I wore underwear, I feel stuffy, but it's warmer than not wearing it.'
c'ipc'ip 'unsatisfied'	jəmsek ha-ljə-nik'a c'ipc'ip-in ha-nte, ha-ko-nin sip ^h -ko kəii dyeing do-Fut-thus unsatisfied-Top do-but do-Con-Top wish-and almost i-njən-il c ^h am-as'-kətin two-year-Acc patient-Past-SE 'I don't feel comfortable about dyeing, but I want to dye my hair. I have been waiting for two years.'

2. Coronal-final stems

pisip 'similar'	wancəni kat ^h ci-nin an ^h -təlato pisis-in he-jaci completely same-Top not-though similar-Top do-should 'Even if (they) are not completely identical, they should be similar.'
k'ek'is 'clean'	kileto k'ek'is-in he poi-canha though clean-Top do look-SE 'It looks clean, though.'
t'at'is 'warm'	i os-i mojaŋ-in ip'ici an ^h -ciman t'at'is-in he this clothing shape-Top pretty not-but warm-Top do 'This clothing looks ugly but it keeps me warm.'
t'əst'əs 'aboveboard'	kjəŋki-e-nin cja-s'-ciman, kileto t'ət'əs-in he game-Loc-Top lose-Past-but though aboveboard-Top do 'We lost the game, but we are open and aboveboard nonetheless.'
himus 'pleased'	cəŋmal coha-ha-nin kə-l po-ni himus-in ha-nte jaksok-il really like-do-Part thing-Acc look-Con pleased-Top do-but promise-Acc cik ^h i-lk'a kəkəŋ-ija keep-Con concern-SE 'I feel pleased by seeing them happy, but I am at the same time afraid that they might break their promise.'
nikis 'relaxed'	kileto səŋk'jək-i nikis-in he. though personality-Nom relaxed-Top do 'The personality is relaxed, though.'
iicəs 'dignified'	nai-ka əli-əsə kiləhci, heŋtoŋ-i iicəs-in he age-Nom young-Con though, behavior-Nom dignified-Top do 'He is young, but he behaves in a dignified manner.'
ε ^h is 'pitiful'	jəŋhwa-sok nam-njə-kan-ii salaŋ-i ε ^h is-in he movie-inside man-woman-between-Poss love-Nom pitiful-Top do 'The love between the man and the lady in the movie is pitiful.'
p'utis 'satisfied'	il-il mac ^h iko na-nik'a p'utis-in ha-ne work-Acc finish-Con come-because satisfied-Top do-SE 'Since I've finished the work, I feel satisfied.'
c'alis 'stimulating'	ki jəŋhwa c'alis-in ha-nte sucun-in əps-ə the movie stimulating-Top do-but level-Top nonexistent-SE 'The movie is stimulating, but it's low quality.'
pantis 'decent'	salam-i pantis-in he person-Nom decent-Top do 'The person is decent in manner.'

3. vela-final stems

simkak 'serious'	jəksi pjəŋ-i simkak-in ha-ciman koc ^h i-l su-nin is'-ə yet illness-Nom serious-Top do-but treat-Part thing-Top existent-SE 'The illness is serious, but it's treatable.'
palamcik 'desirable'	wənc ^h ik-təlo sa-nin ke palamcik-in ha-ciman t'elon principle-following live-Part thing desirable-Top do-but sometimes

	juŋt ^h oŋsəŋ-to p ^h iljo-he flexibility-too need-do 'Living by principles is desirable, but sometimes flexibility is needed as well.'
solcik 'honest'	namca-c ^h inku-ka solcik-in ha-nte maim-i jak-he man-friend-Nom honest-Top do-but mind-Nom weak-do 'The boy friend is honest, but he is weak-minded.'
nəknək 'enough'	ki acuməni sok-i nəknək-in hɛ the aunt inside enough-Top do 'The lady is generous.'
t'akt'ak 'hard'	saŋc ^h ə puwi-ka taŋcaŋ t'akt'ak-in ha-kes'-ciman, kot na-il-kəja injury part-Nom now hard-Top do-Fut-but, soon recover-Fut-SE 'The injured part is hardened right now, but it will be cured soon.'
t'okt'ok 'clever'	ilim-to ip'i-ko t'okt'ok-in ha-nte cəŋ-il p ^h johjən-ha-l cul name-too pretty-and clever-Top do-but emotion-Acc express-do-Part way molla not-know 'Her name is pretty, and she is clever, but she doesn't know how to express her emotion.'
k'imc'ik 'horrible'	ki jəŋhwa k'imc'ik-in ha-nte kiletə tasi po-ko sip ^h -ne the movie horrible-Top do-but though again look-Con wish-SE 'Although the movie was horrible, I still want to see it again.'
kipkjək 'drastic'	jocim kihu pjənhwa-ka kipkjək-in ha-ciman, tɛc ^h elo t'at'is-he recently weather change-Nom drastic-Top do-but, generally warm-do 'Recently, the weather change is drastic, but it's generally warm.'
makmak 'desolate'	ap ^h ilo i-njən kun-seŋhwal-il tə ha-l ke makmak-in hɛ future two-year army-life-Acc more do-Fut thing desolate-Top do 'I feel desolate about serving two more years in the army.'
t'ətɪls'ək 'noisy'	imotil-i nol-lə o-si-ni t'ətɪls'ək-in ha-kes'-ci aunts-Nom play-Con come-Hon-Con noisy-Top do-Fut-SE 'Since aunts are coming over, it will be noisy.'
katik 'full'	oscaŋ-ɛ os-i katik-in ha-nte maksəŋ ip-il kəs-in wardrobe-Loc clothes-Nom full-Top do-but timely wear-Acc thing-Top əps-ə nonexistent-SE 'My wardrobe is full of clothes, but I would like to wear none of them.'
s'iks'ik 'lively'	s'iks'ik-in hɛ-poi-ciman jəcənhi ap ^h -ajo lively-Top do-look-but still sick-SE 'He looks lively, but he is still sick.'

Acknowledgments

I would like to thank Sangjin Hwang and Soo-yun Park for their valuable assistance in corpus and survey data collection for this project. I am also grateful to Yoonjung Kang and the anonymous reviewers of the Journal of East Asian Linguistics for their useful suggestions and detailed comments.

This paper has greatly benefited from the advice and comments from many people. I would like to thank Adam Albright, Stuart Davis, Bruce Hayes, Jin-hyung Kim, Dan

Silverman, Donca Steriade, participants of my Phonology II class at Seoul National University and audiences at the 2006 meeting of the Phonology-Morphology Circle of Korea, Hanyang University, Seoul, Korea and the 2007 Annual Meeting of Linguistics Society of America, Anaheim, USA. I also thank Hyopil Shin, Seoncheol Kim, Beommo Kang and Hyewon Choi for helping me utilize the Sejong corpus and some useful survey data. Finally, I thank my experimental subjects for their time.

References

- Ahn, Sang-Cheol. 2001. *An introduction to Korean phonology*. Seoul: Hanshin.
- Albright, Adam. 2002a. *The identification of bases in morphological paradigms*. Ph.D dissertation, UCLA.
- Albright, Adam. 2002b. Islands of reliability for regular morphology: Evidence from Italian. *Language* 78.4: 684-709.
- Albright, Adam. 2005. The morphological basis of paradigm leveling. In *Paradigms in Phonological Theory*, ed. Laura Downing, Tracy Alan Hall, Renate Raffelsiefen, 17-43. Oxford: Oxford University Press.
- Albright, Adam. 2008. Explaining universal tendencies and language particulars in analogical change. In *Language universals and language change*, ed. Jeff Good, 144-181. Oxford University Press.
- Albright, Adam and Bruce Hayes. 2002. Modeling English past tense intuitions with minimal generalization. In *Proceedings of the 6th meeting of the ACL special interest group in Computational Phonology*, ed. M. Maxwell, 58-69. Philadelphia: ACL.
- Albright, Adam and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental Study. *Cognition* 90: 119-161.
- Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32.1: 45-86.
- Berko, Jean. 1958. The child's learning of English morphology. *Word* 14. 150-177.
- Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10. 425-455.
- Bybee, Joan. 2001. *Phonology and Language use*. Cambridge University Press.
- Choi, Hyewon. 2004. *A survey of standard pronunciation III [phyocunpalum silthecosa III]* Seoul: The National Academy of Korean Language.
- Choi, Kyung-Ae. 2002. [t]~[s] Alternation in Korean Loanwords, *Studies in Phonetics, Phonology and Morphology* 8.2: 289-302.
- Davis, Stuart and Hyunsook Kang. 2006. English loanwords and the word-final [t] problem in Korean. *Language Research* 42.2: 253-274.
- Han, Eunjoo. 2002. Optimal Paradigms in Korean nominal inflection. *Studies in Phonetics, Phonology and Morphology* 8.2: 303-322.
- Hayes, Bruce. 1998. On the richness of paradigms, and the insufficiency of underlying representations in accounting for them. Handout for Stanford talk, Apr. 2, 1998.
- Hayes, Bruce and Zsuzsa Cziraky Londe. 2006. Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology* 23: 59-104.
- Jun, Jongho and Jeehyun Lee. 2007. Multiple stem-final variants in Korean native nouns and loanwords. *Eoneohag* 47: 159-187.
- Kang, Beom-mo and Kim Hung-gyu. 2004. *Frequency analysis of Korean morpheme and word usage 2*. [hankwuke hyengtayso mich ehwi sayong pintouy pwunsek 2] Seoul: Institute of Korean culture, Korea University.
- Kang, Eunji, Ho-Young Lee and Juwon Kim. 2004. The phonetic realization of syllable codas in Korean. *Malsori* 49, Journal of the Korean Society of Phonetic Science and Speech Technology: 1-30.

- Kang, Yoonjung. 2003a. Sound changes affecting noun-final coronal obstruents in Korean. In *Japanese/Korean Linguistics* 12, ed. W. McClure, 128-139. Stanford: CSLI
- Kang, Yoonjung. 2003b. The emergence of phonetic naturalness in an analogical change: evidence from Korean nouns. Handout for LANYU Forum.
- Kang, Yoonjung. 2005. The emergence of the unmarked in an analogical change. Handout from the workshop on phonological typology, Seoul National University, December 9-10, 2005.
- Kang, Yoonjung. 2007. Frequency effects and regularization in Korean nouns. Handout from the workshop on variation, gradience and frequency in phonology, Stanford, July 6-8, 2007.
- Kenstowicz, Michael. 1996. Base-identity and uniform exponence: alternatives to cyclicity. In *Current trends in phonology: models and methods*, ed. J. Durand and B. Laks, 363-94. Paris X and Salford: University of Salford Publications.
- Kim, Chang-seop. 2001. On the relationship of 'Xha-' and 'X+reul ha-'. *Language Research* 37.1: 63-85.
- Kim, Chang-seop. 2008. *A study on the Korean morphology*. [written in Korean: hankwuke hyengtaelon yengu] Seoul: Taehaksa.
- Kim, Hung-gyu and Beom-mo Kang. 2000. *Frequency analysis of Korean morpheme and word usage 1*. [hankwuke hyengtayso mich ehwi sayong pintouy pwunsek 1] Seoul: Institute of Korean Culture, Korea University.
- Kim, Hyunsoon. 2001. A phonetically based account of phonological stop assibilation. *Phonology* 18: 81-108.
- Kim, Jin-hyung. 2005. A reconsideration of phonological leveling: A case of noun inflection in Korean. *Studies in Phonetics, Phonology and Morphology* 11.2: 259-274.
- Kim, Seoncheol. 2003. *A survey of standard pronunciation II* [phyocunpalum silthecosa II] Seoul: The National Academy of Korean Language.
- Kim-Renaud, Young-Key. 1986. *Studies in Korean Linguistics*. Seoul: Hanshin.
- Ko, Heejeong. 2006. Base-output correspondence in Korean nominal inflection. *Journal of East Asian Linguistics* 15: 195-243.
- Ko, Kwang-Mo. 1989. Explaining the noun-final change t>s in Korean. *Eoneohag* 11: 3-22.
- Martin, S. E. 1992. *A reference grammar of Korean*. Tokyo: Charles E. Tuttle.
- Mikheev, Andrei. 1997. Automatic rule induction for unknown-word guessing. *Computational Linguistics* 23: 405-423.
- Oh, Jaehyuk. 2006. *A study on sound changes in noun-final consonants* [cheonmal caumeui kyoche hyunsange taehan yeonkuw], Korea University, MA thesis.
- Oh, Jaehyuk and Jiyoung Shin. 2007. A Study on sound changes in noun-final consonants and collision of the phonetic realization. *Korean Linguistics* 34: 209-232.
- Park, Sunwoo. 2006. *Paradigm uniformity effects in Korean phonology*. PhD dissertation, Korea University, Seoul, Korea.
- Pierrehumbert, Janet 2001. Stochastic phonology. *GLOT* 5(6). 1-13.
- Pierrehumbert, Janet. 2003. Probabilistic phonology: discrimination and robustness. In *Probabilistic Linguistics*, ed. Bod et al., 177-228. Cambridge: The MIT Press.
- Pierrehumbert, Janet. 2006. The statistical basis of an unnatural alternation. In *Papers in*

- Laboratory Phonology* 8, ed. Louis Goldstein, Douglas Whalen and Catherine Best, 81-106. Berlin/New York: Mouton de Gruyter.
- Silverman, Daniel. (2009) Neutralization and anti-homophony in Korean. *Journal of Linguistics*. doi: 10.1017/S0022226709990247.
- Sohn, Ho-Min. 2001. *The Korean language* (Cambridge Language surveys) Cambridge: Cambridge University Press.
- Sohn, Hyang-Sook. 2001. Optimization of word-final coronals in Korean loanword adaptation. In *Proceedings of HILP 5*, ed. Caroline Fery, Antony Dubach Green, and Ruben van de Vijver, 159-177. Potsdam: University of Potsdam.
- Steriade, Donca. 2000. Paradigm uniformity and the phonetics-phonology boundary. In *Papers in Laboratory Phonology 5*, ed. Michael Broe and Janet Pierrehumbert, 313-334. Cambridge: Cambridge University Press.
- You, Jaewon. 1985. A reverse dictionary of modern Korean. Seoul: Chengumsa.
- Zuraw, Kie. 2000. *Patterned exceptions in phonology*. PhD dissertation, UCLA.
- Zuraw, Kie. 2002. Aggressive reduplication. *Phonology* 19.3: 395-539.
- Zuraw, Kie. 2007. The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog infixation. *Language* 83: 277-316.