

# *Korean n-insertion: a mismatch between data and learning\**

**Jongho Jun**

Seoul National University

---

This paper explores the variable application of n-insertion in Korean. Several tendencies emerge from the distribution of n-insertion in existing Korean words, using data drawn from a dictionary and from surveys and experimental studies. Most, but not all, of these tendencies are mirrored in the results of experiments involving novel words, suggesting that Korean speakers are aware of the differential influence of certain phonological factors on the probability of the application of n-insertion. The observed patterns of variation are analysed within the framework of a probabilistic version of Optimality Theory. In addition, the paper shows that the main aspects of n-insertion are motivated by the requirement of perceptually minimal modification, employing an analysis based on Steriade's P-map theory. The observed data *vs.* learning mismatch is also attributed to the lack of a perceptual basis for the tendency in question.

---

## 1 Introduction

Some phonological processes apply variably. The rate of application of such variable processes is often influenced by phonological factors similar to those which condition the occurrence of corresponding categorical processes (Kiparsky 1993, Anttila 1997, Zuraw 2000, 2002, 2010, Hayes & Londe 2006, Hayes *et al.* 2009). For instance, optional t/d-deletion in English is more likely to apply before a consonant (e.g. *west side*) than before a vowel (e.g. *west end*) (e.g. Guy 1980, Kiparsky 1993, Coetzee 2004). In variable nasal substitution in Tagalog, a prefix-final nasal is more likely to fuse with a voiceless obstruent than a voiced one

\* E-mail: [JONGHO@SNU.AC.KR](mailto:JONGHO@SNU.AC.KR).

I would like to thank Donca Steriade, Kie Zuraw and the anonymous reviewers for their suggestions and detailed comments. This paper has greatly benefited from advice and comments from many people. I am grateful to Adam Albright, Edward Flemming, Michael Kenstowicz, Eonsuk Ko, Heejeong Ko, Minkyung Lee, Sunwoo Park, participants in my *Phonology II* class at Seoul National University, audiences at the winter meeting of the Society of Modern Grammar, Daegu University, Kyongsan (February 2014) and at *NELS 45* (November 2014). I also thank Nayoung Park for her valuable assistance in survey data collection. Finally, I thank the participants in my experiment for their time.

(Zuraw 2010). It is typically the case in phonological research that such tendencies are initially observed in existing words, and then tested on loanwords or nonce words. If similar tendencies are mirrored in such novel words, it can be confirmed that they are part of speakers' internalised grammars. As discussed by Coetzee & Pater (2011), various probabilistic constraint-based theories and their related learning models have been proposed to explain variable phonological patterns and their acquisition (see e.g. Boersma 1997, Boersma & Hayes 2001, Goldwater & Johnson 2003, Boersma & Pater *forthcoming*).

In order to establish whether speakers learn all the tendencies which can be observed in existing words of their native language, the present study investigates variable n-insertion in Seoul Korean: /n/ is optionally inserted after a syllable-final consonant and before a high front vocoid /i j/ of the following syllable (e.g. /com-jak/ [comjak] ~ [comnjak] 'mothball'). Based on data from three different sources, a dictionary, previous surveys and experiments, and my own survey, I will show that the rate of n-insertion in existing Seoul Korean words is influenced by a number of phonological factors. The main tendencies found are that n-insertion is more likely before the glide /j/ than before the vowel /i/, and less likely after an obstruent or velar nasal than after other sonorant consonants. In addition, n-insertion is more likely when /j/ is followed by a high vowel, and when the morpheme preceding the insertion site is longer. Results of a wug test show that most of the tendencies observed from the existing words are mirrored in Seoul Korean speakers' production of and intuitions about novel words, suggesting that they are aware of these tendencies. However, the high rate of insertion for longer morphemes preceding the insertion site is not mirrored in novel words. This data *vs.* learning mismatch suggests that speakers do not internalise all the statistically prominent patterns of a language (Hayes *et al.* 2009, Becker *et al.* 2011, Hayes & White 2013).

An additional goal of the present study is to develop a plausible formal account of Korean n-insertion, relying crucially on P-map theory (Steriade 2001, 2009). I will adopt optimality-theoretic (OT) constraints (Prince & Smolensky 1993, McCarthy & Prince 1995) to account for the occurrence of n-insertion and the tendencies observed. Characterising probabilistic patterns of n-insertion, the constraints are assigned numerical weights within the framework of maximum entropy (maxent) grammar (Goldwater & Johnson 2003, Hayes & Wilson 2008, Hayes & White 2013). I conducted a learning simulation to demonstrate that the grammar posited for the analysis of n-insertion can be learned from the Korean data, and that the grammar learned can in fact explain the observed probabilistic patterns. Analysing the n-insertion data will show that the main aspects of Korean n-insertion are motivated by the requirement of perceptually minimal modification, and can be accounted for within Steriade's P-map theory. Specifically, /n/ is perceptually weak in the insertion site, and thus confusable with zero, and even more so in the context where n-insertion is likely to take place. This context-

dependent perceptibility is reflected in a fixed ranking (or relative weights) of relevant faithfulness constraints. The observed data *vs.* learning mismatch will also be attributed to the lack of a perceptual basis for the tendency related to the length of the morpheme preceding the insertion site.

This paper is organised as follows. In the rest of this section, I first describe the basic characteristics of Seoul Korean *n*-insertion and then introduce three phonological tendencies involved in the application of *n*-insertion, discussing Hwang's (2008) experiments involving novel words. §2 shows that these tendencies, along with others, are observed in existing Seoul Korean words, based on the distribution of *n*-insertion found in the data from my own survey, a dictionary and a number of previous studies. §3 presents a basic analysis of *n*-insertion employing OT constraints, while illustrating how characteristic patterns of *n*-insertion can be understood within P-map theory. §4 provides a learning simulation using a maxent learner. In §5, I show that the results of an experiment involving novel words are relatively well matched with existing word patterns in crucial respects, supporting the view that Seoul Korea speakers have stochastic phonological knowledge that originates from the patterns of existing words. §6 discusses how the aims of the present study are achieved, and what problems remain. §7 concludes and summarises this paper.

### 1.1 The basic pattern

In a dialect of Korean spoken in the Seoul-Gyeonggi area (henceforth Seoul Korean), /n/ is inserted at the juncture of two morphemes,  $M_1$  and  $M_2$ , when  $M_1$  ends in a consonant and  $M_2$  begins in a high front vocoid /i j/, as summarised in (1), where  $C_1$  is a  $M_1$ -final consonant.

(1) *n*-insertion in Seoul Korean

$$\emptyset \rightarrow n / C_1)_{M_1} \text{---} M_2(i/j)$$

This phenomenon has been much discussed in the literature on Korean phonology and morphology (Kim 1970, Choi 1971, Kim-Renaud 1974, Im 1981, Huh 1984 and many others).

$M_1$  and  $M_2$  may form an affixed word, compound or syntactic phrase, as shown in (2).<sup>1</sup>

<sup>1</sup> Intervocalic stops in Korean are voiced by intersonorant voicing. Thus a more accurate phonetic transcription of /ipul/ would be [ibul]. In addition, the liquid in Korean is realised as a lateral [l] except when it is a single onset, which is realised as a flap, transcribed as [r] in the traditional literature. These allophonic variations, as well as the palatality of the epenthetic /n/ shown in (3), are not reflected in the transcriptions in the remainder of this paper unless they are relevant to the discussion at hand. In contrast, the representations include forms derived by other regular processes such as obstruent nasalisation, shown in (4) below, and coda neutralisation, where coda obstruents are neutralised to homorganic unreleased lenis stops. Finally, /c/, /<sup>h</sup>/ and /' / stand for the coronal affricate, aspirated and glottalised (or tense) respectively.

(2) a. *Pre-|j| insertion*

<i>prefix-stem</i>	/təs-jaŋmal/	[tənnjaŋmal]	‘anklet socks’
<i>compound</i>	/com-jak/	[comŋjak]	‘mothball’
<i>phrase</i>	/mæk-in#jəs/	[mækinnjət]	‘taffy that (someone) ate’

b. *Pre-|i| insertion*

<i>prefix-stem</i>	/hot <sup>h</sup> -ipul/	[honnipul]	‘unlined comforter’
<i>compound</i>	/som-ipul/	[sommipul]	‘cotton sheet’
<i>phrase</i>	/han#il/	[hannil]	‘thing that (someone) has done’

As illustrated in (3), the epenthetic consonant /n/ is phonetically realised as a palatalised coronal sonorant, [ɲ] or, after a lateral, [ɺ], due to allophonic palatalisation (accompanied by lateralisation when C<sub>1</sub> is a liquid).<sup>2</sup>

## (3) a.

	<i>n-insertion</i>	<i>palatalisation</i>	
/com-jak/	[com.njak]	[com.ɲ(j)ak]	‘mothball’
/som-ipul/	[som.ni.pul]	[som.ɲi.pul]	‘cotton sheet’

## b.

	<i>n-insertion</i>	<i>lateralisation</i>	<i>palatalisation</i>	
/al-jak/	[al.njak]	[al.ljak]	[aɺ.ɺ(j)ak]	‘pill’
/sol-ip <sup>h</sup> /	[sol.nip]	[sol.lip]	[soɺ.ɺip]	‘pine needles’

With n-insertion, a non-liquid C<sub>1</sub> is typically realised as a nasal, due to the nasalisation of an obstruent before a nasal.

## (4)

	<i>n-insertion</i>	<i>obstruent nasalisation</i>	
/kiəp-jesan/	[ki.əp.nje.san]	[ki.əm.nje.san]	‘corporation budget’
/pat <sup>h</sup> -ilaŋ/	[pat.ni.laŋ]	[pən.ni.laŋ]	‘ridge of a field’

When C<sub>1</sub> is a liquid underlyingly, it is realised as a lateral, as can be seen in (3b).

## 1.2 Variation

In addition to the conditions mentioned above on the application of n-insertion, it has been argued in the literature (Huh 1984: 267, Ki 1990: 130–131, Ko 1992: 32, Han 1993: 124–125, Kim *et al.* 2002: 41, 44, Hong 2006: 400) that n-insertion occurs only when M<sub>2</sub> is a stem or root which can be an independent word. There are lexical exceptions (see e.g. Ko 1992: 32, Kim *et al.* 2002: 46–53, Bae 2003: 241–242). Not all /i j/-initial M<sub>2</sub> stems trigger n-insertion, as shown in (5), where SE = sentence ender (from Ko 1992: 32).

<sup>2</sup> With palatalisation, the triggering palatal glide is completely deleted or significantly weakened (O. Kang 2003: 460, Lee & Lee 2006).

(5) *Exceptional blocker*

/mas-is'-ta/	[masitt'a]	*[mannitt'a]	'delicious'
taste-exist-SE			
/mæs-is'-ta/	[mæsitt'a]	*[männitta']	'stylish'
style-exist-SE			

On the other hand, certain /j/-initial M<sub>2</sub> suffixes may trigger n-insertion (Ko 1992: 37, Kim *et al.* 2002).

(6) *Exceptional triggers*a. *Sino-Korean suffixes*

/pisaŋ-jon/	[pisaŋnjon]	'for emergency'
/kwancəl-jəm/	[kwancəlljəm]	'arthritis'

b. *Native Korean polite sentence ender /-jo/*

/camsi-man-jo/	[camsimannjo]	'Wait a moment.'
/kiləm-jo/	[kiləmnjo]	'of course'

More importantly, it is usually the case that the application of n-insertion is not obligatory (Kim-Renaud 1974: 150, Ko 1992: 32, Kwack 1992: 84, Han 1993: 124–125, Lee 1996: 167–171, Hong 1997: 175, Kim *et al.* 2002: 46, Bae 2003: 240–243, Oh 2006: 119, Ahn 2009: 279), as shown in (7) (from Kwack 1992, Ko 1992, Han 1993).

(7) /sək-jak/	[səkjak]	~ [sɛŋjak]	'colour weakness'
/pɛk-jəl/	[pɛkjəl]	~ [pɛŋjəl]	'white heat'
/hot <sup>h</sup> -ipul/	[hotipul]	~ [honnipul]	'unlined comforter'
/pat <sup>h</sup> -ilaŋ/	[patilaŋ]	~ [pannilaŋ]	'rows in a field'
/mos#icə/	[moticə]	~ [monnicə]	'cannot forget'

The probability of n-insertion varies across speakers and words, as will become clear below.

The present study examines the factors influencing the rate of this optional n-insertion. The next section discusses Hwang's (2008) productivity test, the results of which reveal that the frequency of n-insertion differs crucially depending on the type of C<sub>1</sub> consonant and following vocoids.

### 1.3 Tendencies in novel words

Hwang (2008) carried out a production experiment on n-insertion in Seoul Korean. Experimental stimuli were composed of two morphemes, M<sub>1</sub> and M<sub>2</sub> (e.g. /hiphap + junip<sup>h</sup>om/ 'hiphop uniform'), as shown in (8). M<sub>1</sub> was a loanword ending in one of the consonants /m n ŋ l p s k/. M<sub>2</sub> was a loanword or nonce word beginning with /i/ or /j/.

(8) *Experimental stimuli*

M <sub>1</sub>	<i>Loanwords</i> /sillim/ 'slim' /khillin/ 'clean' /wək <sup>h</sup> iŋ/ 'working' /simol/ 'small' /hiphap/ 'hiphop' /siwis/ 'sweet' /pillek/ 'black'	M <sub>2</sub>	<i>Loanwords</i> /iəp <sup>h</sup> on/ 'earphone' /junip <sup>h</sup> om/ 'uniform'  <i>Nonce words</i> /icipa/ /jucenol/
----------------	--	----------------	---

Twenty Seoul Korean speakers were asked to read aloud the written stimuli in a sentence. The occurrence of n-insertion was judged based on the author's acoustic analysis and three listeners' perception of the recordings of the subjects' productions. The results are summarised in Table I. Insertion rates differed, depending on the M<sub>2</sub>-initial vocoid. Almost no insertion was found before /i/: only 0.7% insertion (2 out of 280 production outputs). In contrast, n-insertion applied frequently before /j/: 49.2% (124 out of 252). In addition, insertion rates before /j/ differed according to the M<sub>1</sub>-final C<sub>1</sub>: /n m/ > /l/ > /k/ > /p s/ > /ŋ/. Notice that the probability of n-insertion is much lower after the obstruents and the velar nasal than after the other sonorants. These C<sub>1</sub>-dependent differences can be better seen in Table II.

		C <sub>1</sub>						
		m	n	l	p	s	k	ŋ
vocoid	i	0 (0/40)	2.5 (1/40)	2.5 (1/40)	0 (0/40)	0 (0/40)	0 (0/40)	0 (0/40)
	j	91.7 (33/36)	94.4 (34/36)	78.8 (26/33)	17.5 (7/40)	17.5 (7/40)	35.0 (14/40)	11.1 (3/27)

Table I

Results of Hwang's (2008) experiment with novel words, giving rate of insertion as a percentage of the number of inserted outputs divided by the total number of outputs. The total number in each cell varies because outputs were excluded from the analysis when the listeners' perceptual judgements failed to reach a certain level of consensus on the occurrence of n-insertion (see Hwang 2008 on the details of the judgement process.)

A Pearson's  $\chi^2$  test with Yates' continuity correction was performed to see whether these differences were statistically significant. Results showed that the relevant differences, i.e. son-obs ( $\chi^2 = 93.275$ ,  $p < 2.2e-16$ ) and son-ŋ ( $\chi^2 = 61.123$ ,  $p = 5.361e-15$ ), were significant.<sup>3</sup>

<sup>3</sup> In Table I, the differences between /m n/ and /l/ ( $\chi^2 = 4.687$ ,  $p = 0.095$ ) and between /p s/ and /k/ ( $\chi^2 = 4.565$ ,  $p = 0.102$ ) are not statistically significant.

son	88·57
obs	22·33
ŋ	11·11

Table II

Insertion rate (%) in Hwang's (2008) experiment, by C<sub>1</sub> type  
(son = sonorant except /ŋ/, obs = obstruent).

Thus the three phonological tendencies in (9) can be established from the results of Hwang's experiments on Seoul Korean n-insertion in novel words.

(9) *Phonological tendencies in novel words*

a. *Syllabicity effect*

n-insertion is more likely before /j/ than before /i/.

b. *Obstruency effect*

n-insertion is less likely after obstruents than after sonorants other than /ŋ/.

c. *Velar nasal effect*

n-insertion is less likely after /ŋ/ than after other sonorant consonants.

Where do these asymmetries come from? Why do Seoul Korean speakers apply n-insertion to novel words less frequently before /i/ and after obstruents and /ŋ/? One hypothesis based on recent research on phonological variation (Ernestus & Baayen 2003, Hayes & Londe 2006, Hayes *et al.* 2009, Zuraw 2010) is that Seoul Korean speakers use their knowledge about existing Korean words in the production of novel words. The same tendencies are expected to be found in Seoul Korean speakers' production of and intuitions about existing Korean words. The following section discusses the n-insertion patterns in existing Seoul Korean words.

## 2 Tendencies in existing words

In this section I investigate patterns of n-insertion in existing Seoul Korean words, based on data from the three sources mentioned above.

### 2.1 n-insertion in the dictionary

Seoul Korean is considered to be the standard dialect of Korean. Thus Korean dictionaries are primarily composed of Seoul Korean words. It is usually the case that the dictionary of a language contains many words that are unknown even to most native speakers of that language. Thus when we use dictionary data in linguistic research, we should exclude such words that are never or very rarely used in real speech situations.

For this reason, I constructed a dictionary database containing only words attested in the Sejong text corpus of 5.5 million words.<sup>4</sup>

I started with Kang & Kim's (2004) list of words attested in the Sejong text corpus (116,417 nouns, 5,253 verbs, 1,105 adjectives and 4,925 adverbs). From this list, I selected words meeting the minimal requirement for the application of n-insertion, i.e. those with orthographic sequences of a syllable-final consonant followed by /i/ or /j/ (4,997 in total).<sup>5</sup> Finally, words were excluded if they were (i) not listed in the *Standard Korean dictionary* (Kwuklip kwuke yenkwuwen 1999), or (ii) listed, but classified as non-standard dialectal or archaic forms in the same dictionary (henceforth *SKD*). The resulting database (3,239 words) contains Seoul Korean words which occurred at least once in the Sejong corpus and were also listed as standard Korean words in the *SKD*.

Each word in the database was encoded for the occurrence or absence of n-insertion, based on the pronunciation given in the *SKD*. Morphological complexity was additionally encoded on the basis of the criteria in (10).

(10) *Criteria for morphological complexity*

- a. Words were classified as multimorphemic if at least one of the following conditions was met:
  - i. Either  $M_1$  or  $M_2$  could occur as an independent word.
  - ii. There was a morpheme boundary marker between  $M_1$  and  $M_2$  in the *SKD*, regardless of whether  $M_2$  is a stem.
- b. Words were classified as monomorphemic if neither of the above conditions was met.

Table III shows the distribution of n-insertion in the resulting data according to morphological complexity.

This suggests that n-insertion applies only to multimorphemic words, with a very small number of exceptions; this is consistent with the conventional characterisation of Seoul Korean n-insertion discussed above. The reason that monomorphemic words have so little n-insertion is probably that if the word were typically pronounced with /n/, it would be spelled with the corresponding symbol.<sup>6</sup> In what follows, only multimorphemic words are considered.

<sup>4</sup> Available at <http://www.sejong.or.kr/>.

<sup>5</sup> In standard Korean orthography, the phonemic characters are grouped into blocks, each of which corresponds to a syllable, and thus basic syllable divisions can be seen in the written words.

<sup>6</sup> One might wonder how n-insertion could arise in monomorphemic contexts. Monomorphemic words with n-insertion are all Sino-Korean. Sino-Korean words consist of syllable-size units, each of which corresponds at least in etymology to a Chinese character with a particular meaning. Each such syllable-size unit is usually spelled as a separate syllable block in Korean orthography, regardless of syllable structure in the pronunciation. For instance, the Sino-Korean monomorphemic word /kəm.jəl/ 'censorship' is spelled <kəm.jəl>, reflecting its etymological structure, i.e. a combination of /kəm/ 'examine' and /jəl/ 'count'. In the spelling, the final consonant of the initial syllable is followed by a high front vocoid, the environment for optional n-insertion. For this word, the dictionary

	monomorphemic		multimorphemic	
insertion	0.05	1	23.21	305
variable	0.31	6	0.38	5
no insertion	99.64	1918	76.41	1004
<i>total</i>		1925		1314

Table III

Distribution of n-insertion (%) in the *SKD*, by morphological complexity. The number of words in each category is also shown. Variable forms are those which are listed in the dictionary both with and without insertion.

The distribution of n-insertion according to the type of  $M_2$ -initial vocoid in multimorphemic words is given in Table IV.

	/i/		/j/	
insertion	9.26	92	66.56	213
variable	0.20	2	0.94	3
no insertion	90.54	900	32.50	104
<i>total</i>		994		320

Table IV

Distribution of n-insertion (%) in the *SKD* in multimorphemic words, by  $M_2$ -initial vocoid type.

Fewer than 10% of words with /i/-initial  $M_2$  undergo n-insertion, as opposed to 67.5% of words with /j/-initial  $M_2$ . The difference between /i/ and /j/ is statistically significant (Pearson's  $\chi^2$  with Yates' continuity correction:  $\chi^2 = 456.957$ ,  $p < 2.2e-16$ ). Thus the syllabicity effect is confirmed.

The distribution of n-insertion according to  $C_1$  and  $M_2$ -initial vocoid types is presented in Table V. Words with a compound marker (traditionally called *sai-sios*) were excluded from this calculation, since it is not clear which category the compound marker belongs to. The compound marker should probably not be classified as either obstruent or sonorant, although the symbol for /s/ is used in standard Korean orthography.

For words with /i/-initial  $M_2$ , the ratio of n-inserted forms is higher when  $C_1$  is a sonorant other than /ŋ/ than when  $C_1$  is an obstruent or /ŋ/. Only the difference between /ŋ/ and the other sonorants is statistically significant (Pearson's  $\chi^2$  test with Yates' continuity correction:  $\chi^2 = 4.804$ ,  $p = 0.028$ ), confirming the velar nasal effect. However, for words with

---

shows two pronunciations, with and without n-insertion, i.e. [kəm.ŋjəl] and [kə.mjəl].

	/i/	/j/
son	11·51	65·33
obs	8·20	66·67
η	5·17	69·07

Table V

Insertion rate (%) in the *SKD* in multimorphemic words, by  $C_1$  and  $M_2$ -initial vocoid type.

/j/-initial  $M_2$  when  $C_1$  is a sonorant other than /η/, the ratio of n-inserted forms is lowest, and the relevant differences, i.e. son–obs and son–η, are not statistically significant. Consequently, the obstruency and velar nasal effects are not confirmed in words with /j/-initial  $M_2$ , where n-insertion is relatively frequent.

To summarise this section, only the syllabicity effect is clearly confirmed by the distribution of Seoul Korean words in the dictionary database, whereas the remaining two effects were not found, at least among words with /j/-initial  $M_2$ , the majority of which are subject to n-insertion. Only words with /i/-initial  $M_2$ , where n-insertion applies infrequently, show a statistically significant velar nasal effect. We should notice that most words in the *SKD* are given as forms with either insertion or non-insertion; only a very small number of words have both forms. Not surprisingly, variation is underrepresented in the *SKD*, as will become clear in the next section.

## 2.2 n-insertion in previous studies

Previous studies on n-insertion in Seoul Korean have employed a sizeable pool of native speakers, as shown in Table VI.

	subjects	test words
Choi (2002)	210	55
Kim (2003)	350	61
Kook <i>et al.</i> (2005)	209	182
<i>total</i>	769	268

Table VI

Numbers of subjects and test words employed in previous surveys and experimental studies. The total number of words is 268, rather than 298, because some words were used in more than one study. All the overlapping words appear in Kook *et al.* (2005).

Choi (2002) and Kim (2003) base their results on phonetic transcriptions of the speech of subjects who read experimental sentences. In order to construct a database, I took the number of their reported responses, which was roughly equal to the number of the subjects who produced inserted or non-inserted forms for each test word. The responses for some of the test words are shown in Table VII.

		insertion	no insertion	source
/twespak-ima/	'bulging forehead'	4	346	Kim
/pul-jənsok/	'discontinuity'	115	228	Kim
/kusok-jəŋcaŋ/	'arrest warrant'	91	119	Choi
/k'oc <sup>h</sup> -isil/	'flower dew'	165	185	Kim
/jəl-jətəl/	'eighteen'	331	19	Kim
/sanəp-jəŋ/	'industrial'	205	10	Choi

Table VII

Response frequencies for some of the test words employed by Choi (2002) and Kim (2003).

This simple data collection was not possible for the results reported by Kook *et al.* (2005), who consulted Seoul Korean speakers about their preferred pronunciations. They provide only percentages, not raw frequencies, for male and female subjects separately.

I first obtained the mean percentage insertion by averaging the percentages of male and female subjects, and then multiplied it by the total number of subjects, 209. This number was taken as the frequency of forms with insertion, and the result of subtracting it from the total number of subjects was taken as that of forms without insertion. The number of total responses for each test item is not always 209, since it is possible that subjects might have preferred multiple forms for some test items and failed to respond to others. But, with no better option available, I adopted this data-conversion method.

As indicated in Table VII, substantial variation can be observed in the data derived from the previous studies. In many cases, speakers did not agree on the occurrence (or absence) of n-insertion for a particular test word. More importantly, insertion rate (i.e. the relative usage frequency of the inserted form indicated by the number of inserted responses divided by the number of both inserted and non-inserted responses) varies greatly across words.

Let us now consider whether the variable n-insertion data from the previous studies reveal the three phonological tendencies under consideration, i.e. the syllabicity, obstruency and velar nasal effects. I calculated the insertion rates in Table VIII for different M<sub>2</sub>-initial vocoid types from the results of the previous studies.

	/i/	/j/
monomorphemic	<i>unavailable</i>	45·41 41
multimorphemic	44·71 90	71·20 137

*Table VIII*

Insertion rate (%) in previous studies, by M<sub>2</sub>-initial vocoid type. The number of words in each category is also shown.

Notice that all test words with /i/-initial M<sub>2</sub> are multimorphemic. Among the multimorphemic words, insertion rate is higher when the M<sub>2</sub>-initial vocoid is a glide than when it is a vowel. This difference is statistically significant (Pearson's  $\chi^2$  test with Yates' continuity correction:  $\chi^2 = 3597.215$ ,  $p < 2.2e-16$ ). Thus the syllabicity effect is confirmed.

In order to examine the obstruency and velar nasal effects, I now focus on multimorphemic words, since there are no monomorphemic words in the studies under consideration with C<sub>1</sub> obstruents, and thus they are not balanced with respect to C<sub>1</sub> type. In addition, for the reasons mentioned in §2.1, words containing a compound marker are excluded. Table IX shows insertion rates for different C<sub>1</sub> and M<sub>2</sub>-initial vocoid types.

	/i/	/j/
son	37·19	79·42
obs	53·82	68·60
ŋ	43·44	58·97

*Table IX*

Insertion rate (%) in previous studies, by C<sub>1</sub> and M<sub>2</sub>-initial vocoid types.

In Table IX the insertion rates for words with /j/-initial M<sub>2</sub> are consistent with the obstruency and velar nasal effects, since the rates are lower when C<sub>1</sub> is an obstruent or /ŋ/ than when it is a sonorant other than /ŋ/. But the insertion rates for words with /i/-initial M<sub>2</sub> are not at all consistent with the two effects, since the rate is lowest when C<sub>1</sub> is a sonorant consonant other than /ŋ/.

For a systematic investigation of the factors influencing the rate of n-insertion, I fitted a logistic regression model on the frequency data of multimorphemic words with /j/-initial M<sub>2</sub> with the glm function from the MASS package (Venables & Ripley 2002) in R (R Development Core Team 2014). The dependent variable is binary: n-inserted or not (reference level). The independent variables are the following factors: C<sub>1</sub>

type (son (reference), obs,  $\eta$ ),  $M_1$  syllable count ( $\sigma_1$ ),  $M_2$  syllable count ( $\sigma_2$ ),  $M_1$  origin (origin<sub>1</sub>: native (reference), sino, loan) and  $M_2$  origin (origin<sub>2</sub>: native (reference), sino).<sup>7</sup> The results are shown in Table X.

	estimate	SE	$p$ ( $> z $ )
(Intercept)	1.755	0.071	<0.001***
$C_1$ ( $\eta$ )	-1.054	0.036	<0.001***
$C_1$ (obs)	-0.594	0.032	<0.001***
$\sigma_1$	0.256	0.027	<0.001***
$\sigma_2$	-0.269	0.021	<0.001***
origin <sub>1</sub> (sino)	0.549	0.059	<0.001***
origin <sub>2</sub> (sino)	-0.869	0.063	<0.001***

Table X

Logistic regression results for data from previous studies: coefficients. Significance is indicated by \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ) and \*\*\* ( $p < 0.001$ ).

Positive and negative estimates indicate that the relevant factors encourage and discourage n-insertion respectively. All the factors adopted in the model affect insertion rate significantly. When  $C_1$  is an obstruent or  $/\eta/$  (as opposed to other sonorants), n-insertion is discouraged, confirming the obstruency and velar nasal effects. n-insertion is encouraged if what precedes the insertion site is longer, and discouraged if what follows it is longer. If what precedes the insertion site is Sino-Korean, n-insertion is encouraged, and if what follows it is Sino-Korean, it is discouraged.<sup>8</sup>

To summarise this section, the syllabicity effect is confirmed by the data from previous studies. The velar nasal and obstruency effects are also confirmed by the results for multimorphemic words with  $/j/$ -initial  $M_2$ , although they were not found for words with  $/i/$ -initial  $M_2$ . Since n-insertion frequently takes place before  $/j/$ , and its application before  $/i/$  is only a minor part of n-insertion, I will only consider words with  $/j/$ -initial  $M_2$  in exploring the velar nasal and obstruency effects. (Note that the results of both Hwang's 2008 experiments on novel words, discussed in §1.3, and my own, which will be presented in §5, show that n-insertion is

<sup>7</sup>  $M_1$  syllable count ( $\sigma_1$ ) and  $M_2$  syllable count ( $\sigma_2$ ) indicate the syllable count of each of the two parts of a word, i.e. before and after the syllable-initial  $/j/$ , regardless of its morphological composition. Thus, if a word consists of more than two morphemes,  $M_1$  or  $M_2$  may consist of multiple morphemes. NATIVE and SINO indicate native and Sino-Korean words respectively.

<sup>8</sup> These two lexical origin effects, i.e. higher insertion rate with Sino-Korean  $M_1$  and lower rate with Sino-Korean  $M_2$ , are consistent with Kim *et al.*'s (2002: 52–53) statement that n-insertion seems to be obligatory for words with Sino–native morpheme combinations but it is less so in words with native–Sino combinations. But, as will be seen in the next section, my own survey shows different results, with insertion rate being lower for words with Sino-Korean  $M_1$  than those with native Korean  $M_1$ .

unproductive before /i/ in Seoul Korean.) The syllabicity effect is confirmed by both sets of data, but the obstruency and velar nasal effects are confirmed only by the data from previous studies, not by the dictionary database. Notice that the results of the previous studies show that Seoul Korean speakers may apply n-insertion even to the words that have no inserted forms in the *SKD*, such as those in [Table VII](#) above, e.g. /pul-jansok/ and /kusok-jaŋcaŋ/. We can conclude that the occurrence of n-insertion is underreported in the *SKD*. Since the results of previous research do not display this problem, they should be considered more reliable than the *SKD*. However, the number of test words and phrases considered in these studies, ranging from 55 to 182, is not large enough to explore all the effects under consideration. Given that the words and phrases are complex and heterogeneous, involving the substantial number of factors listed in [Table X](#), the inventory used in the previous studies is too small. In order to ensure the presence and robustness of the velar nasal and obstruency effects, I therefore carried out a survey of Seoul Korean speakers' intuitions on n-insertion, employing a larger but more uniform set of words. Since it has become obvious that multimorphemic words with /j/-initial  $M_2$  frequently undergo n-insertion, I focused on them in the survey reported on in the next section.

### 2.3 A study of existing Korean words

I carried out a survey on n-insertion amongst 22 Seoul Korean speakers, employing 304 multimorphemic words with /j/-initial  $M_2$ .<sup>9</sup> Both inserted and non-inserted forms of the words were presented in standard Korean orthography as a questionnaire. For each test word, the participants were instructed to indicate their preferred pronunciation from options such as those in (11) for /com-jak/ and /<sup>h</sup>taŋ-jak/ 'herbal decoction'.

(11)		/com-jak/	/ <sup>h</sup> taŋ-jak/
a.	<i>insertion</i>	com.njak	<sup>h</sup> taŋ.njak
b.	<i>no insertion (resyllabification)</i>	co.mjak	—
c.	<i>no insertion (alignment)</i>	com.jak	<sup>h</sup> taŋ.jak

If  $C_1$  was /ŋ/, only two options, (a) and (c), were given, since there is no way to represent the second option in the standard Korean orthography, and it is generally assumed that /ŋ/ in onset position is prohibited in the phonology (see §3.2 for related discussion). Participants were allowed to give more than one possibility, and, if their preferred pronunciation was not given, they were asked to write it in a blank space on questionnaire.<sup>10</sup>

<sup>9</sup> This was the total number of words with /j/-initial  $M_2$  that I classified as multimorphemic in a pilot investigation of my dictionary database. The test words had the following  $C_1$  consonants: ŋ (93 words), n (78), l (42), m (18), k (49), p (15), s (3), h (2), c (1), p<sup>h</sup> (1), t<sup>h</sup> (1) and compound marker (1). After the survey was completed, I performed a more careful investigation, adding 16 words to the database.

<sup>10</sup> Most of the written responses involved a change in the part of the word which was not the focus of the study. For instance, for the test word /cəŋkwəŋ-jək/ 'desire for

One reason for conducting a self-evaluation survey using written forms is that both data collection and analysis are much easier and faster than in a production experiment. Hwang (2008: 14–16) states that it was often very difficult in the production experiment described in §1.3 to determine the presence or absence of n-insertion in the output purely on the basis of auditory and/or visual spectrogram inspection. This difficulty is mainly due to the acoustic and perceptual similarity of /j/ and /nj/. (In §3.3, I will consider some important aspects of Korean n-insertion, including the fact that, because of this similarity, /n/, rather than any other consonant, is inserted before a high front vocoid, but not other vocoids.) In addition, in Kang's (1997: 143–146) experimental study there was no crucial difference in the reported rates of a variable phonological process between the self-evaluation and production tests. The results of the survey reported here may be as reliable as those of a production experiment.

In the analysis of the results, I excluded the responses to test words with a compound marker in  $C_1$  position, for the reasons mentioned in §2.1. The total number of test words considered in the analysis is 303.

Like the studies discussed in §2.2, the results of the current survey show substantial variation across speakers and words. To establish whether obstruency and velar nasal effects were present, insertion rates were calculated for different  $C_1$  types, as shown in Table XI.

son		obs		ŋ	
48.77	138	41.14	72	38.59	93

Table XI

Insertion rate (%) in the current survey of existing Korean words, by  $C_1$  type.

Insertion rate was lower after an obstruent or /ŋ/ than after other sonorants. This difference is statistically significant, as established by a mixed effects logistic regression model analysis. The results of the present survey were fitted with the `lmer` function from the `lme4` package (Bates *et al.* 2011) in R (R Development Core Team 2014). The dependent variable was binary, i.e. n-inserted or not. Each subject and each test word was included as a random intercept. Fixed factors are as above, with the addition of  $M_2$ -initial vocoid (ja (reference), jə, je, jo, ju), and log-transformed token frequency in the Sejong corpus. The resulting fixed effects are shown in Table XII.

These results suggest that n-insertion is significantly discouraged when  $C_1$  is an obstruent or /ŋ/, as opposed to the other sonorants. Thus both the velar nasal and obstruency effects are confirmed.

---

power', some participants provided forms like [cəŋk'wənnjok], tensing the post-sonorant /k/. Such written responses were classified as an example of the corresponding category. Thus [cəŋk'wənnjok] was counted as a case of insertion.

	estimate	SE	<i>p</i> (>  <i>z</i>  )
(Intercept)	-1.482	0.560	0.008**
$C_1$ ( $\eta$ )	-0.494	0.208	0.017*
$C_1$ (obs)	-0.713	0.224	0.001**
jV (jə)	0.102	0.223	0.647
jV (je)	1.065	0.778	0.171
jV (jo)	0.230	0.291	0.427
jV (ju)	0.966	0.317	0.002**
$\sigma_1$	1.458	0.156	<0.001***
$\sigma_2$	-0.197	0.145	0.173
origin <sub>1</sub> (loan)	-0.420	0.806	0.601
origin <sub>1</sub> (sino)	-0.658	0.325	0.043*
origin <sub>2</sub> (sino)	-0.169	0.374	0.651
frequency	-0.099	0.088	0.260

Table XII

Mixed effects logistic regression results in the current survey of existing Korean words (fixed effects).

Some additional significant effects were found. First, n-insertion was significantly more frequent before /ju/ (i.e. when /j/ is followed by a high vowel) than before other vocoid sequences.<sup>11</sup> This asymmetry, which I will refer to as the HEIGHT EFFECT, can also be seen in Table XIII, where insertion rates are given for different vocoid sequences at the beginning of  $M_2$ .

ja	jə	je	jo	ju
39.83 73	42.45 149	47.25 4	44.30 43	57.56 34

Table XIII

Insertion rate (%) in the current survey of existing Korean words, by  $M_2$  initial vocoid sequence.

Second, n-insertion is significantly encouraged with a longer  $M_1$ , consistent with the results of previous work, as reported in Table X. More specifically, n-insertion is less likely with words with monosyllabic  $M_1$  than those with disyllabic  $M_1$ ; I will refer to this as the LENGTH EFFECT. The relevant insertion rates are shown in Table XIV.

<sup>11</sup> Notice that /u/ is the only high vowel which can follow /j/; the other high vowels, /i i/, do not occur in this environment.

monosyllabic	disyllabic	trisyllabic	quadrisyllabic	pentasyllabic
29·10 140	55·96 154	67·95 7	31·82 1	72·73 1

Table XIV

Insertion rate (%) in the current survey of existing Korean words, by  $M_1$  length.

In addition, n-insertion is discouraged when  $M_1$  is Sino-Korean, rather than native Korean. Although this lexical origin effect is statistically significant at the 0·05 level, the difference in insertion rate involved is very small (native Korean 45·96% *vs.* Sino-Korean 43·13%). Moreover, the corresponding results from previous studies given in Table X show the opposite tendency, i.e. a higher insertion rate with Sino-Korean  $M_1$ . On the assumption that the effect in question is at best minor, I will not consider the origin of  $M_1$  to be a significant factor in the application of n-insertion.

All other factors, i.e. the length and origin of  $M_2$  and token frequency, do not affect the rate of n-insertion significantly.

## 2.4 Summary

The significant tendencies in the distribution of n-insertion in existing Korean words are summarised in (12).

### (12) *Tendencies in Seoul Korean n-insertion*

#### a. *Syllabicity effect*

Insertion is more likely before /j/ than before /i/.

#### b. *Obstruency effect*

Insertion is less likely after an obstruent  $C_1$ .

#### c. *Velar nasal effect*

Insertion is less likely after a velar nasal  $C_1$ .

#### d. *Height effect*

Insertion is more likely before /j/ if it is followed by a high vowel.

#### e. *Length effect*

Insertion is more likely with a longer  $M_1$ .

The results of my own survey provide support for all these tendencies, with the exception of the syllabicity effect, which was not tested. The syllabicity effect was significant in my dictionary database as well as in previous studies, in which the height and length effects were also confirmed.<sup>12</sup> The remaining two effects, i.e. the obstruency and velar nasal effects, were supported by the previous studies, but not by the dictionary database.

<sup>12</sup> This is based on the results of a logistic regression analysis, not shown here, where the height of a vowel following /j/ and the syllable count of  $M_1$  are included as independent variables. The length effect in the data from previous studies can be seen in Table X.

I conclude that all five effects hold true for existing Seoul Korean words, on the assumption that the results of the present survey and previous studies are more reliable in detecting the obstruency and velar nasal effects than the dictionary database, for the reasons mentioned at the end of §2.2.

### 3 Analysis

As discussed by Coetzee & Pater (2011), variation data have been analysed in a number of constraint-based theories, e.g. Partially Ordered Constraints Theory (Kiparsky 1993, Anttila 1997), Stochastic OT (Boersma 1997, 1998, Boersma & Hayes 2001), Noisy Harmonic Grammar (Boersma & Pater *forthcoming*) and maxent grammar (Goldwater & Johnson 2003, Wilson 2006, Hayes & Wilson 2008). With the exception of Partially Ordered Constraints Theory, which has no learning model, all these OT and OT-like theories adopt constraints with a numerical value to encode probability distributions over variants. I analyse the observed patterns of Seoul Korean n-insertion using constraints which are assigned numerical weights. This section discusses the constraints which are responsible for the occurrence of Seoul Korean n-insertion and the observed asymmetric phonological tendencies. The specific weights of the constraints will be determined in the next section, which reports the results of a learning simulation using a maxent learner.

#### 3.1 What drives insertion?

When /n/ is not inserted,  $M_1$ -final  $C_1$  consonants can be either resyllabified into the onset of the  $M_2$ -initial syllable or aligned with the end of the  $M_1$ -final syllable, as shown in (13).

- (13)                      a. *resyllabification*                      b. *alignment*  
                               /com-jak/                      co.m) <sub>$M_1$</sub> jak                      com.) <sub>$M_1$</sub> jak  
                               /som-ipul/                      so.m) <sub>$M_1$</sub> i.pul                      som.) <sub>$M_1$</sub> i.pul

The resyllabified forms in (13a) have misaligned morpheme–syllable edges. More specifically,  $M_1$ -final consonants are not right-aligned with a syllable edge, violating the constraint in (14).

- (14) ALIGNMORPHEME-R (ALIGN-R)

The right edge of a morpheme coincides with the right edge of a syllable.

Alignment constraints of this type have been employed for the analysis of n-insertion by Shin (1997), Park (2005), Lee (2006), Lee & Lee (2006), Hwang (2008) and Ahn (2009). The idea that n-insertion occurs to prevent resyllabification is suggested by Kim-Renaud (1974: 148–149), Im (1981: 7) and Ki (1990: 85, §3.5).

Notice that ALIGN-R is active in Korean, independently of n-insertion. In Korean, an intervocalic consonant (with the possible exception of /ŋ/) is always syllabified as an onset, e.g. /aka/ [a.ka] ‘baby’. But when a morpheme boundary intervenes, especially in the case of compounding, a morpheme-final intervocalic consonant is variably syllabified between an onset and a coda, as shown in (15) (Park 2001: 730, citing Lee 1992: 42).<sup>13</sup>

- (15) /sikol-ai/ [si.ko.ra.i] ~ [si.kol.a.i] ‘country-bred child’  
 cf. /sikol/ ‘country’, /ai/ ‘child’  
 /mul-ori/ [mu.ro.ri] ~ [mul.o.ri] ‘wild duck’  
 cf. /mul/ ‘water’, /oli/ ‘duck’

Although resyllabification is in general preferred to alignment, as will be discussed in §4, aligned forms such as [si.kol.a.i] and [mul.o.ri] are possible. The occurrence of the aligned forms, i.e. blocking of resyllabification, can be attributed to ALIGN-R, regardless of whether there is an intervocalic consonant in an n-insertion context.

The aligned forms in (13b) satisfy ALIGN-R, but violate other constraints, including \*CODA (‘syllables do not have codas’). In addition, when the M<sub>2</sub>-initial vocoid is /i/, as in [som.i.pul], ONSET (‘syllables must have onsets’) is obviously violated. When the M<sub>2</sub>-initial vocoid is /j/, as in [com.jak], the violation of ONSET is less clear. If /j/ is syllabified in the onset, ONSET is satisfied, but if it is syllabified in the nucleus, ONSET is violated. Both syllabifications have been proposed in the literature on Korean phonology. Many previous studies (Kim-Renaud 1974, Kim 1986, Sohn 1987, Kang 1991, Kim & Kim 1991, Park 2001, K.-S. Kang 2003, Yun 2004) argue that glides in Korean are basically syllabified in the nucleus. However, at least some of them (Park 2001, K.-S. Kang 2003, Yun 2004) further argue, or assume, that, although glides in Korean are initially syllabified in the nucleus, they may move into the onset position, especially when it is empty.<sup>14</sup> /j/ in the aligned forms would then end up in onset position, violating the constraint in (16) (adapted from Rubach 2000: 274) banning onset vocoids.

(16) VOCOID-NUCLEUS (V-NUC)

Every [–consonantal] segment must be in the nucleus.

Based on the previous studies on Korean glides discussed above, I assume that both syllabifications of /j/ are possible.

Finally, forms with n-insertion such as [som.ni.pul] and [com.njak] would violate DEP(n) (‘no insertion of /n/’) and \*CODA. Table XV summarises which constraints would be violated by each of the possible output forms.

<sup>13</sup> As has been stated in the literature on Korean phonology (see e.g. Jun 1995, Shin *et al.* 2000), the occurrence of resyllabification can be identified on the basis of the allophonic variation of a singleton liquid, where a lateral [l] and a flap [ɾ] are syllabified in the coda and onset respectively.

<sup>14</sup> (18) below provides an example of a case in which /j/ in Korean behaves as an onset consonant, satisfying ONSET.

variant	/som-ipul/	/com-jak/	constraints violated
insertion	som.nipul	com.njak	DEP(n), *CODA
resyllabification	so.mipul	co.mjak	ALIGN-R
alignment	som.ipul	com.[ja] <sub>Nuc</sub> k com.j[a] <sub>Nuc</sub> k	ONSET, *CODA V-NUC, *CODA

*Table XV*

Possible variant forms and constraints violated.

Variable application of n-insertion can be explained by adopting variable ranking or similar weights of the above constraints. Notice that all these constraints (except for DEP(n), which will be discussed in detail below) are relatively well established independently of Korean n-insertion. Moreover, the above set of constraints is sufficient to explain how the main variants may occur. Thus, for the analysis of the basic pattern of Korean n-insertion, it is not necessary to adopt any additional constraints, such as a syllable-contact constraint prohibiting rising sonority over a syllable boundary. (For analyses of Korean n-insertion based on syllable contact, see Shin 1997, Chung 2001, Lee 2004 and Lee & Lee 2006.)

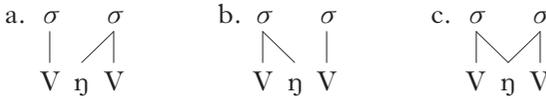
### 3.2 The obstruency and velar nasal effects

This section discusses the constraints responsible for the obstruency and velar nasal effects. The obstruency effect is due to automatic obstruent nasalisation in Korean, by which an obstruent becomes a nasal before a nasal. If n-insertion applied to words with obstruent  $C_1$ ,  $C_1$  would become a nasal (e.g. /hæk-jəlljo/ [hɛŋjəlljo] ‘nuclear fuel’), violating IDENT[son] (‘do not change the value of [sonorant]’). Since inserted forms of words with  $M_1$ -final sonorants (e.g. /com-jak/ [com.njak]) would not violate IDENT[son], other things being equal, n-insertion would be less likely after an obstruent than after a sonorant. Hwang (2008) adopts IDENT[son] for the same purpose.

Let us now consider how to explain the velar nasal effect. It has been assumed in the literature on Korean phonology (e.g. Kim 1986, Chung 2001, Hwang 2002, Park 2008, Kim 2011) that, unlike other consonants, /ŋ/ cannot initiate a syllable. This assumption, which can be captured by the constraint \*[ŋ] (‘no velar nasal in the onset’), is based on the fact that /ŋ/ can neither initiate a word nor follow a consonant word-medially. The absence of /#ŋV/ and /VCŋV/ sequences, along with the presence of /Vŋ#/ and /VŋCV/ sequences, suggests that /ŋ/ in Korean needs to be licensed in the coda. Thus it is generally accepted that a velar nasal occurring between vowels cannot be syllabified in the onset (17a). However, syllabifying it in the coda, as in (17b), which is widely accepted in the

literature, is not the only option for satisfying coda licensing. An intervocalic velar nasal can be ambisyllabic, as in (17c).<sup>15</sup>

(17) Possible syllable structures for intervocalic /ŋ/



The ambisyllabicity of Korean /ŋ/ gains support from an impressionistic description of the intervocalic velar nasal (Huh 1984: 208), which claims that when /ŋ/ occurs between vowels, it is difficult to determine whether it is the coda of a preceding syllable or the onset of a following syllable. Some phonetic and phonological facts in Korean show that /ŋ/ does not pattern with other unambiguous coda consonants, which is inconsistent with the prediction based on the structure in (17b).

The duration of Korean nasals varies greatly, depending on their position in the syllable. In general, they are much longer in the coda than in the onset. As shown in Table XVI, from Hwang (2002), labial and alveolar nasals occurring before a consonant, i.e. in coda position, have much longer durations than those occurring between vowels, i.e. in onset position.

	V__CV	V__V
/m/	179.12 (43.01)	66.27 (9.78)
/n/	170.77 (36.54)	80.22 (9.08)
/ŋ/	171.47 (38.10)	87.15 (28.74)

Table XVI

Mean durations (in ms) and standard deviations (in parentheses) of nasals in Korean (from Hwang 2002). C = lenis stop.

The duration of a velar nasal occurring in the V\_\_CV coda position is not significantly different from that of other nasal codas. Notice that intervocalic /ŋ/ is much shorter than both its preconsonantal counterpart and other nasals which are unambiguously in the coda. Thus intervocalic /ŋ/ can hardly be grouped with other coda nasals with respect to duration, casting doubt on its syllabification exclusively in the coda in (17b).

Hiatus-avoidance phenomena in suffixal allomorph selection in Korean provide additional evidence against the syllabification of intervocalic /ŋ/ in

<sup>15</sup> For ambisyllabic /ŋ/ to satisfy \*[ŋ], the constraint definition needs to be revised as follows: ‘the velar nasal is licensed in the coda’. See Itô (1986, 1989) and Itô & Mester (1994) for discussion of how multiply linked elements can satisfy positional licensing constraints.

the coda. Some suffixes have different allomorphs depending on whether the stems preceding them end in a consonant or a vowel, as illustrated in (18).

(18) *Phonologically conditioned suffixal allomorphs*

a. <i>nominative case marker</i>	-i / C __	son-i	'hand-NOM'
	-ka / V __	ch <sup>h</sup> a-ka	'car-NOM'
b. <i>vocative case marker</i>	-a / C __	tasom-a	'(name)-VOC'
	-ja / V __	kiho-ja	'(name)-VOC'

For stems ending in a vowel, the choice of the consonant-initial allomorphs /-ka/ and /-ja/ can be attributed to ONSET. If intervocalic /ŋ/ is syllabified in the coda, /ŋ/-final stems would be followed by consonant-initial allomorphs. As shown in (19), this is not the case.

(19) *Phonologically conditioned suffixal allomorphs for /ŋ/-final stems*

a. <i>nominative case marker</i>	-i	paŋ-i	'room-NOM'
b. <i>vocative case marker</i>	-a	salan-a	'love-VOC'

The choice of vowel-initial allomorphs for /ŋ/-final stems strongly suggests that ONSET is satisfied by a stem-final velar nasal, which is thus not syllabified exclusively in the coda.

Given the above phonetic and phonological facts, I consider intervocalic /ŋ/ in Korean to be ambisyllabic. This suggests that the constraint prohibiting an ambisyllabic velar nasal (\*AMBI-ŋ) has only a minor effect, compared to the corresponding constraint for other consonants (\*AMBI-C). M<sub>1</sub>-final /ŋ/ would then be ambisyllabic before M<sub>2</sub>-initial high front vowels, unless n-insertion applies. Given that the ambisyllabic /ŋ/ may satisfy both ONSET and ALIGN-R (see Itô & Mester 1994, 1999 on non-crisp alignment), words with /ŋ/ in C<sub>1</sub> would be likely to surface as such, resisting n-insertion. Accordingly, the velar nasal effect may arise.

As mentioned above, my analysis of Korean n-insertion is couched within a probabilistic OT theory. Specifically, I adopt a maxent grammar, where weighted constraints interact to assign probabilities to possible outputs. The specific weights and probabilities will be determined through the learning simulation presented in §4; the OT tableaux in (20) illustrate crucial points of the analysis of obstruency and velar nasal effects presented above. In each tableau, five crucial candidate forms are given, involving (i) insertion, (ii) resyllabification, (iii) alignment, with /j/ in the onset, (iv) alignment, with /j/ in the nucleus, and (v) ambisyllabicity (represented by {C}). Dominant, i.e. high-weighted, constraints are separated from the rest of the constraints by a solid line. Constraints with relatively similar weights are indicated by dashed vertical lines, and those which play no role in the comparison between forms with different C<sub>1</sub> consonants are shaded. The obstruency effect can be seen by comparing (20a.i) and (20b.i). The form with n-insertion with an underlying obstruent C<sub>1</sub> in (20b.i) violates IDENT[son], due to obstruent nasalisation, whereas that

with a sonorant C<sub>1</sub> in (20a.i) does not. Thus, other things being equal, the rate of occurrence of the former is lower than that of the latter.<sup>16</sup>

(20) a. 

/com-jak/	*[ŋ]	*AMBI-C	ID [son]	*AMBI-ŋ	ALIGN-R	ONS	V-NUC	*CODA	DEP (n)
i. com.njak								**	*
ii. co.mjak					*			*	
iii. com.jak							*	**	
iv. com.[ja] <sub>N</sub> k						*		**	
v. co{m}jak		*!						**	

b. 

/hɛk-jəllo/									
i. hɛŋ.njəl.ljo			*					**	*
ii. hɛ.kjəl.ljo					*			*	
iii. hɛk.jəl.ljo							*	**	
iv. hɛk.[jə] <sub>N</sub> l.ljo						*		**	
v. hɛ{k}jəl.ljo		*!						**	

c. 

/t <sup>h</sup> aŋ-jak/									
i. t <sup>h</sup> aŋ.njak								**	*
ii. t <sup>h</sup> a.ŋjak	*!				*			*	
iii. t <sup>h</sup> aŋ.jak							*	**	
iv. t <sup>h</sup> aŋ.[ja] <sub>N</sub> k						*		**	
v. t <sup>h</sup> a{ŋ}jak				*				**	

The analysis of the velar nasal effect is a little more complicated. There are two differences between the analysis of a velar nasal C<sub>1</sub> and that of any other sonorant C<sub>1</sub>, as can be seen by comparing (20a) and (20c). Since \*AMBI-ŋ, the constraint which bans ambisyllabic velar nasals, is low-weighted, but the corresponding constraint for other consonants is dominant, the candidate with ambisyllabic C<sub>1</sub> can compete with the n-insertion candidate only when C<sub>1</sub> is /ŋ/. Thus, consistent with the velar nasal effect, n-insertion is relatively less favoured when C<sub>1</sub> is /ŋ/ than when it is any other sonorant consonant. However, if C<sub>1</sub> is /ŋ/, the candidate with resyllabification (20c.ii) would never occur as an attested variant, due to dominant \*[ŋ]. The absence of this candidate would raise the relative frequency of the competing candidates, including the one with n-insertion. This is in

<sup>16</sup> In line with previous studies on Korean glides, I assume that /j/ is syllabified in the nucleus unless its movement into an onset is required by the ONSET constraint. Thus, of the two possible syllabifications of M<sub>2</sub>-initial /j/ in aligned forms such as [com.j[a]<sub>Nuc</sub>k] and [com.[ja]<sub>Nuc</sub>k], I consider the one with [j] in the onset to be the attested one. This is based on the fact in Korean that the presence of [j] can satisfy ONSET, as can be seen in glide formation, where stem-final /i/ becomes [j] to avoid hiatus in verbal inflection (e.g. /t'ɛli-ə/ [t'ɛ.rjə] \*[t'ɛ.ri.ə] 'hit-IMP') and in the vocative suffixal allomorphy illustrated in (18) (e.g. [kiho-ja], \*[kiho-a]).

conflict with the velar nasal effect. Consequently, the velar nasal effect can be obtained by keeping the weight of \*AMBI-ŋ low enough to cancel out the frequency/preference change caused by the absence of the form with resyllabification.

### 3.3 P-map effects

Why do Seoul Korean speakers insert a coronal nasal, rather than any other consonant? And, why is it inserted only before high front vocoids, not before other vocoids? I first give a brief introduction to Steriade's (2001, 2009) P-map theory, which proposes that it is segments with low perceptibility that are typically inserted or deleted, due to the fixed ranking of faithfulness constraints reflecting the relevant perceptibility scale. I then provide arguments for the perceptual weakness of the coronal nasal before high front vocoids in Korean. The syllabicity and height effects will also be explained in a similar way.

P-map theory is motivated by the claim that only perceptually tolerated modifications to input forms are accepted (Kohler 1990). A segment can be inserted or deleted only when its insertion or deletion is not perceptually noticeable. Specifically, the choice of an epenthetic segment is determined on the basis of a context-dependent hierarchy of perceptual similarity between input and output. The segment most confusable with zero in the relevant context is predicted to be the epenthetic segment. The reason for [ʔh] and the glides [j w] (adjacent to their homorganic vowels) being frequent epenthetic segments across languages is that they do not induce coarticulatory changes on neighbouring segments, and thus input–output pairs like V–ʔV, V–hV and V–GV (where G is homorganic with V) are the most similar. The choice of epenthetic [ʔ] can be illustrated by the P-map and ranking of DEP constraints in (21), where  $\Delta(a-b)$  = the perceptual difference between *a* and *b*.

- (21) a. *P-map*  
 $\Delta(\emptyset-p) / \_ V, \Delta(\emptyset-t) / \_ V, \Delta(\emptyset-k) / \_ V > \Delta(\emptyset-ʔ) / \_ V$   
 b. *Ranking*  
 $\text{DEP}(p,t,k) / \_ V \gg \text{DEP}(ʔ) / \_ V$

If /n/ occurs before high front vocoids, it undergoes allophonic palatalisation, failing to induce coarticulatory changes on the following vocoids. Thus, the input–output pairs i–ni (or, more precisely, i–ɲi) and j–nj (or j–ɲj) are perceptually similar pairs, just like V–GV, where G is homorganic with V. Since formant transitions from [ɲ] to [i j] are flat (as if there is no preceding consonant), the perceptual difference of the relevant pairs i–ni and j–nj ( $\Delta(\emptyset-n) / \_ \{i, j\}$ ) must be smaller than that of non-homorganic pairs involving different inserted segments like i–mi and j–mj ( $\Delta(\emptyset-m) / \_ \{i, j\}$ ) and of those involving different vocoids like a–na, e–na and w–nw ( $\Delta(\emptyset-n) / \_ \{a, e, w\}$ ). This suggests not only that insertion of /n/ before high front vocoids is perceptually less salient than insertion of segments like /p m/ in the same environment, but also that insertion of /n/ is

less salient before high front vocoids /i j/ than before other vocoids, such as /a e/ and /w/. In addition, /i j/ are high vocoids with a low-frequency resonance, and are acoustically like nasals.<sup>17</sup> Given that the insertion of a segment which is similar to its neighbouring segment is less noticeable than the insertion of a dissimilar segment, addition of /n/ before nasal-like high vocoids /i j/ would be perceptually less prominent than addition of a true oral segment such as /p t k/ before /i j/, and of /n/ before a non-high truly oral vocoid such as /a e o/.<sup>18</sup> In P-map theory, these perceptibility differences can be reflected in the fixed ranking (or relative weights) of the DEP constraints in (22), explaining why it is the coronal nasal that is inserted before high front vocoids.

$$(22) \left. \begin{array}{l} \text{DEP(p)} / \_ \{i,j\}, \text{DEP(m)} / \_ \{i,j\} \dots \\ \text{DEP(n)} / \_ a, \text{DEP(n)} / \_ e, \text{DEP(n)} / \_ w \dots \end{array} \right\} \gg \text{DEP(n)} / \_ \{i,j\}$$

DEP(n)/\_ {i, j} needs to be further parameterised for the explanation of the syllabicity and height effects. In Korean, unlike /ni/, the sequence /nj/ is phonetically realised almost as a single segment [ɲ] (O. Kang 2003, Lee & Lee 2006).<sup>19</sup> Thus the input–output pair j–nj is perceptually almost identical to j–ɲ, which must be more similar than i–ni, since the input–output difference is less than a single segment in the former, but the entire segment [ɲ] in the latter. This perceptibility difference ( $\Delta(\emptyset\text{-n})/\_ i > \Delta(\emptyset\text{-n})/\_ j$ ) can be incorporated into the ranking (or relative weights) of the relevant DEP constraints in (23), accounting for the syllabicity effect.

$$(23) \text{DEP(n)} / \_ i \gg \text{DEP(n)} / \_ j$$

Finally, the degree of perceptual modification involved in pre-/j/ n-insertion also differs depending on the height of the vowel following /j/. /j/ normally coarticulates with a following vowel. When /j/ is followed by a high vowel, its low-frequency resonance, which it shares with nasals, can be fully maintained, making it perceptually more similar in nasality to its corresponding output with n-insertion, i.e. [ɲ]. But when it is followed by a non-high vowel, which lacks resonance in the low-frequency range, /j/'s low-frequency resonance is not maintained for as long, making it less nasal-like, so that the relevant input–output pair is less similar. The height effect can be attributed to this perceptual similarity of the high vowel input–output pair, which determines the relative weights of the relevant DEP constraints in (24).

<sup>17</sup> ‘High vowels, such as /i/ and /u/, normally have a low frequency resonance and, therefore, are acoustically more like nasals’ (Borden & Harris 1984: 180). Since /j/ is spectrally very much like /i/, I assume that /j/ is also acoustically like nasals.

<sup>18</sup> Like the high front vocoids /i j/, the high back vocoids /u w/ are acoustically nasal-like, and thus, addition of a nasal /n/ would be probably less prominent before them than before non-high vocoids. But, when an alveolar nasal /n/ precedes /u w/, it would induce coarticulatory changes. Thus the input–output pairs u-nu and w-nw would not be so similar pairs as i-ni and j-nj.

<sup>19</sup> The deletion of /j/ is described as optional by O. Kang (2003: 460), but is assumed to be obligatory by Lee & Lee (2006).

(24) DEP(n) / \_\_j [V, -high]  $\gg$  DEP(n) / \_\_j [V, +high]

In summary, the choice of the epenthetic and contextual segments in n-insertion and the syllabicity and height effects can be analysed in terms of a context-dependent perceptibility hierarchy: /n/ is confusable with zero in the n-insertion context, and even more so in the context where n-insertion is highly frequent.

#### 4 Learning simulation

In this section, I show not only that the OT grammar presented in the previous section can be learned from the Korean data, but also that the learned grammar can in fact explain the observed asymmetric patterns of Korean n-insertion. Specifically, using a maxent learner implemented in the maxent grammar tool (Hayes 2009), I conducted a learning simulation. A maxent grammar (see Goldwater & Johnson 2003 for a detailed description) adopts constraint weights, like Harmonic Grammar (Smolensky 1986, Legendre *et al.* 1990), but, unlike standard OT, defines a probability distribution over the possible outputs. Constraint weights are determined to maximise the probability of the observed output forms. In the present learning simulation, weights of the constraints presented in the previous section were set on the basis of the training data, which were essentially composed of the results of my own survey on n-insertion. To test the insertion rates predicted, the grammar learned was applied to a test set of words which was the same as the training set.<sup>20</sup> The survey data had responses to words with /j/-initial M<sub>2</sub>, but not /i/-initial M<sub>2</sub>, and thus constraints responsible for the syllabicity effect (such as DEP(n)/\_\_i) were excluded from the simulation. For simplicity, the constraint prohibiting ambisyllabicity for consonants other than /ŋ/ (\*AMBI-C), which is obviously high-weighted in Korean, was also excluded.

The frequencies of inserted and non-inserted forms (resyllabified or aligned) were taken from the relevant distribution in the results of the current survey, shown in Table XVII.<sup>21</sup>

Notice that the grammar learned must reflect not just n-insertion patterns, but also the general phonology of Korean. It should not contradict what is known about Korean morphophonology. In order to incorporate the general resyllabification rate in Korean into the learning simulation, the training data had to include the resyllabified and aligned forms of words with /VC<sub>1</sub>+V<sub>2</sub>/, where V<sub>2</sub> is not a high front vocoid (e.g. /sikol-ai/).

One indication of the resyllabification rate in Korean can be found in Jun's (1995: 242, 251) experiment on Korean speakers' production of

<sup>20</sup> See the Appendix for the training items which were given to the learner. The testing items were exactly same as the training items, except that all their output frequencies were zero.

<sup>21</sup> As discussed in §3.2, when C<sub>1</sub> is /ŋ/, it was considered as ambisyllabic in all non-inserted responses.

vocoid type	C <sub>1</sub> type	total	response type	n	%
j [V, -high]	son	2763	insertion	1295	46·87
			resyllabification	935	33·84
			alignment	533	19·29
	obs	1397	insertion	531	38·01
			resyllabification	595	42·59
			alignment	271	19·40
ŋ	1875	insertion	715	38·13	
		non-insertion	1160	61·87	
j [V, +high]	son	337	insertion	217	64·39
			resyllabification	31	9·20
			alignment	89	26·41
	obs	200	insertion	126	63·00
			resyllabification	28	14·00
			alignment	46	23·00
ŋ	224	insertion	95	42·41	
		non-insertion	129	57·59	

Table XVII

Training data for Korean n-insertion.

word-final /l/ before a vowel-initial word. She reports that the flap, which can be interpreted as a resyllabified /l/, occurs in 65%–98% of cases across an Accentual Phrase boundary and 70%–100% within an Accentual Phrase. Assuming that most, if not all, of the words in the training data are within an Accentual Phrase, I took the relevant average value, i.e. 85%, as the rate of resyllabification for the words with /VC<sub>1</sub>+V<sub>2</sub>/ (V<sub>2</sub> ≠ i/j) in the training data. In the learning and training data, words of this type (e.g. /sikol-ai/), where M<sub>2</sub>-initial vocoids can be any vowels and glides other than /i j/, should be much more frequent than those shown in Table XVII, where the M<sub>2</sub>-initial vocoid is always /j/. Thus I simply multiplied the total number of survey responses, i.e. 6,796, by 10, and added the resulting number of words (67,960) with /VC<sub>1</sub>+V<sub>2</sub>/ to the training data: of these, 85% (57,766) were resyllabified (e.g. [si.ko.ra.i]), whereas 15% (10,194) were aligned (e.g. [si.kol.a.i]).

The complete training data given to the learner can be seen in the Appendix. The learner was run on the training set for 1,000,000 cycles, with the default setting for the preferred weight value of the constraint, and with no fixed rankings (or, more precisely, relative weights), including those discussed in the previous section. The resulting constraint weights are presented in Table XVIII.

constraint	weight
*[ŋ]	14.36
ONSET	3.01
ALIGN-R	1.65
V-NUC	1.57
DEP(n) / __j [V, -high]	0.64
*CODA	0.41
IDENT[son]	0.41
*AMBI-ŋ	0.02
DEP(n) / __j [V, +high]	0.00

*Table XVIII*

Constraint weights returned by the learner.

Let us consider how these learned weights can explain the patterns under consideration. Recall that /VC+V/ sequences in Korean are syllabified as [V.CV], violating ALIGN-R, or [VC.V], violating ONSET and \*CODA. The general preference for the former can be captured by the weights of the relevant constraints: the sum of the weights of ONSET and \*CODA exceeds the weight of ALIGN-R. The height effect (i.e. n-insertion is less likely with a non-high vowel following /j/) can be explained by the greater weight of DEP(n) / \_\_j [V, -high] in comparison with DEP(n) / \_\_j [V, +high]. The velar nasal effect can be captured by the very low weight of \*AMBI-ŋ, since the velar nasal in C<sub>1</sub>, without n-insertion, would be likely to be ambisyllabic, satisfying both ONSET and ALIGN-R. Finally, the obstruency effect can be explained by the non-zero weight of IDENT[son], which is violated only when n-insertion takes place in words with an obstruent C<sub>1</sub>.

Table XIX shows the encoded probability distributions (which can be considered as the prediction of the grammar learned) for each combination of C<sub>1</sub> type and vowel height, along with the corresponding observed proportion of inserted forms (i.e. the input to the simulation).

The maxent model successfully reproduces the relative frequencies reflecting the crucial effects, i.e. obstruency, velar nasal and height, confirming the validity of the proposed analysis.

M <sub>2</sub> -initial sequence	j [V, -high]			j [V, +high]		
	son	obs	ŋ	son	obs	ŋ
input (observed)	46.87	38.01	38.13	64.39	63.00	42.41
prediction (learned)	49.18	39.17	33.87	64.67	54.91	49.21

*Table XIX*

Insertion rates (%) in the learning simulation: comparison of the survey data (observed) with the maxent model (learned).

Finally, note that the P-map-based fixed rankings, or relative weights, of the constraints discussed in the previous section were not employed in the present simulation. Nonetheless, the constraint weights in Table XVIII are consistent with them. This is probably because the distribution of n-insertion in existing Korean words already reflects the requirement of minimal perceptual modification.

## 5 Speakers' knowledge of n-insertion tendencies

In this section, I address the question of whether and how much the learned grammar presented in the previous section does in fact reflect Korean speakers' mental grammar. Specifically, I conducted an experiment on n-insertion in Seoul Korean, using novel words, to find out whether speakers are aware of the tendencies observed in existing words and, if so, which. This wug test is basically a replication of Hwang's (2008) productivity test, with the addition of two factors which turned out to influence the insertion rate of existing Seoul Korean words in my own survey: (i) length of  $M_1$ , and (ii) height of the vowel following  $M_2$ -initial /j/.

All test words consisted of two morphemes, a loanword  $M_1$  and a wug stem  $M_2$ .<sup>22</sup>  $M_1$ , which was either monosyllabic or disyllabic, ended in one of seven consonants, /m n ŋ l p s k/, as shown in (25).  $M_2$  began with one of /i ju ja/. The total number of test items was 84 (2 (syllable count) × 7 (coda) × 3 (vocoid type) × 2 (repeating blocks)). The same number of control items (with vowel-final  $M_1$  or /a e/-initial  $M_2$ ) was employed. In the test form, no two words which were similar in some crucial aspect were adjacent. For instance, no two test items were shown successively, but were always separated by a control item.

### (25) *Experimental tokens*

$M_1$	<i>Test</i>		<i>Control</i>
	<i>monosyllabic</i>	<i>disyllabic</i>	
	/s'əm/ 'some'	/sillim/ 'slim'	/hɛp <sup>h</sup> i/ 'happy'
	/t <sup>h</sup> en/ 'ten'	/k <sup>h</sup> illin/ 'clean'	/jellou/ 'yellow'
	/k <sup>h</sup> iŋ/ 'king'	/wək <sup>h</sup> iŋ/ 'working'	/pillu/ 'blue'
	/t <sup>h</sup> ol/ 'tall'	/simol/ 'small'	/silpə/ 'silver'
	/t <sup>h</sup> ap/ 'top'	/hiphap/ 'hiphop'	/lɛti/ 'red'
	/has/ 'hot'	/siwis/ 'sweet'	
	/k <sup>h</sup> ik/ 'kick'	/pillek/ 'black'	
$M_2$	<i>Test</i>	<i>Control</i>	
	/icipa/	/acepa/	
	/jucenol/	/ekenol/	
	/jacenal/		

<sup>22</sup> Where possible, I used Hwang's (2008) test and control items in the present experiment. When no word was available from Hwang's list, I either used different items or created new ones. Specifically, all disyllabic loanword  $M_1$  and /ju i/-initial wug stems are from Hwang's lists, but all monosyllabic loanword  $M_1$  and /ja a e/-initial wug stems were added.

Two orders were prepared: one set was the reversed version of the other. 37 speakers participated in the test, none of whom had participated in the survey of existing Seoul Korean words discussed in §2.3. 24 subjects were tested on one set and 13 on the other. In the test form, both inserted and non-inserted forms for each word were presented, in standard Korean orthography. The experimenter told the participants that the words were made-up compound nouns for new chemical products. The participants were instructed to choose a pronunciation of each of the given compounds, from options such as those in (26).

(26)		/sʰəm+jucenol/	/kʰiŋ+jucenol/
a. <i>insertion</i>		sʰəm.nju.ce.nol	kʰiŋ.nju.ce.nol
b. <i>no insertion (resyllabification)</i>		sʰə.mju.ce.nol	–
c. <i>no insertion (alignment)</i>		sʰəm.ju.ce.nol	kʰiŋ.ju.ce.nol

As in the survey of existing words, when  $C_1$  was /ŋ/, only options (a) and (c) were given. Participants were allowed to give more than one possibility, and, if their preferred pronunciation was not given, they were asked to write it in.

In the analysis of the test results, I excluded the responses of two subjects in which n-insertion was observed even more frequently among control items than test items, along with those of three subjects in which no insertion was indicated at all. I analysed the responses of the remaining 32 subjects.

The insertion rates according to  $C_1$  type were calculated over the responses to the test words, i.e. those with /i ju ja/-initial  $M_2$ , as shown in Table XX.

son	obs	ŋ
30·55	18·15	14·91

Table XX

Insertion rate (%) in the wug test, by  $C_1$  type.

Insertion rate was lower after obstruents and /ŋ/ than after other sonorant consonants. This difference is statistically significant, as can be seen in a mixed effects logistic regression model analysis.

As before, the results of the present test were fitted with the lmer function from the lme4 package in R. The dependent variable is binary, inserted or not (reference). The independent variables are:  $C_1$  type (son (reference), obs, ŋ),  $M_2$ -initial vocoid type (V: ja (reference), ju, i) and  $M_1$  syllable count ( $\sigma_1$ ). Random intercepts were set for item and subject.<sup>23</sup>

<sup>23</sup> More complex models, with interactions or random slopes, failed to converge. See note 24 for the results of a regression model with interactions included but random effects excluded.

The results are shown in Table XXI.

	estimate	SE	<i>p</i> (>  <i>z</i>  )
(Intercept)	-0.115	0.476	0.809
C <sub>1</sub> (η)	-1.412	0.378	<0.001***
C <sub>1</sub> (obs)	-1.072	0.257	<0.001***
V (i)	-2.860	0.340	<0.001***
V (ju)	0.553	0.267	0.038*
σ <sub>1</sub>	-0.346	0.237	0.144

Table XXI

Mixed effects logistic regression results for the wug test.

These results suggest that n-insertion is significantly discouraged when C<sub>1</sub> is an obstruent or /ŋ/ (as opposed to the other sonorants), and the M<sub>2</sub>-initial vocoid is /i/ (as opposed to /ja/). In addition, n-insertion is significantly encouraged when /j/ is followed by a high vowel. Thus all three main effects, syllabicity, obstruency and velar nasal, are confirmed, as well as the height effect. The lower insertion rate before /i/ can also be seen in Table XXII, where the insertion rates for /i/ and the control vowels /a e/ at the beginning of M<sub>2</sub> are almost identical, suggesting that pre-/i/ n-insertion is unproductive.

	ju	ja	i	a	e
test	36.32	27.59	5.00	<i>n/a</i>	<i>n/a</i>
test and control	30.23	22.79	3.75	4.69	5.68

Table XXII

Insertion rate (%) in the wug test, by M<sub>2</sub> initial vocoid sequence (*n/a* indicates that an item was not available).

Insertion rate is higher with monosyllabic M<sub>1</sub> (25.20; cf. disyllabic M<sub>1</sub> = 20.95).<sup>24</sup> This is the opposite of the tendency in existing words, as can be seen from the results of the current existing word survey and the data from previous studies. Recall that in the results of the current survey on existing Korean words, the insertion rates of monosyllabic and disyllabic M<sub>1</sub> were 29.10 and 55.96 respectively. This suggests that the speakers failed to learn the tendency involving the M<sub>1</sub> length of existing words.

<sup>24</sup> This difference was not significant in the mixed effects logistic regression model analysis shown in Table XXI. But it was significant (*p* = 0.0027) in the logistic regression model with interactions between C<sub>1</sub> and M<sub>2</sub>-initial vocoid types, not shown here (using the glm function from the MASS package in R; Venables & Ripley 2002). In this analysis, no significant difference in the obstruency and velar nasal effects was found between the two crucial M<sub>2</sub>-initial vocoid types, i.e. /ja/ and /ju/.

Let us consider how well these wug-test results match the predictions of the maxent model learning simulations presented in §4. Table XXIII shows the insertion rates from the wug-test results, along with the predictions of the maxent model.

M <sub>2</sub> -initial sequence	j [V, -high]			j [V, +high]		
C <sub>1</sub> type	son	obs	η	son	obs	η
maxent-model prediction	49·18	39·17	33·87	64·67	54·91	49·21
wug-test result	35·96	22·11	18·32	45·92	30·10	25·58

*Table XXIII*

Insertion rate (%) in the wug test, by C<sub>1</sub> and M<sub>2</sub>-initial vocoid types.

These rates are consistently lower after obstruents and /η/, and higher before /j [+high]/, and show a relatively good match between the wug-test results and the predictions of the learned grammar, suggesting that the learned grammar can explain not only existing word patterns, as shown in §4, but also novel word patterns.

To summarise the results of the current experiment with novel words, all three main effects, syllabicity, obstruency and velar nasal, are confirmed, replicating Hwang's (2008) experimental results.<sup>25</sup> In addition, insertion rate is higher before /ju/ than before /ja/, confirming the height effect. However, the insertion rate is higher for shorter M<sub>1</sub>, which is the opposite of the tendency for existing words. Thus the length effect is not confirmed. These results suggest that Seoul Korean speakers learn many prominent patterns among existing words through phonological generalisations, but not all prominent patterns of existing words can be learned.

## 6 Discussion

The present study has two aims: (i) to find out whether speakers learn all the tendencies which can be observed in existing words in their native language, and (ii) to develop a plausible formal account of Korean n-insertion, relying crucially on P-map theory. In this section, I discuss how the two aims have been achieved, and what problems remain, along with the question of how this study differs from previous studies on Korean n-insertion.

<sup>25</sup> The size of the obstruency and velar nasal effects is much larger in Hwang's experimental results, as can be seen by comparing the rates in Tables II and XX. However, as can be seen in Table XI, the size of the corresponding effects in existing Korean words is not as large either. Thus the results of my experiment with novel words are closer to the results of the survey on existing Korean words than those of Hwang's, although the exact source of the difference in the effect size between the two experiments is unclear.

## 6.1 Previous studies

As mentioned in the introduction, there has been a great deal of research on n-insertion in Korean. Most of it has discussed both gradient and categorical patterns, but none of it has provided an analysis of gradient patterns. As far as I know, no previous study has provided a detailed description and analysis of variation involved in n-insertion in both existing and novel words and phrases. Zuraw (2011), which inspired the present study, provides a systematic investigation of the distribution of variant output forms of n-insertion. But the focus of her study is *sai-sios*, not n-insertion. The data from *sai-sios* and n-insertion are considered together, which I think might be inappropriate, since *sai-sios* is basically a compound marker, whereas n-insertion can occur freely in other constructions, such as affixed words and phrases (Im 1981: 6).<sup>26</sup> As discussed in §1.3, Hwang (2008) investigates Korean speakers' production of novel words, while providing an analysis of the experiment results within the framework of Partially Ordered Constraints Theory (Anttila 1997). He does not provide a detailed description of variation in existing words, and thus a true comparison between existing and novel word patterns is not possible.

## 6.2 P-map effects

In §3.3, I provided a P-map account of n-insertion. Based on Steriade's (2001, 2009) P-map theory, I explained (i) the choice of the epenthetic consonant and basic application conditions of n-insertion (i.e. why /n/, as opposed to other consonants, is inserted before /i j/, as opposed to other vocoids), (ii) the syllabicity effect (i.e. why insertion is more likely before a glide /j/ than a vowel /i/) and (iii) the height effect (i.e. why insertion is more likely when /j/ is followed by a high vowel than by a non-high vowel). Based on the discussion of acoustic and perceptual factors involved in the context of n-insertion, I argued that /n/ in Korean is perceptually confusable with zero in the n-insertion context, and even more so in the context where n-insertion is highly productive.

In order to complete my P-map analysis of n-insertion, several potential problems need to be addressed, as pointed out by anonymous reviewers. First, the P-map account presented in §3.3 focuses on typical cases of n-insertion, in which the epenthetic consonant appears as a palatalised nasal [ɲ]. When C<sub>1</sub> is /l/, the epenthetic consonant /n/ undergoes lateralisation, becoming [l], which differs in a number of acoustic aspects from [n].<sup>27</sup> Thus, unlike [n]-insertion, [l]-insertion might be perceptually obtrusive,

<sup>26</sup> See Ko (1992) for discussion of why n-insertion should be considered distinct from the *sai-sios* phenomenon.

<sup>27</sup> Phonotactically, the lateral [l] does not occur in postconsonantal contexts unless the preceding consonant is also a lateral. In such contexts, the liquid phoneme /l/ typically undergoes nasalisation, regardless of whether the preceding consonant is nasal or not, e.g. /tehak-lo/ [tehaŋno] 'university street' (see Jun 2000 for cases where this l-nasalisation is not obligatory). Thus, [l] hardly appears in the n-insertion contexts other than the post-/l/ one.



as the one for the choice of the deletion target, it is predicted that ‘the segments most likely to be inserted are also most likely to be deleted’ (Steriade 2009: 175). In Korean, as discussed in §3.3, /n/ is inserted before a high front vocoid, because it is the most confusable with zero in that environment. For the same reason, /n/ is the target of deletion before a high front vocoid at the beginning of the word.

Is there a causation relationship between n-deletion and n-insertion? Ko (1992) proposes a rule-inversion account of n-insertion in which word-initial n-deletion is reanalysed as a word-medial n-insertion rule. Note too the similarity between the insertion/deletion of /n/ before /i j/ in Korean and the insertion/deletion of postvocalic /r/ in English dialects, which has also been analysed as a case of rule inversion (Vennemann 1972). Ko’s proposal is criticised by Kim *et al.* (2002: 58), who point out that the contexts of the n-insertion and n-deletion rules are not perfectly complementary, since n-insertion is confined to postconsonantal position. I agree with Kim *et al.* that word-initial n-deletion was not the direct cause of n-insertion. However, it is quite possible that alternations involved in n-deletion might enhance the confusability between /n/ and zero before high front vocoids, leading to n-insertion. This possibility is based on the assumption that the P-map can be constructed from language learners’ experience not only of (universal and language-specific) phonetics, as discussed in §3.3, but also of alternations.<sup>29</sup>

Finally, the analysis presented in §3 does not differentiate between word-initial and medial morphemes beginning with high front vocoids. Specifically, the ranking (or relative weights) ONSET, V-NUC ≫ DEP(n) would lead to n-insertion not only after a consonant (/com-jak/ → [com.njak]), but also, incorrectly, at the beginning of a word (/jaku/ → [jaku], \*[njaku]) ‘baseball’). Two possible solutions can be suggested. First, word-initial n-insertion might be blocked by the constraint banning a coronal nasal before high front vocoids at the beginning of the prosodic word. This constraint is independently needed to explain word-initial n-deletion, as discussed above. One problem is that, as mentioned above, recent loanwords are not subject to this constraint, and thus its synchronic status is questionable. The second solution, which is more in line with the rest of the proposed analysis of n-insertion, is P-map-based. The inserted /n/ after a consonant would at least partially be hidden by the preceding consonant, and thus n-insertion would be perceptually less noticeable after a consonant than at the beginning of a word, where word-initial /n/ would be realised with almost no loss of its perception cues. In P-map theory, this perceptibility difference suggests that the DEP constraint prohibiting word-initial n-insertion (DEP(n) /  $\omega$ [\_\_ V]) must be ranked above that for postconsonantal n-insertion (DEP(n) / C \_\_ V). The absence of word-initial n-insertion can be attributed to high-weighted DEP(n) /  $\omega$ [\_\_ V]. In contrast, word-medial n-insertion might be due to

<sup>29</sup> Thanks to the associate editor for pointing out this possibility to me.

low-weighted DEP-(n) / C \_\_ V. Therefore, the DEP(n) constraints presented above for the analysis of Korean n-insertion belong to the DEP(n) / C \_\_ V constraint family. For instance, the correct form of the constraint DEP(n) / \_\_j adopted in the proposed analysis would be DEP(n) / C \_\_j. Either of these solutions can explain why n-insertion does not occur word-initially.

### 6.3 Do speakers learn all statistically prominent patterns?

The results of the present study suggest that Seoul Korean speakers are aware of most of the tendencies in the distribution of existing words, and use this knowledge when they apply n-insertion to novel words. As shown in §3 and §4, the internalised grammar and how it is learned can be formalised and implemented in a probabilistic constraint-based grammar, such as a maxent grammar, and its related learning model.

However, the results of the present study also suggest that speakers do not know all statistically prominent patterns in existing words. When  $M_1$  is disyllabic, rather than monosyllabic, insertion rate is higher in existing words, but lower in novel words. This means that speakers are not aware of the length effect, i.e. the higher rate of insertion with longer  $M_1$ , in existing Korean words.

Why do Korean speakers fail to learn the tendency about the length of  $M_1$ ? Given that only phonologically natural patterns can be learned easily (Becker *et al.* 2011, Hayes *et al.* 2009, Hayes & White 2013), the failure to learn the length effect would indicate that the length of  $M_1$  is not a phonologically natural factor which can condition the application of n-insertion or affect the probability that n-insertion applies. This seems to be true, especially under the proposed P-map account, in which the P-map determines many important aspects of n-insertion, including the basic application conditions and factors affecting its probability.

As discussed in the previous section, this P-map effect is mostly concerned with the elements of  $M_2$ . There does not seem to be a plausible reason for a longer  $M_1$  enhancing the confusability between /n/ and zero at the beginning of  $M_2$ . More generally, I cannot find any phonetic or phonological grounding for the length effect. Accordingly, I attribute Korean speakers' failure to learn the length effect to the lack of phonological naturalness involved.

It should also be noticed that the results of the previous studies report non-negligible amounts of n-insertion (44.7% on average) when the  $M_2$ -initial vocoid is a vowel /i/. But the results of the wug test suggest that the productivity of n-insertion before /i/ is no different from that of vowels such as /a e/, which are known not to trigger n-insertion in Korean. Hwang's (2008) results are no different in this respect from mine. Is this then another case in which speakers failed to learn a statistically prominent pattern? The answer is not clear. The test words employed

in the previous studies constitute a very small subset of the relevant Seoul Korean words, especially when the  $M_2$  initial vocoid is /i/. Notice that the dictionary database discussed in §2.1, which reports n-insertion in fewer than 10% of the words with /i/-initial  $M_2$ , includes many more words than previous studies, which focused on words likely to undergo n-insertion. It is thus probable that most of the words not employed in those studies are unlikely (or less likely) to undergo n-insertion. The actual rate of n-insertion in existing Seoul Korean words with /i/-initial  $M_2$  must be much lower than the average insertion rate calculated from the results of previous studies. Further research is needed to determine whether the n-insertion rate for existing Korean words with /i/-initial  $M_2$  is low enough to be ignored by Korean speakers.

## 7 Conclusion

In this paper, I have explored variation in Seoul Korean n-insertion, with the following two aims: (i) to find out whether speakers learn the tendencies which can be observed in existing words in their native language, and (ii) to develop a formal account of Korean n-insertion. I first discussed a number of tendencies in the distribution of n-insertion in existing Seoul Korean words, on the basis of data from my own survey, the *SKD* and previous studies. These tendencies can be summarised as follows: (i) n-insertion is more likely before a glide than before a vowel; (ii) n-insertion is less likely after obstruents than after sonorants, except /ŋ/; (iii) n-insertion is less likely after /ŋ/ than after other sonorants; (iv) n-insertion is more likely when /j/ precedes a high vowel; (v) n-insertion is more likely when the morpheme preceding the insertion site is longer.

I then provided a constraint-based analysis of n-insertion, crucially relying on P-map theory. Based on this analysis, I conducted a maxent-model learning simulation of Seoul Korean n-insertion, taking the distribution of the responses in my own survey as the training data. The results of the learning simulation demonstrate that the proposed grammar can be learned from the Korean data, and that it can explain the observed quantitative patterns of n-insertion.

Finally, I conducted an experiment using novel words, specifically loanwords and nonce words. The results confirmed most of the effects found in existing Korean words, but the length effect was not confirmed. From this data *vs.* learning mismatch, it seems that language learners are not aware of all the prominent tendencies in existing words. To explain the mismatch, I argued that the length of  $M_1$  is not a phonologically natural factor which can condition the application of n-insertion, or influence the probability that it applies. Specifically, I attributed the data *vs.* learning mismatch to the lack of perceptual basis for the length effect.

**Appendix: Training input file***Abbreviations:*

M = sonorant consonants other than /ŋ/

P = obstruents

V<sub>2</sub> = vocoids other than /i j/

{C} = ambisyllabic consonant

training items			constraint violations								
input	output	frequency	*CODA	ONSET	ALIGN-R	V-NUC	*AMBI-ŋ	DEP(ŋ) / _[j+h]	DEP(ŋ) / _[j-h]	Id[-son]	*[ŋ]
VM-ja	Vm.nja	1295	1						1		
	V.mja	935			1						
	Vm.ja	533	1			1					
	Vm.[ja] <sub>Nuc</sub>	0	1	1							
VP-ja	Vm.nja	531	1						1	1	
	V.pja	595			1						
	Vp.ja	271	1			1					
	Vp.[ja] <sub>Nuc</sub>	0	1	1							
V <sub>ŋ</sub> -ja	V <sub>ŋ</sub> .nja	715	1						1		
	V. <sub>ŋ</sub> ja	0			1						1
	V{ <sub>ŋ</sub> }ja	1160	1				1				
	V <sub>ŋ</sub> .[ja] <sub>Nuc</sub>	0	1	1							
VM-ju	Vm.nju	217	1					1			
	V.mju	31			1						
	Vm.ju	89	1			1					
	Vm.[ju] <sub>Nuc</sub>	0	1	1							
VP-ju	Vm.nju	126	1					1		1	
	V.pju	28			1						
	Vp.ju	46	1			1					
	Vp.[ju] <sub>Nuc</sub>	0	1	1							
V <sub>ŋ</sub> -ju	V <sub>ŋ</sub> .nju	95	1					1			
	V. <sub>ŋ</sub> ju	0			1						1
	V{ <sub>ŋ</sub> }ju	129	1				1				
	V <sub>ŋ</sub> .[ju] <sub>Nuc</sub>	0	1	1							
VC-V <sub>2</sub>	V.CV <sub>2</sub>	57766			1						
	VC.V <sub>2</sub>	10194	1	1							

## REFERENCES

- Ahn, Mee-Jin (2009). /n/-insertion and onset simplification in Kyungsang Korean. *Studies in Phonetics, Phonology and Morphology* 15. 263–282.
- Anttila, Arto (1997). Deriving variation from grammar. In Frans Hinskens, Roeland van Hout & W. Leo Wetzels (eds.) *Variation, change and phonological theory*. Amsterdam & Philadelphia: Benjamins. 35–68.
- Bae, Ju-Chae (2003). *The pronunciation of Korean*. [*Hankwukeuy palum*.] Seoul: Samkyong Munhwasa.
- Bates, Douglas M., Martin Maechler & Ben Bolker (2011). Package ‘lme4’ (Version 0.999375-39): linear mixed-effects models using S4 classes. Available at <http://cran.r-project.org/web/packages/lme4/lme4.pdf>.
- Becker, Michael, Nihan Ketzet & Andrew Nevins (2011). The surfeit of the stimulus: analytic biases filter lexical statistics in Turkish laryngeal alternations. *Lg* 87. 84–125.
- Boersma, Paul (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21. 43–58.
- Boersma, Paul (1998). *Functional phonology: formalizing the interactions between articulatory and perceptual drives*. PhD dissertation, University of Amsterdam.
- Boersma, Paul & Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *LI* 32. 45–86.
- Boersma, Paul & Joe Pater (forthcoming). Convergence properties of a Gradual Learning Algorithm for Harmonic Grammar. In John J. McCarthy & Joe Pater (eds.) *Harmonic Grammar and Harmonic Serialism*. London: Equinox.
- Borden, Gloria J. & Katherine S. Harris (1984). *Speech science primer: physiology, acoustics, and perception of speech*. 2nd edn. Baltimore: Williams & Wilkins.
- Casali, Roderic F. (1997). Vowel elision in hiatus contexts: which vowel goes? *Lg* 73. 493–533.
- Choi, Hye-Won (2002). *A survey of standard pronunciation*. [*Phyocunpalum silthaycosa*.] Seoul: National Institute of the Korean Language.
- Choi, Hyun-Bae (1971). *Our grammar*. [*Ulimalpon*.] 4th (posthumous) edn. Seoul: Cengumsa.
- Chung, Chin Wan (2001). The behavior of velar nasal and syllabification. *Studies in Phonetics, Phonology and Morphology* 7:1. 177–189.
- Coetzee, Andries W. (2004). *What it means to be a loser: non-optimal candidates in Optimality Theory*. PhD dissertation, University of Massachusetts, Amherst.
- Coetzee, Andries W. & Joe Pater (2011). The place of variation in phonological theory. In John Goldsmith, Jason Riggle & Alan Yu (eds.) *The handbook of phonological theory*. 2nd edn. Malden, Mass. & Oxford: Wiley-Blackwell. 401–434.
- Ernestus, Mirjam & R. Harald Baayen (2003). Predicting the unpredictable: interpreting neutralized segments in Dutch. *Lg* 79. 5–38.
- Goldwater, Sharon & Mark Johnson (2003). Learning OT constraint rankings using a Maximum Entropy model. In Jennifer Spenador, Anders Eriksson & Östen Dahl (eds.) *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. Stockholm: Stockholm University. 111–120.
- Guy, Gregory R. (1980). Variation in the group and the individual: the case of final stop deletion. In William Labov (ed.) *Locating language in time and space*. New York: Academic Press. 1–36.
- Han, Eunjo (1993). The phonological word in Korean. In Patricia M. Clancy (ed.) *Japanese/Korean Linguistics*. Vol. 1. Stanford: CSLI. 117–129.
- Hayes, Bruce (2009). Maxent grammar tool. Software package. Available at <http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool>.
- Hayes, Bruce & Zsuzsa Cziráky Londe (2006). Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology* 23. 59–104.

- Hayes, Bruce & James White (2013). Phonological naturalness and phonotactic learning. *LI* 44. 45–75.
- Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *LI* 39. 379–440.
- Hayes, Bruce, Kie Zuraw, Péter Siptár & Zsuzsa Londe (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Lg* 85. 822–863.
- Hong, Soonhyun (1997). *Prosodic domains and ambisyllabicity in Optimality Theory*. PhD dissertation, University of Pennsylvania.
- Hong, Soonhyun (2006). /n/-insertion in native Korean and Sino-Korean revisited. *Studies in Phonetics, Phonology and Morphology* 12. 391–413.
- Huh, Woong (1984). *Korean phonology*. [*Kwukeumwunhak*.] Seoul: Cengumsa.
- Hwang, Sangjin (2008). *Korean speakers' knowledge of /n/-insertion: p-map approach*. MA thesis, Seoul National University.
- Hwang, Yeon-Shin (2002). *An experimental study of the Korean nasal*. [*Hankwuke piu-meykwwanhan silhemumsenghakcekyenkwu*.] MA thesis, Seoul National University.
- Im, Hong-Pin (1981). On the solution of the sai-sios problem. [Sai-sios mwunceyyu haekyelul wihaye.] *Kukehak* 10. 1–35.
- Itô, Junko (1986). *Syllable theory in prosodic phonology*. PhD dissertation, University of Massachusetts, Amherst.
- Itô, Junko (1989). A prosodic theory of epenthesis. *NLLT* 7. 217–259.
- Itô, Junko & Armin Mester (1994). Reflections on CodaCond and Alignment. In Jason Merchant, Jaye Padgett & Rachel Walker (eds.) *Phonology at Santa Cruz 3*. Santa Cruz: Linguistics Research Center. 27–46.
- Itô, Junko & Armin Mester (1999). Realignment. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.) *The prosody–morphology interface*. Cambridge: Cambridge University Press. 188–217.
- Johnson, Keith (1997). *Acoustic and auditory phonetics*. Cambridge, Mass.: Blackwell.
- Jun, Jongho (2000). Preliquid nasalization. *Korean Journal of Linguistics* 25:2. 191–208.
- Jun, Sun-Ah (1995). Asymmetrical prosodic effects on the laryngeal gesture in Korean. In Bruce Connell & Amalia Arvaniti (eds.) *Phonology and phonetic evidence: papers in laboratory phonology IV*. Cambridge: Cambridge University Press. 235–253.
- Kang, Beom-mo & Kim Hung-gyu (2004). *Frequency analysis of Korean morpheme and word usage*. [*Hankwuke hyengtayso mich ehwi sayong pintouy pwunsek*.] Vol. 2. Seoul: Institute of Korean Culture, Korea University.
- Kang, Hyeon-Seok (1997). *Phonological variation in glides and diphthongs of Seoul Korean: its synchrony and diachrony*. PhD dissertation, Ohio State University.
- Kang, Kyung-Shim (2003). The status of onglides in Korean: evidence from speech errors. *Studies in Phonetics, Phonology and Morphology* 9. 1–15.
- Kang, Ongmi (2003). *Korean phonology*. [*Hankwuke umwunlon*.] Seoul: Thaeaksa.
- Kang, Yongsoon (1991). *Phonology of consonant–vowel interaction: with special reference to Korean and dependency phonology*. PhD dissertation, University of Illinois, Urbana-Champaign.
- Ki, Se-kwan (1990). *A phonological study of /l/-deletion and /n/-addition in Korean word-formation*. [*Kwuke tanehyengsengeyseyu liul thallakkwa niun chemkaey tayhan umwunloncek yenku*.] PhD dissertation, Wonkwang University, Iksan.
- Kim, Chin-Wu (1970). Boundary phenomena in Korean. *Papers in Linguistics* 2. 1–26.
- Kim, Chin-Wu & Hyoung-Youb Kim (1991). The character of Korean glides. *Studies in the Linguistic Sciences* 21:2. 113–125.
- Kim, Jong-mi (1986). *Phonology and syntax of Korean morphology*. PhD dissertation, University of Southern California, Los Angeles.
- Kim, Seoncheol (2003). *A survey of standard pronunciation*. [*Phyocwunpalum silthay-cosa*.] Vol. 2. Seoul: National Institute of the Korean Language.

- Kim, Youngshin (2011). *An acoustic, aerodynamic and perceptual investigation of word-initial denasalization in Korean*. PhD dissertation, University College London.
- Kim, Yu-Pum, Sunwoo Park, Byoung-Sup Ahn & Bongwon Lee (2002). An overview of studies of Korean /n/-insertion. [Niun saphyensanguy yenkwusacek kemtho.] *Emwunnoncip* 46. 43–71.
- Kim-Renaud, Young-Key (1974). *Korean consonantal phonology*. PhD dissertation, University of Hawaii. Published 1991, Seoul: Hanshin.
- Kiparsky, Paul (1993). An OT perspective on phonological variation. Handout from Rutgers Optimality Workshop.
- Ko, Kwang-mo (1992). n-epenthesis and sai-sios in Korean. [Niun chemkawa saisio-seytayhan yenkwu.] *Eoneohak* 14. 31–51.
- Kohler, Klaus J. (1990). Segmental reduction in connected speech in German: phonological facts and phonetic explanation. In W. J. Hardcastle & A. Marchal (eds.) *Speech production and speech modelling*. Dordrecht: Kluwer. 69–92.
- Kook, Kyung-A, Ju-Won Kim & Ho-Young Lee (2005). A study of n-insertion preferences in Korean. [Senhoto cosalulthonghan niun chemkahyensanguy silhyenyang-sang yenkwu.] *Malsori* 53. 37–60.
- Kwack, Dong-Kee (1992). *A phonological study on Korean prosodic structures*. [Wunyultanwiew uyhan kwuke umwunhyensanguy pwunsek.] PhD dissertation, Seoul National University.
- Kwuklip kwuke yenkwuwun [National Institute of the Korean Language] (ed.) (1999). *Standard Korean dictionary*. [Pyocwun kwuke taysacen.] Seoul: Doosan Dong-A. Available at <http://stdweb2.korean.go.kr>.
- Lee, Hoyoung (1996). *Korean phonetics*. [Kukeumsenghak.] Seoul: Thaeaksa.
- Lee, Jin-Seong (1992). *Phonology and sound symbolism of Korean ideophones*. PhD dissertation, Indiana University.
- Lee, Minkyung (2006). Gyeongsang Korean n-insertion revisited. *Studies in Phonetics, Phonology and Morphology* 12. 623–641.
- Lee, Ponghyung (2004). Korean sonorant modifications as domain initial strengthening. *Korean Journal of Linguistics* 29. 607–633.
- Lee, Yongsung & Minkyung Lee (2006). n-insertion as y-devocalization in Korean. *Korean Journal of Linguistics* 31. 413–440.
- Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky (1990). Harmonic Grammar: a formal multi-level connectionist theory of linguistic well-formedness: an application. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*. Hillsdale: Erlbaum. 884–891.
- McCarthy, John J. & Alan Prince (1995). Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*. Amherst: GLSA. 249–384.
- Oh, Mira (2006). Reexamination of environments for /n/-insertion. [Niun-sapip hwan-kenguy caykemtho.] *The Linguistic Association of Korea Journal* 14:3. 117–135.
- Park, Jong-Hee (2001). The status of onglide ‘j’ in Korean. [Olum hwalum ‘j’uy umcel naypu wichi.] *Korean Journal of Linguistics* 26. 715–733.
- Park, Sunwoo (2005). Positional markedness and /n/-insertion in Korean. [Wichicek yuphosengkwa hankwukeuy niunsapip.] *Studies in Phonetics, Phonology and Morphology* 11. 323–335.
- Park, Sunwoo (2008). An acoustic study of Korean place assimilation. [Hankwuke wichtonghwayu silhemumsenghakcek pwunsek.] *Linguistic Research* 25:2. 45–65.
- Prince, Alan & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder. Published 2004, Malden, Mass. & Oxford: Blackwell.
- R Development Core Team (2014). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at <http://www.r-project.org>.

- Rubach, Jerzy (2000). Glide and glottal stop insertion in Slavic languages: a DOT analysis. *LI* 31. 271–317.
- Shin, Jiyoung, Minjung Kim & Keeho Kim (2000). The phonetic realization of the accentual phrase and resyllabification in Korean. [Hankwuke kangseykwuuy umsengecek silhyen yangsangkwwa cayumcelhwa.] *Korean Journal of Linguistics* 25. 383–403.
- Shin, Seung-Hoon (1997). *Constraints within and between syllables: syllable licensing and contact in Optimality Theory*. PhD dissertation, Indiana University.
- Smolensky, Paul (1986). Information processing in dynamical systems: foundations of Harmony Theory. In D. E. Rumelhart, J. L. MacClelland & the PDP Research Group (eds.) *Parallel Distributed Processing: explorations in the micro-structure of cognition*. Vol. 1. Cambridge, Mass.: MIT Press. 194–281.
- Sohn, Hyang-Sook (1987). *Underspecification in Korean phonology*. PhD dissertation, University of Illinois at Urbana-Champaign.
- Steriade, Donca (2001). Directional asymmetries in place assimilation: a perceptual account. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*. San Diego: Academic Press. 219–250.
- Steriade, Donca (2009). The phonology of perceptibility effects: the P-map and its consequences for constraint organization. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: studies in honor of Paul Kiparsky*. Cambridge, Mass.: MIT Press. 151–179.
- Venables, W. N. & B. D. Ripley (2002). *Modern applied statistics with S*. 4th edn. New York: Springer.
- Vennemann, Theo (1972). Rule inversion. *Lingua* 29. 209–242.
- Wilson, Colin (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* 30. 945–982.
- Yun, Youngdo (2004). *Glides and high vowels in Korean syllables*. PhD dissertation, University of Washington.
- Zuraw, Kie (2000). *Patterned exceptions in phonology*. PhD dissertation, University of California, Los Angeles.
- Zuraw, Kie (2002). Aggressive reduplication. *Phonology* 19. 395–439.
- Zuraw, Kie (2010). A model of lexical variation and the grammar with application to Tagalog nasal substitution. *NLLT* 28. 417–472.
- Zuraw, Kie (2011). Predicting sai-siot in Korean compound nouns: phonological and non-phonological factors. Paper presented at the 21st Japanese/Korean Linguistic Conference, Seoul National University.