



# Morphophonological gradience in Korean n-insertion

JONGHO JUN 

RESEARCH

]u[ubiquity press

## Abstract

This study addresses the questions of what factors may have a gradient effect in application of a morphophonological process, how they interact, and which of the gradient effects speakers are aware of, by investigating the variation patterns of Korean n-insertion. An analysis is performed on the results of two surveys on speakers of two dialects of Korean, Seoul and Kyungsang, one using existing words and the other using novel Korean words. From the results of the survey on n-insertion in existing Korean words, I have found several gradient tendencies involving a variety of factors including phonological and morphological ones and their interactions. Such factors include morphological category, etymology and length of component morphemes, sonorancy and place of articulation of the consonant preceding the insertion site, height of the initial vowel of the morpheme following the insertion site and dialects of speakers. None of these factors are absolute conditions for the occurrence of n-insertion, contrary to the previous studies on Korean n-insertion, and they have gradient effect, contributing to the overall probability of n-insertion. Consequently, Korean n-insertion provides a clear case where previous categorical proposals do not match with gradient reality, lending support to quantitative theories of morphophonology.

CORRESPONDING AUTHOR:

**Jongho Jun**

Seoul National University, KR

[jongho@snu.ac.kr](mailto:jongho@snu.ac.kr)

KEYWORDS:

Morphophonology; gradient phonology; Korean n-insertion; lexicon-learning mismatch

TO CITE THIS ARTICLE:

Jun, Jongho. 2021.  
Morphophonological  
gradience in Korean  
n-insertion. *Glossa: a journal  
of general linguistics* 6(1):  
40. 1–40. DOI: [https://doi.  
org/10.5334/gjgl.1401](https://doi.org/10.5334/gjgl.1401)

A morphophonological process is typically conditioned by phonological and morphological factors. Such factors categorically distinguish words that can undergo the process from those that cannot. Some other factors, mainly non-linguistic ones such as word frequency, may have a gradient effect. They affect the process only probabilistically within the set of potential target words. However, recent research (Zuraw 2000, 2002, 2010; Hayes et al. 2009; Coetzee & Pater 2011; Hayes & White 2013; McPherson & Hayes 2016; Jurgec 2016; Moore-Cantwell & Pater 2016; Zuraw & Hayes 2017; and Zymet 2018) shows that phonological and morphological factors may have a gradient effect, contributing to the overall probability of the occurrence of a morphophonological process. The present study addresses the questions of what factors may have such a gradient effect in application of a morphophonological process, how they interact, and which of the gradient effects speakers are aware of, by investigating the variation patterns of Korean n-insertion.

For this purpose, I investigated the results of two surveys on speakers of two dialects of Korean, Seoul and Kyungsang, one using existing Korean words and the other using novel Korean words. The results of the survey with existing words show that a variety of factors including phonological and morphological ones interact to give rise to several interesting gradient tendencies involving n-insertion. This is consistent with the recent research on gradient morphophonology, cited above, but contrary to the traditional literature on Korean n-insertion which argue that most of such factors are absolute conditions for the occurrence of n-insertion.

In order to establish whether Korean speakers are aware of the tendencies observed in existing words, I explored the results of a novel word survey. Several tendencies observed in existing words were mirrored in the results of a novel word survey. Thus, Korean speakers are aware of the differential influence of the relevant factors on the probability of the application of n-insertion. There were, on the other hand, some apparent mismatches between the results of the surveys on existing and novel words, which I attribute to the lack of phonological naturalness involved.

The present work seeks to complement that of Jun (2015) by further exploring what factors affect the distribution of n-insertion in Korean and how they interact. Although this study can be considered as an extension of Jun's study on Seoul Korean n-insertion, it would make a distinct contribution by using a larger set of data and providing a more elaborate data analysis. Specifically, the database of the present study consists of the results of two surveys on Korean n-insertion, one on Seoul Korean speakers, conducted by Jun, and the other on Kyungsang Korean speakers, conducted in the present study. I will analyze the data, using an expanded set of factors including the dialect of the participants and morphological categories of component morphemes, which were not considered in Jun's work. Thus, the present study provides a more comprehensive experimental investigation of Korean n-insertion, addressing a wider set of empirical and theoretical issues. A variety of factors, which have been proposed to have a categorical effect in the previous studies on Korean n-insertion, turned out to have only a gradient effect in this study. In addition, the present analysis of Korean n-insertion can lead to better understanding of whether and how speakers learn gradient lexical patterns.

The organization of this paper is as follows. Section 2 provides background information on Korean n-insertion, beginning with a description of the basic patterns of Korean n-insertion. It then discusses conditioning factors for the occurrence of n-insertion which previous studies argue to have categorical effects. Sections 3–4 describe surveys on Seoul and Kyungsang Korean speakers. Section 3 concerns the survey on existing Korean words whereas section 4 the survey on novel Korean words. Effects of conditioning factors will be examined both separately and in combination. Section 5 provides discussions of how to analyze Korean n-insertion, the extension of existing word patterns to novel words, dialectal variation and remaining problems. The final section concludes the present study.

In Korean, /n/ is inserted at the juncture of two morphemes when the first morpheme ( $M_1$ ) ends with any consonant, called here  $C_1$ , and the following morpheme ( $M_2$ ) begins in a high front vocoid, /i, j/, as illustrated in (1) and (2).

- (1) Korean n-insertion: rule form  
 $\emptyset \rightarrow n / C_1 ]_{M_1} - [_{M_2} i/j$
- (2) Korean n-insertion: examples<sup>1</sup>
- /som-ipul/ [sɔmnipul] ‘cotton sheet’
  - /com-jak/ [comnjak] ‘mothball’

The epenthetic /n/ is phonetically realized as a palatalized nasal [ɲ] after a non-liquid  $C_1$  consonant, and thus phonetically more accurate transcriptions of the examples in (2) would be [sɔmnipul] and [comnjak].<sup>2</sup> This is due to allophonic palatalization in Korean where /n, l/ become palatalized before high front vocoids /i, j/. When  $C_1$  is a liquid, the epenthetic /n/ undergoes both palatalization and lateralization where /n/ becomes [l] after a lateral: e.g. /al-jak/ [aɭɭjak] ‘pill’. When n-insertion occurs, the  $C_1$  consonant is always realized as a nasal due to obstruent nasalization in Korean where an obstruent becomes a nasal before a nasal (e.g. /hot<sup>h</sup>-ipul/ [honnipul] ‘unlined comforter’), except when  $C_1$  is underlyingly a liquid which surfaces as a lateral, as mentioned above.

The basic conditions for application of n-insertion, illustrated above with (1), are not enough to figure out when to apply and when not to apply n-insertion. There are a number of words which meet the basic conditions but do not undergo n-insertion. Many additional conditioning factors have been proposed in the previous literature for the purpose of properly restricting the domain of application of n-insertion. The crucial factors include morphological category of component morphemes, syllabicity of  $M_2$ -initial vocoid, and sonorancy of  $C_1$ . I will now provide a brief review of relevant previous studies.

## 2.1 Morphology

It has been reported in the literature that n-insertion does or does not apply depending on the type of morpheme involved. Morphemes preceding and following the insertion site, called here  $M_1$  and  $M_2$ , have been argued to be subject to different restrictions on the occurrence of n-insertion. I will first discuss previous proposals on the restriction of  $M_1$  and then those of  $M_2$ .

### 2.1.1 Morphology of $M_1$

n-insertion is more pervasive and widespread in Kyungsang Korean than Seoul Korean. The main difference lies in the type of triggering  $M_1$  morpheme. According to Han (1994: 114–130), as summarized in (3), n-insertion in Seoul Korean occurs when  $M_1$  is a free stem (3a, b) or a prefix (3c), but does not occur when  $M_1$  is a bound root (3d, e) which is mostly Sino-Korean.

- (3) Category of  $M_1$  vs. composition (O: n-insertion, X: non-n-insertion, based on Han 1994)<sup>3</sup>

	$M_1$ - $M_2$	Seoul	Kyungsang	Examples
a.	<b>stem-stem</b>	O	O	/puak-il/ [puakɲil] ‘kitchen work’
b.	<b>stem-root/suffix</b>	O	O	/sikjoŋ-ju/ [sikjoŋnju] ‘cooking oil’
c.	<b>prefix-stem</b>	O	O	/hot <sup>h</sup> -ipul/ [honnipul] ‘unlined comforter’
d.	<b>root-root</b>	X	O	/min-jo/ [minjo ~ minnjo] ‘folk song’
e.	<b>root-stem</b>	X	O	/il-joil/ [iljoil ~ illjoil] ‘Sunday’ /an-jak/ [anjak ~ annjak] ‘eye drops’

1 Allophonic variations such as inter-sonorant voicing (i.e. lenis stops become voiced between sonorants) and palatalization (see main text) are not reflected in the transcription in the remainder of this paper unless they are relevant to the discussion at hand. In contrast, the representations include forms derived by neutralizing processes such as obstruent nasalization (see main text), coda neutralization (i.e. coda obstruents are neutralized into homorganic unreleased lenis stops), and post-obstruent tensing (i.e. a lenis obstruent becomes tense after an obstruent). /c/, /<sup>h</sup>/ and /ʔ/ stand for the coronal affricate, aspirated and glottalized (or tense), respectively.

2 With palatalization, the triggering palatal glide is deleted or at least significantly weakened (Kang 2003; Lee & Lee 2006; Ahn 2008).

3 Han (1994: 123) classifies Sino-Korean monosyllabic  $M_2$ 's in words like /sikjoŋ-ju/ ‘cooking oil’ as a root. But they may alternatively be classified as a suffix. In this study, I adopt the latter classification, following the Standard Korean dictionary (Kwuklip kwuke yenkwuwen 1999; henceforth SKD).

In contrast, n-insertion in Kyungsang Korean occurs even after a root (3d, e), as well as after a stem and a prefix (3a–c). Thus, the morphological domain of n-insertion is narrower for Seoul Korean compared to Kyungsang Korean.

To summarize, the dialectal difference with respect to morphological category of  $M_1$  suggests that a root  $M_1$  is a weaker trigger of n-insertion than a non-root, namely, stem or prefix,  $M_1$ .

### 2.1.2 Morphology of $M_2$

It has been mentioned in many previous studies on Seoul Korean n-insertion that n-insertion occurs only when  $M_2$  is a stem (or root) which can be an independent word (Huh 1984; Han 1993, 1994; Ko 1992; Kim et al. 2002; Hong 2006; Cho 2016; and others). Some of those previous studies (for instance, Ko 1992: 37) consider this free  $M_2$  requirement to hold only for native Korean words. This etymological restriction makes sense given that n-insertion may occur before Sino-Korean /j/-initial suffixes, as illustrated in (3b) where /-ju/ ‘oil’ is a Sino-Korean suffix. To summarize, the previous studies on the role of morphological category of  $M_2$  in Korean n-insertion suggest that n-insertion occurs only before a free morpheme, at least in native Korean words.

## 2.2 Syllabicity of $M_2$ -initial vocoid

It is widely assumed in the traditional literature on Korean phonology and morphology that n-insertion occurs not only before a high front vowel /i/ but also before a palatal glide /j/. However, some previous studies have argued either that the vowel /i/ is not a trigger of n-insertion, or that /i/ is less likely to trigger n-insertion than the glide /j/. Lee & Lee (2006: 422–4), Hong (2006: 397) and Ahn (2009: 263) deny the existence of synchronic pre-/i/ insertion in Korean, thus claiming that /j/ is the sole trigger of n-insertion. Lee (1996: 168), Bae (2003: 241), and Oh (2006: 125–9) acknowledge the existence of pre-/i/ insertion, but describe it as less frequent, or less naturally occurring, than pre-/j/ insertion. To summarize, /j/ is a stronger trigger of n-insertion than /i/, although it is controversial whether the difference in strength is categorical or gradient.

## 2.3 Sonorancy of $C_1$

Some previous studies have argued that the sonorancy of  $C_1$  determines the likelihood that n-insertion will apply. According to Cho (1995: 610–11) and Cho & Iverson (1997: 702), Korean n-insertion applies obligatorily after a sonorant  $C_1$  and optionally after an obstruent  $C_1$ .

A similar  $C_1$  sonorancy asymmetry has been argued to hold for Sino-Korean words consisting of monosyllabic roots. In the description of Seoul Korean n-insertion presented by Oh (2006: 121–2) and followed by Ahn (2008: 384–5), Sino-Korean words consisting of monosyllabic roots undergo n-insertion only after a sonorant (e.g. /kʌm-jʌl/ [kʌmnjʌl] ‘censorship’), not after an obstruent (e.g. /pɛk-jʌl/ [pɛkjʌl] ‘white heat’). The same type of  $C_1$  sonorancy asymmetry in Sino-Korean words can also be found in Lee’s (2006: 635–6, (24) vs. (27)) description of Kyungsang Korean n-insertion.

To summarize, the previous studies on the role of sonorancy of  $C_1$  in the application of Korean n-insertion suggest that a sonorant  $C_1$  is a stronger trigger of n-insertion than an obstruent  $C_1$ , although the exact strength and domain of the sonorancy asymmetry still need to be explored.

## 2.4 Exceptions and variation

As discussed above, traditional analyses of Korean n-insertion have attempted to explain the variation in n-insertion by restricting the domain of rule application in terms of the morphological category of  $M_1$  and  $M_2$ , syllabicity of  $M_2$ -initial vocoids, and sonorancy of  $C_1$ , which may interact with additional factors such as dialect, etymology and/or length of the morphemes involved, as summarized in (4).

(4) Summary: conditioning factors for the occurrence and non-occurrence of n-insertion (sino = Sino-Korean, mono = monosyllabic)

		n-insertion	non-n-insertion
a.	morphology	M <sub>1</sub> prefix, free stem, bound root (Kyungsang)	bound root (Seoul)
		M <sub>2</sub> free stem, bound root/suffix (sino)	bound root/suffix (native)
b.	syllabicity of M <sub>2</sub> - initial vocoids	(i) before /j/ (ii) before /j/ (more frequent), /i/ (less frequent)	(i) before /i/
c.	C <sub>1</sub> sonorancy	(i) sonorant (obligatory), obstruent (optional) (ii) sonorant (mono sino)	(ii) obstruent (mono sino)

They mostly propose such factors to have categorical effects on variation while describing relevant patterns based mainly on the authors' intuition as a native speaker of Korean.

However, some exceptions to each of those factors have already been pointed out in the previous studies based on the authors' intuition, and even a wider range of exceptions were reported in experimental and survey studies on Korean n-insertion (Kim 2000; Choi 2002; Kim 2003). For instance, the free M<sub>2</sub> requirement (i.e. n-insertion applies to words with a free M<sub>2</sub>) is neither a sufficient nor a necessary condition for the occurrence of n-insertion. Not all /i, j/-initial M<sub>2</sub> stems trigger n-insertion: e.g. /mas-is'ta/ \*[mannitt'a], [masitt'a] 'delicious (taste-exist-SE)'. On the other hand, certain (/j/-initial) native Korean M<sub>2</sub> suffixes may trigger n-insertion: e.g. /camsi-man-jo/ [camsimannjo] 'Wait a moment (moment-only-SE)' (Bae 2003; Oh 2006). It seems that none of the conditioning factors proposed in the previous studies completely determine the occurrence or absence of n-insertion. In addition, as mentioned in a number of previous studies on Korean n-insertion (Kim-Renaud 1974/1991; Ko 1992; and many others), n-insertion is often optional. The probability of n-insertion may vary greatly across speakers and words, as can be seen in the results of previous experimental/survey studies on Korean n-insertion (Choi 2002; Kim 2003; and Jun 2015), and as is also evident from the data of this study presented in sections 3–4. The widespread exceptions and variation involved make Korean n-insertion look quite irregular. Nonetheless, it is still possible that the factors proposed in the previous literature may have gradient effect, contributing to the overall probability of n-insertion. In fact, such possibilities have been mentioned for some of the factors in the previous studies: for instance, a higher frequency or preference of n-insertion before /j/ than before /i/ (Lee 1996; Bae 2003; and Oh 2006) and a greater likelihood of n-insertion after a sonorant C<sub>1</sub> than after an obstruent C<sub>1</sub> (Han 1994). Unfortunately, the few claimed tendencies were mostly based on the authors' intuition, rather than the results of wide-scale experimental or survey studies.

In the next section, I will provide a comprehensive and systematic investigation of n-insertion in existing Korean words to establish what factors affect n-insertion and whether they have categorical or gradient effects. For this purpose, I consider the following expanded set of factors:

- (5) Potential factors affecting Korean n-insertion
- a. syllabicity of M<sub>2</sub>-initial vocoid
  - b. dialect
  - c. morphological category of component morphemes
  - d. etymology of component morphemes
  - e. type of C<sub>1</sub>
  - f. length of component morphemes
  - g. height of a vowel following M<sub>2</sub>-initial /j/, called here V<sub>2</sub>
  - h. word frequency

The type of C<sub>1</sub>, the length of component morphemes, and the height of V<sub>2</sub> are added to the list of factors, mainly because they have been shown to be active in the application of Seoul

Korean n-insertion by some previous studies (Hwang 2008, Jun 2015). Note that the length of a word or morpheme has been identified as a factor affecting the rate of a phonological process such as Turkish final devoicing (Becker et al. 2011). It has, in fact, been shown in Jun (2015) that n-insertion in Seoul Korean is less likely with words with monosyllabic  $M_1$  than those with longer  $M_1$ . As discussed in section 2.2, several previous studies on Korean n-insertion have argued for the effect of  $C_1$  sonorancy. However, the results of Hwang's (2008) experimental study and Jun's (2015) survey on Seoul Korean speakers show that this  $C_1$  sonorancy effect differs depending on the place of articulation of  $C_1$  sonorants. Specifically, n-insertion is less likely after a velar nasal  $C_1$  than after other sonorants. Thus, the present investigation will involve a comparison not only between words with obstruent vs. sonorant (except /ŋ/)  $C_1$ , but also between those with the velar nasal vs. other sonorant  $C_1$ . Jun also reports that the height of  $V_2$  plays a significant role in determining the rate of Seoul Korean n-insertion, and, specifically, that insertion is more likely before a high vowel than before a non-high vowel. Finally, word frequency is added to the list in (5) as it has often been shown to affect the application rate of an optional process (for instance, English final t/d deletion as discussed by Coetzee and Pater 2011).

### 3. n-insertion in existing Korean words

#### 3.1 Data and experimental procedure

In this section, I explore n-insertion in existing Korean words by investigating the results of the surveys on native speakers of two dialects of Korean, Seoul and Kyungsang. For Seoul Korean data, I use the results of the survey conducted by Jun (2015). In that survey, twenty-two Seoul Korean speakers participated. For Kyungsang Korean data, I performed a survey on Kyungsang Korean speakers, using the same method and the same word set as adopted by Jun.

Twenty-three paid Kyungsang Korean speakers participated in the survey (mean age = 25.9 years, eleven females and twelve males). Most of the participants ( $n = 20$ ) were raised in Daegu, and the rest ( $n = 3$ ) in other parts of Northern Kyungsang province. Sixteen participants were recruited from the community at Seoul National University through public advertising, and they took the survey in a quiet room (for about 40–50 minutes). The rest of the participants ( $n = 7$ ), who were recruited through personal referrals, completed and returned the survey via email.

304 multi-morphemic Korean words with /j/-initial  $M_2$  were employed as test words. These formed an exhaustive set of words with /j/-initial  $M_2$  that Jun classified as multi-morphemic in a pilot investigation of his dictionary database.<sup>4</sup> No test words included /i/-initial  $M_2$ , mainly because it was clear from the results of Jun's investigation of the data drawn from a dictionary and from previous surveys (Choi 2002, Kim 2003, Kook et al. 2005) that pre-/i/ insertion is attested, but less frequent in existing Korean words. Most of the test words were nouns ( $n = 286$ ), and the exceptions included thirteen verbs and five adverbs.

A single survey form was used, and thus the tasks were administered in the same order for all participants. In the survey form, each test word along with its inserted and non-inserted forms were presented in standard Korean orthography.<sup>5</sup> For each test word, the participants were instructed to choose their pronunciation from the following three (or two) options:

(6) Options given in the survey form<sup>6</sup>

	e.g. (i)	com-jak 'mothball'	(ii) sotok-jak 'antiseptic'	(iii) t <sup>h</sup> aŋ-jak 'herbal decoction'
a. insertion		com.njak	so.toŋ.njak	t <sup>h</sup> aŋ.njak
b. non-insertion (resyllabified)		co.mjak	so.to.kjak	—
c. non-insertion (aligned)		com.jak	sotok.jak	t <sup>h</sup> aŋ.jak

<sup>4</sup> Jun's database consists of words which occurred at least once in the Sejong corpus and were also listed as standard Korean words in the SKD.

<sup>5</sup> For the test words ( $n = 114$ ) which might be potentially ambiguous or unclear in meaning, their dictionary meanings were also presented in the survey form.

<sup>6</sup> In Korean orthography, the phonemic characters are grouped into blocks, each of which corresponds to a syllable. Both syllable divisions and constituency such as onset and coda can be seen in the written words.

When  $C_1$  was not /ŋ/ as in (6i-ii), two non-inserted forms (6b,c) were given in addition to an inserted one (6a). The resyllabified form (6b) is known as the standard surface form in Korean where an intervocalic consonant occupies an onset position. But, it has been argued in the literature (Lee 1992, Park 2001) that the aligned form (6c) is also possible when a morpheme boundary intervenes between a consonant and a following vocoid. When  $C_1$  was /ŋ/ as in (6iii), only a single non-inserted option (6c) was given since there is no way to represent the resyllabified option (6b) in the standard Korean orthography, and it is generally assumed in Korean phonology that [ŋ] cannot be syllabified (at least exclusively) in onset position. When  $C_1$  was underlyingly an obstruent as in (6ii), the option (6a) with insertion included nasalization of  $C_1$  since obstruent nasalization is obligatory in Korean, as stated in section 2. The participants were allowed to choose more than one, each of which was counted as one data point in the analysis. Also, when their pronunciation was not given as an option, they were asked to write it in the blank space on the survey form. (See Jun (2015) for more details of the methods and the reason for conducting a self-evaluation survey using written forms.)

In order to construct a database for the analysis of n-insertion in existing Korean words, I first combined results of Jun’s survey on Seoul Korean speakers and the results of the present survey on Kyungsang speakers. The resulting data will be analyzed and discussed, using a statistical model and insertion rates (i.e. number of *insertion* responses divided by total number of responses).

I conducted data screening before the main analysis. Responses to certain sets of test words would cause difficulty in the analysis of the effects of the factors, presented above, on Korean n-insertion. First, I excluded the responses to a test word with a compound marker in  $C_1$  position (/twi-s-jeki/ ‘backbiting (back-COMPOUND MARKER-story)’) from the database. The phonological nature of the compound marker is controversial in the literature on Korean phonology and morphology: /s/, /t/, /ʔ/, empty skeletal slot and others.<sup>7</sup> It is thus unclear which category the compound marker belongs to. It should probably not be classified as either obstruent or sonorant, although it is represented with the letter <s> in Korean orthography.

An additional problematic set of test words has the same  $M_2$ -initial suffix sequence /jʌ/, a contracted form of a causative suffix /i/ and a converbal suffix /ʌ/: e.g. /cuk-jʌ-cu-ta/ ‘be killing (die-CAUS.CVB-give-INF)’.<sup>8</sup> This suffix sequence is the only native Korean  $M_2$ -initial suffix included in the test word set. There is almost no variation among responses to the words with it. Most Korean participants did not choose the *insertion* option for them. The insertion rates for the ten test words with /jʌ/ are either zero or close to it: 3 words (0%), 4 words (4%), 2 words (6%), 1 word (13%). Thus, it seems that n-insertion is blocked almost categorically before this suffix. In the analysis of the database, I excluded responses to the test words with /jʌ/ so that they cannot weaken the effects of the factors adopted for investigation. The resulting data consist of 13,484 responses to 293 test words (285 nouns, five adverbs and three verbs).

### 3.2 Raw data analysis

This section considers only overall patterns of the data by calculating the mean insertion rates for each of the factors adopted in the experiment. In the next section, I provide a statistical analysis of the same data.

#### 3.2.1 Dialect

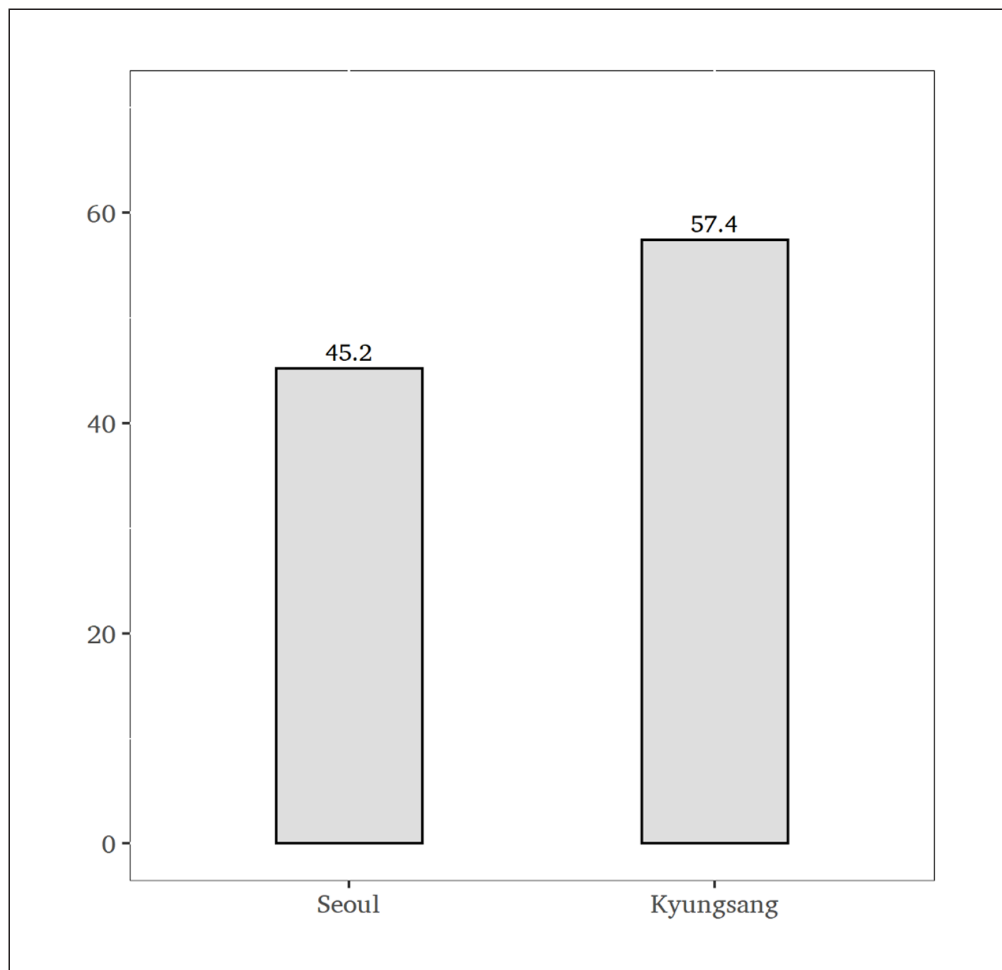
The mean insertion rate for the entire data is 51.4%, and the rate is, on average, higher for Kyungsang speakers than for Seoul speakers, as shown in [Table 1](#) and [Figure 1](#).

Dialect	rates
Seoul	45.2
Kyungsang	57.4

**Table 1** Insertion rate (%) by dialect.

<sup>7</sup> See Jun (2018) for a recent review of relevant literature.

<sup>8</sup> /jʌ/ is derived through glide formation of /i/ occurring before a vowel, /ʌ/ here.



**Figure 1** Insertion rate (%) by dialect.

The observed higher rate of insertion for Kyungsang speakers is consistent with previous studies (section 2.1.1), but it will be shown that the observed difference is not statistically significant.

### 3.2.2 Morphological category

To establish the effect of  $M_1$  morphology, insertion rates were calculated for different morphological categories of  $M_1$ , as shown in [Table 2](#).

$M_1$ morphology	rates
root	33.1 (58)
stem	57.3 (215)
prefix	41.2 (20)

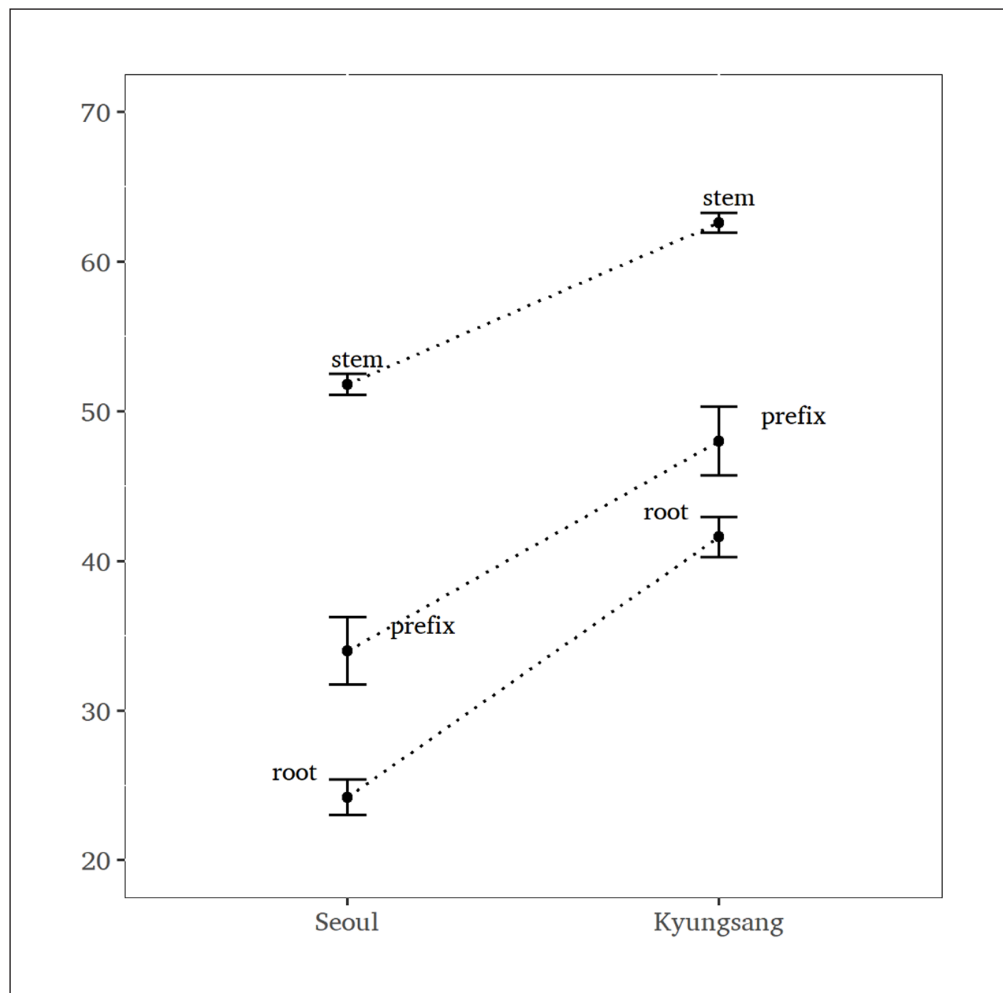
**Table 2** Insertion rate (%) by  $M_1$  morphology. (The number of corresponding words is shown in parentheses.)

Insertion rate was lowest for words with a bound root  $M_1$ , highest for those with a free stem  $M_1$ , and intermediate for those with a prefix  $M_1$ . This observed ranking by  $M_1$  morphological category, root < prefix < stem, did not differ depending on the dialect of the participants, as shown in [Table 3](#) and [Figure 2](#).

$M_1$ morphology	Seoul	Kyungsang
root	24.2	41.6
stem	51.8	62.6
prefix	34.0	48.0

**Table 3** Insertion rate (%) by  $M_1$  morphology and dialect.





**Figure 2** Insertion rate (%) by M<sub>1</sub> morphology and dialect.

In addition, to establish the effect of M<sub>2</sub> morphology, insertion rates were calculated for different morphological categories of M<sub>2</sub>, as shown in [Table 4](#).

M <sub>2</sub> morphology	rates
root	47.3 (62)
stem	51.3 (208)
suffix	64.7 (23)

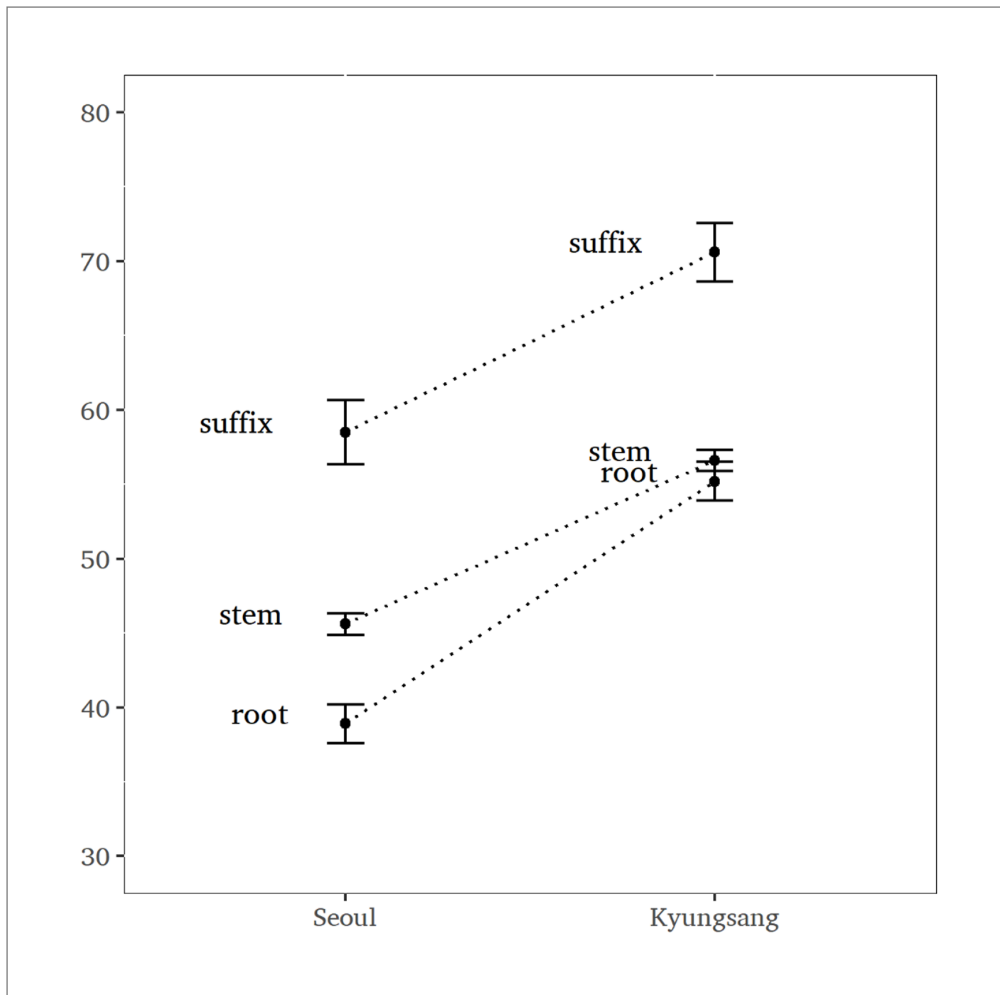
**Table 4** Insertion rate (%) by M<sub>2</sub> morphology. (The number of corresponding words is shown in parentheses.)

Insertion rate was lowest for words with a bound root M<sub>2</sub>, highest for those with a suffix M<sub>2</sub>, and intermediate for those with a free stem M<sub>2</sub>. As can be seen in [Table 5](#) and [Figure 3](#), the observed ranking by M<sub>2</sub> morphological category, root < suffix < stem, did not differ across dialects.

M <sub>2</sub> morphology	Seoul	Kyungsang
root	38.9	55.2
stem	45.6	56.6
suffix	58.5	70.6

**Table 5** Insertion rate (%) by M<sub>2</sub> morphology and dialect.

In summary, n-insertion was more likely with a free stem or an affix than with a bound root, regardless of whether it precedes or follows the insertion site. In addition, the relative frequencies of n-insertion by morphological category did not differ by dialect.



**Figure 3** Insertion rate (%) by M<sub>2</sub> morphology and dialect.

### 3.2.3 Etymology

To establish the effect of M<sub>1</sub> etymology, insertion rates were calculated for different etymological origins of M<sub>1</sub> morphemes, as shown in [Table 6](#).

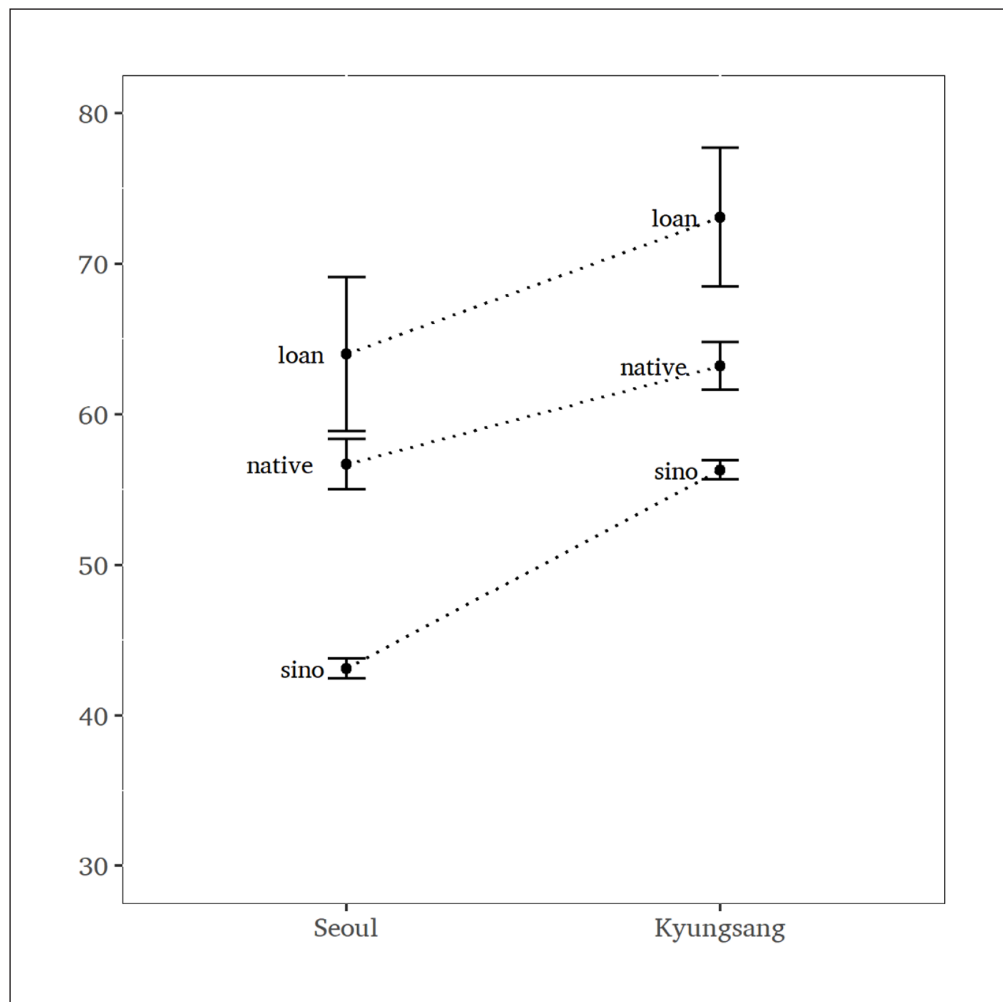
M <sub>1</sub> etymology	rates
sino	49.8 (250)
native	60.0 (39)
loan	68.7 (4)

**Table 6** Insertion rate (%) by M<sub>1</sub> etymology. (loan = loanword)  
 (The number of corresponding words is shown in parentheses.)

Insertion rate was, on average, lower when M<sub>1</sub> is Sino-Korean than native Korean. Although the rate was highest for words with a loanword, the test word set has only four words with a loanword M<sub>1</sub>. As can be seen in [Table 7](#) and [Figure 4](#), the observed ranking by M<sub>1</sub> etymology, Sino-Korean < native Korean (< loanword), did not differ by dialect.

M <sub>1</sub> etymology	Seoul	Kyungsang
sino	43.1	56.3
native	56.7	63.2
loan	64.0	73.1

**Table 7** Insertion rate (%) by M<sub>1</sub> etymology and dialect.



**Figure 4** Insertion rate (%) by M<sub>1</sub> etymology and dialect.

In addition, to establish the effect of M<sub>2</sub> etymology, insertion rates were calculated for different etymological origins of M<sub>2</sub> morphemes, as shown in [Table 8](#).

M <sub>2</sub> etymology	rates
sino	50.4 (267)
native	62.3 (26)

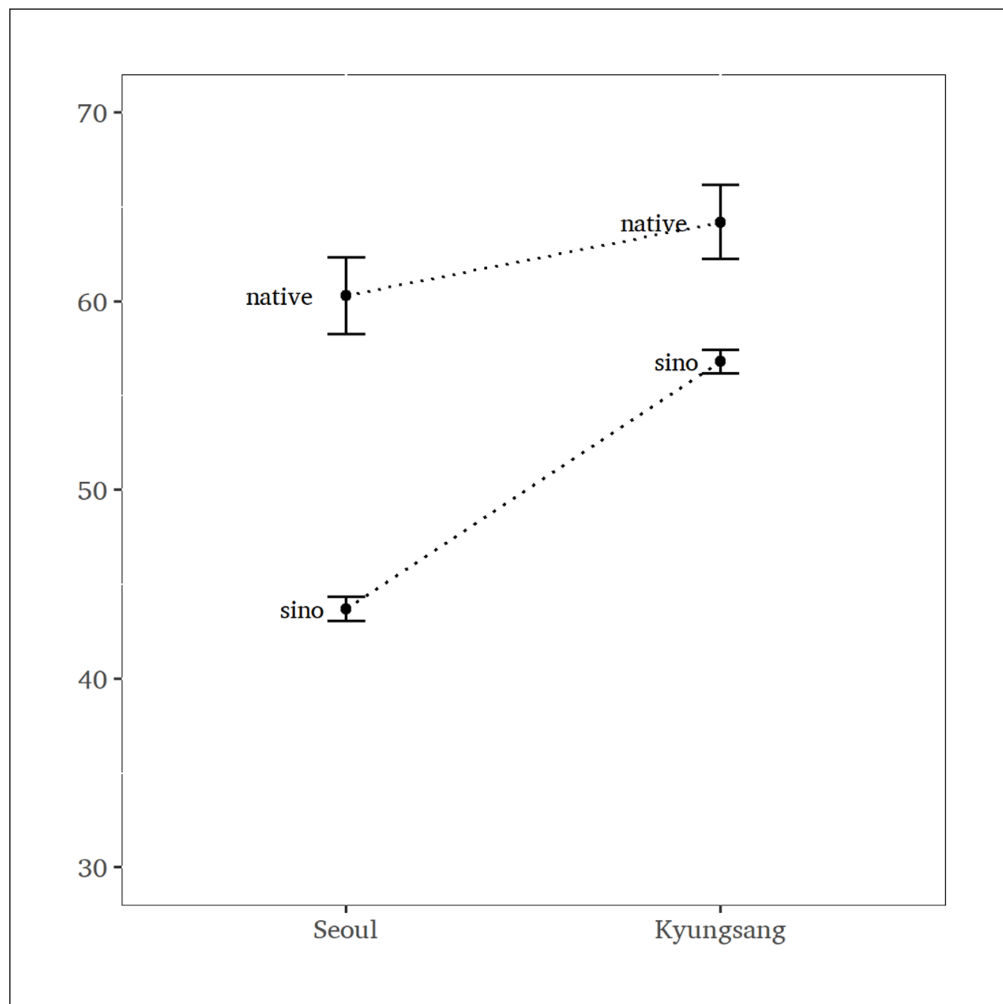
**Table 8** Insertion rate (%) by M<sub>2</sub> etymology. (The number of corresponding words is shown in parentheses.)

Similarly to the results for M<sub>1</sub> etymology, insertion rate was lower when M<sub>2</sub> is Sino-Korean than native Korean. As can be seen in [Table 9](#) and [Figure 5](#), the observed relative rate difference between words with native and Sino-Korean M<sub>2</sub> did not differ by dialect.

M <sub>2</sub> etymology	Seoul	Kyungsang
sino	43.7	56.8
native	60.3	64.2

**Table 9** Insertion rate (%) by M<sub>2</sub> etymology and dialect.

In summary, n-insertion was less likely with a Sino-Korean morpheme than with a native Korean (or loanword), regardless of whether it precedes or follows the insertion site. This tendency was true for both Seoul and Kyungsang speakers.



**Figure 5** Insertion rate (%) by M<sub>2</sub> etymology and dialect.

### 3.2.4 C<sub>1</sub> type

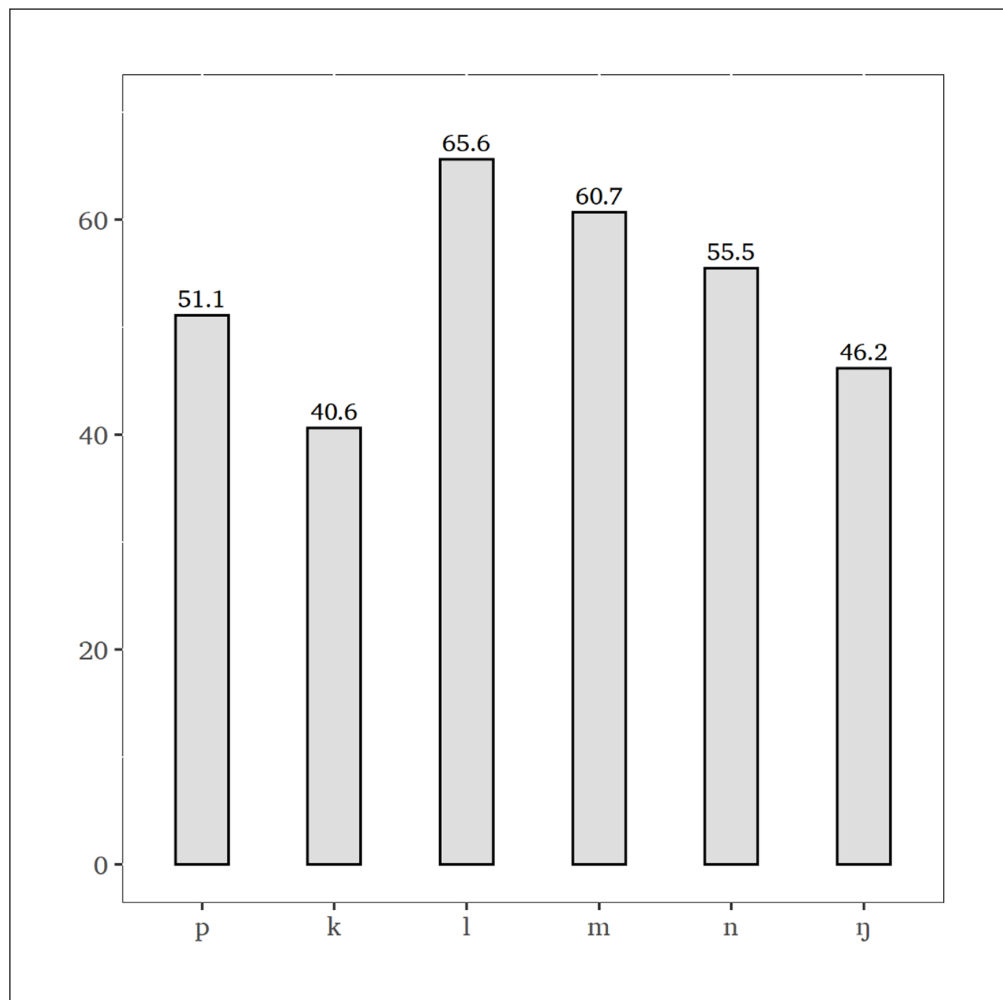
To establish how n-insertion varies depending on the type of C<sub>1</sub>, insertion rates were calculated for different C<sub>1</sub> consonants, as shown in [Table 10](#).

C <sub>1</sub> type	C <sub>1</sub>	rates
obstruent	c	55.6 (1)
	t <sup>h</sup>	64.4 (1)
	p <sup>h</sup>	51.1 (1)
	s	58.1 (3)
	p	51.1 (15)
	k	40.6 (48)
sonorant	l	65.6 (35)
	m	60.7 (18)
	n	55.5 (78)
	ŋ	46.2 (93)

**Table 10** Insertion rate (%) by C<sub>1</sub> consonants. (The number of corresponding words is shown in parentheses.)

Excluding consonants with extremely low frequency (less than or equal to 3), insertion rates are plotted by C<sub>1</sub> consonants in [Figure 6](#).

Insertion rate was lower after an obstruent or velar nasal than after a sonorant other than /ŋ/, as shown in [Table 11](#).



**Figure 6** Insertion rate (%) by C<sub>1</sub> consonants.

C <sub>1</sub> type	rates
obs	44.2 (69)
son	58.9 (131)
ŋ	46.1 (93)

**Table 11** Insertion rate (%) by C<sub>1</sub> type. (obs = obstruent, son = sonorant except /ŋ/; The number of corresponding words is shown in parentheses.)

The observed lower insertion rates with an obstruent or /ŋ/ C<sub>1</sub> are true for both Seoul and Kyungsang speakers, as can be seen in [Table 12](#) and [Figure 7](#).

C <sub>1</sub> type	Seoul	Kyungsang
obs	42.8	45.5
son	51.1	66.3
ŋ	38.6	53.5

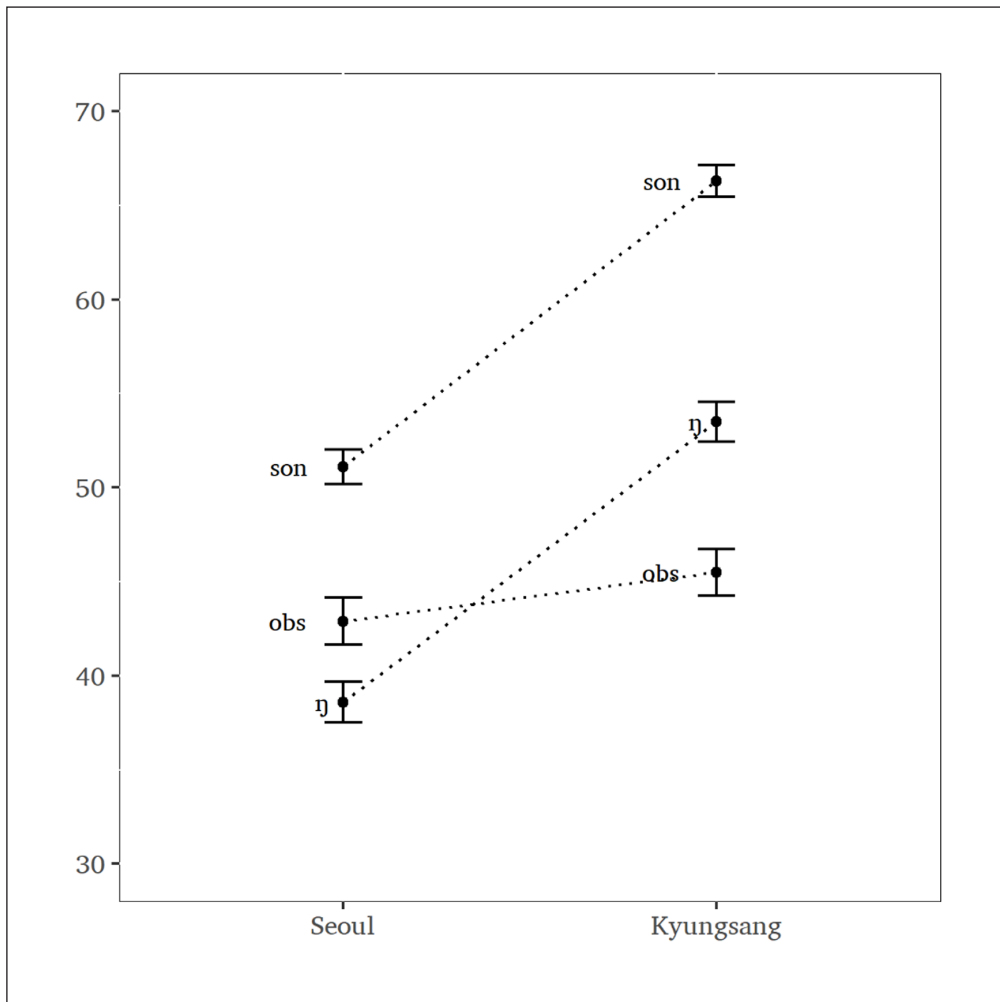
**Table 12** Insertion rate (%) by C<sub>1</sub> type and dialect.

The relative rate difference between words with an obstruent and velar nasal C<sub>1</sub> was switched between the participants of the two dialects: obstruent > /ŋ/ for Seoul, but obstruent < /ŋ/ for Kyungsang.

In summary, n-insertion was less likely after an obstruent or velar nasal than after other sonorant consonants, regardless of the dialect of the participants.

### 3.2.5 Length

To establish how n-insertion varies depending on the length of M<sub>1</sub>, insertion rates were calculated for different numbers of component syllables of M<sub>1</sub>, as shown in [Table 13](#).



**Figure 7** Insertion rate (%) by C<sub>1</sub> type and dialect.

number of M <sub>1</sub> syllables	rates
1	38.9 (130)
2	61.1 (154)
3	71.7 (7)
4	31.9 (1)
5	73.3 (1)

**Table 13** Insertion rate (%) by the number of syllables of M<sub>1</sub>. (The number of corresponding words is shown in parentheses.)

Insertion rate was higher when M<sub>1</sub> is disyllabic than monosyllabic. The insertion rate is even higher when M<sub>1</sub> is trisyllabic, suggesting that as the M<sub>1</sub> length increases, the more likely n-insertion is. But, since words with trisyllabic syllables or longer are relatively few in the test set, I take the above rate distribution as indicating the difference between words with mono and polysyllabic M<sub>1</sub>, collapsing the results for words with polysyllabic M<sub>1</sub>, as shown in [Table 14](#).

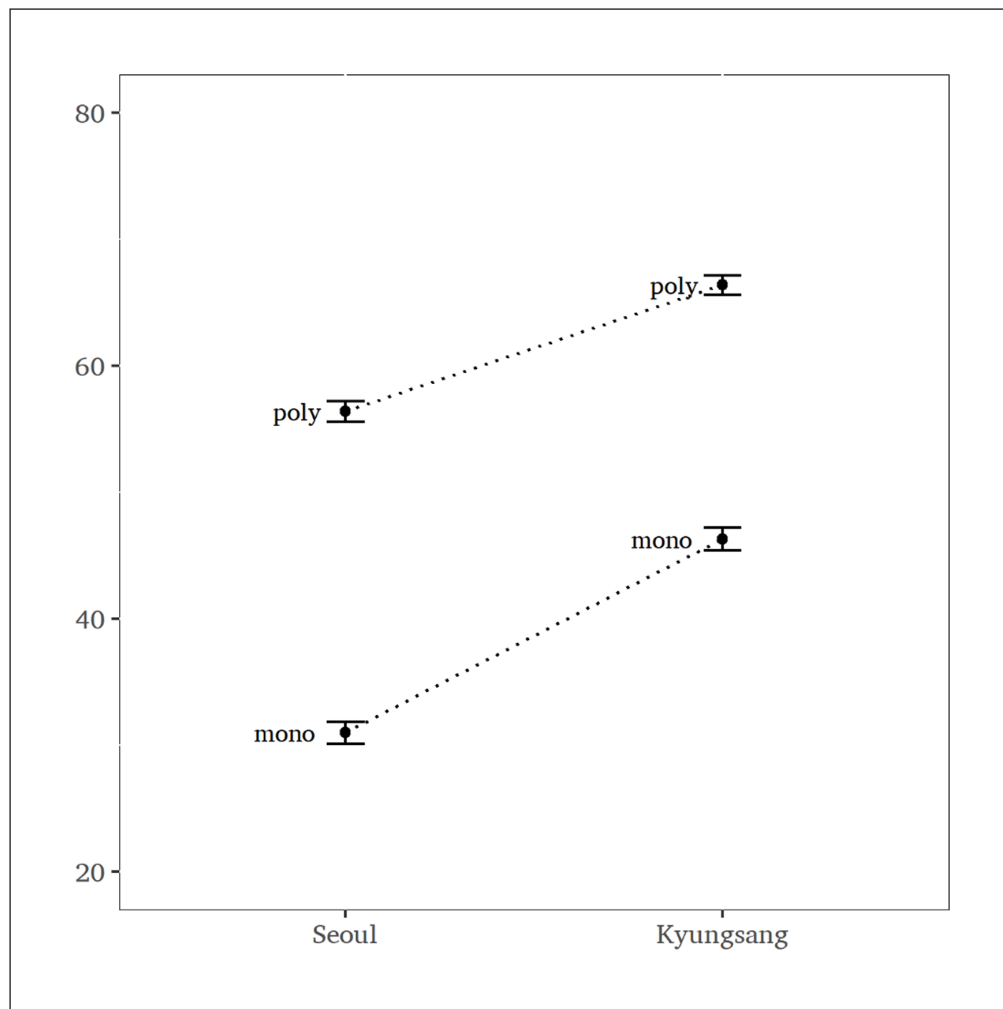
M <sub>1</sub> length	rates
monosyllabic	38.9 (130)
polysyllabic	61.5 (163)

**Table 14** Insertion rate by (%) M<sub>1</sub> length. (The number of corresponding words is shown in parentheses.)

The observed higher rate of insertion with polysyllabic M<sub>1</sub> is true for both Seoul and Kyungsang participants, as shown in [Table 15](#) and [Figure 8](#).

	Seoul	Kyungsang
monosyllabic	31.0	46.3
polysyllabic	56.4	66.4

**Table 15** Insertion rate (%) by M<sub>1</sub> length and dialect.



**Figure 8** Insertion rate (%) by  $M_1$  length and dialect (poly = polysyllabic).

In addition, to establish how n-insertion varies depending on the length of  $M_2$ , insertion rates were calculated for different numbers of component syllables of  $M_2$ , as shown in [Table 16](#).

number of $M_2$ syllables	rates
1	51.2 (148)
2	51.3 (131)
3	58.2 (12)
4	40.8 (2)

**Table 16** Insertion rate (%) by the number of syllables of  $M_2$ . (The number of corresponding words is shown in parentheses.)

Insertion rate does not seem to vary systematically with the length of  $M_2$ . There is almost no difference in average insertion rate between words with mono and disyllabic  $M_2$ . Words with trisyllabic  $M_2$  show somewhat higher rates, but words with trisyllabic  $M_2$  or longer are relatively few in the test set. Thus, as in the analysis of  $M_1$  length effects, I collapsed the results of all words with polysyllabic  $M_2$  into a single category. Its average insertion rate is 51.8%, which is not substantially different from the rate for words with monosyllabic  $M_2$ , i.e. 51.2%, as shown in [Table 17](#).

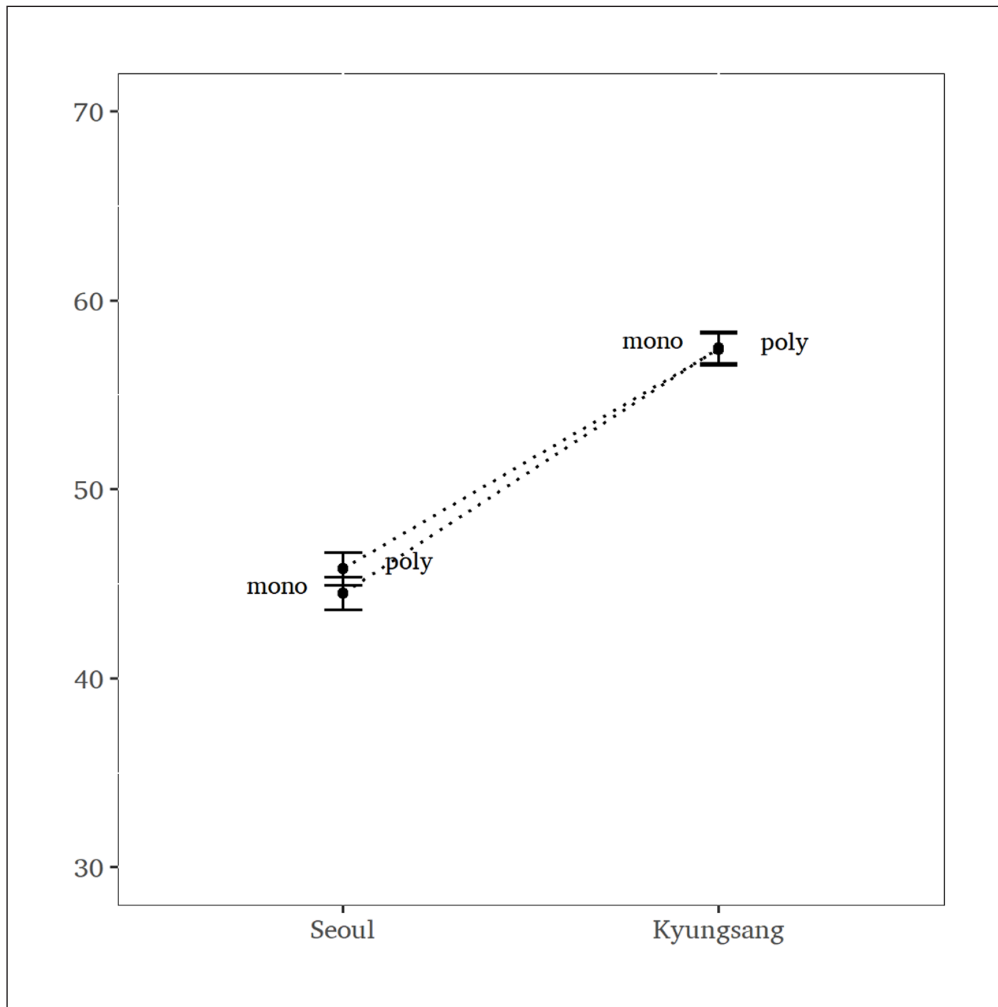
$M_2$ length	rates
monosyllabic	51.2 (148)
polysyllabic	51.8 (145)

**Table 17** Insertion rate (%) by  $M_2$  length. (The number of corresponding words is shown in parentheses.)

The rate (non-)difference between words with mono and polysyllabic  $M_2$  did not vary much by dialect, as shown in [Table 18](#) and [Figure 9](#).

	Seoul	Kyungsang
polysyllabic	45.8	57.4
monosyllabic	44.5	57.5

**Table 18** Insertion rate (%) by  $M_2$  length and dialect.



**Figure 9** Insertion rate (%) by  $M_2$  length and dialect.

In summary, insertion rates differed according to the length of  $M_1$ , not  $M_2$ , regardless of the dialect of the participants. n-insertion was more likely with a polysyllabic  $M_1$  than with a monosyllabic  $M_1$ .

### 3.2.6 $V_2$ height

To establish the effect of the quality of a vowel following /j/, called  $V_2$ , insertion rates were calculated for different  $V_2$  vowels, as shown in [Table 19](#) and [Figure 10](#).

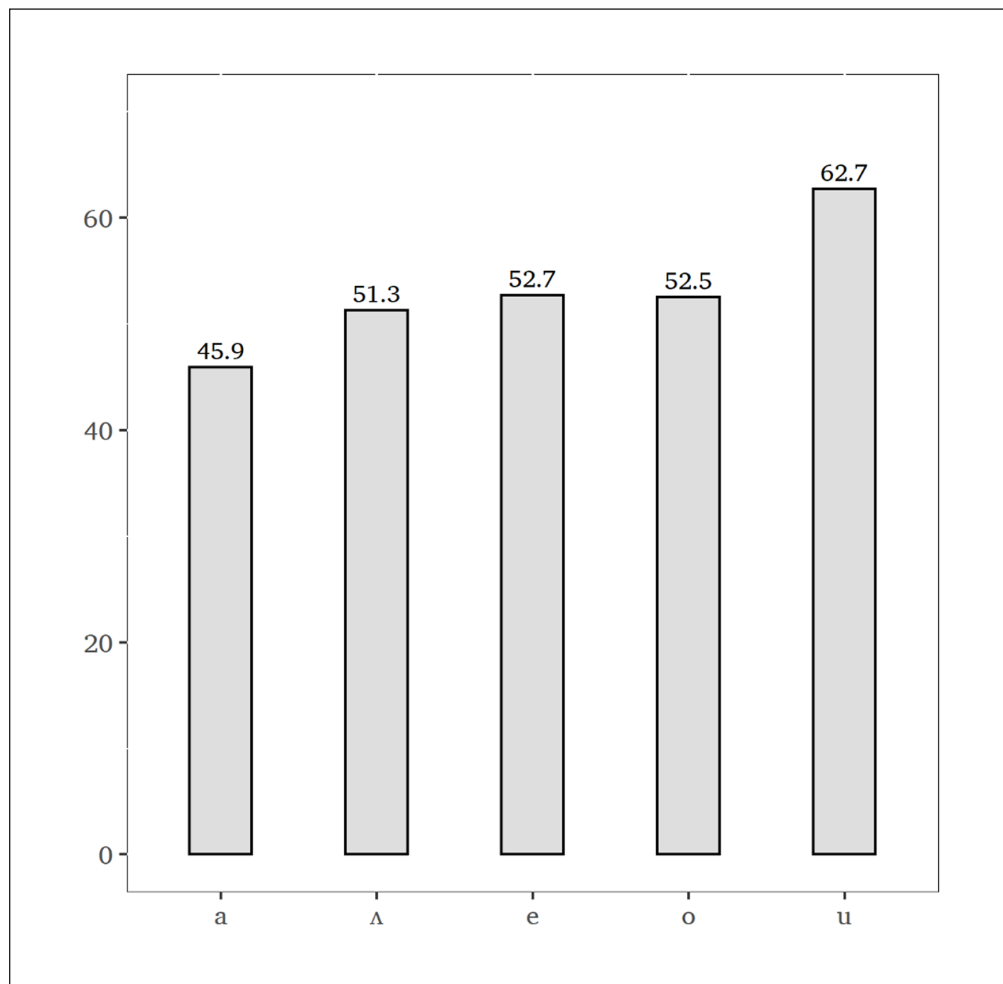
$V_2$	rates
a	45.9 (73)
ʌ	51.2 (139)
e	52.6 (4)
o	52.4 (43)
u	62.7 (34)

**Table 19** Insertion rate (%) by  $V_2$  vowels. (The number of corresponding words is shown in parentheses.)

Insertion rate was higher when  $V_2$  is /u/, namely, a high vowel (62.7%), than when  $V_2$  is non-high (49.9%), as shown in [Table 20](#).<sup>9</sup>

<sup>9</sup> In Korean, /u/ is the only high vowel which can follow /j/ unlike the other high vowels, /i, i/.





**Figure 10** Insertion rate (%) by  $V_2$  vowels.

$V_2$ height	rates
non-high	49.9 (259)
high	62.7 (34)

**Table 20** Insertion rate (%) by  $V_2$  height.

The observed higher rate of n-insertion with a high  $V_2$  is true for both Seoul and Kyungsang participants, as can be seen in [Table 21](#) and [Figure 11](#).

$V_2$ height	Seoul	Kyungsang
non-high	43.6	56.1
high	57.6	67.7

**Table 21** Insertion rate (%) by  $V_2$  height and dialect.

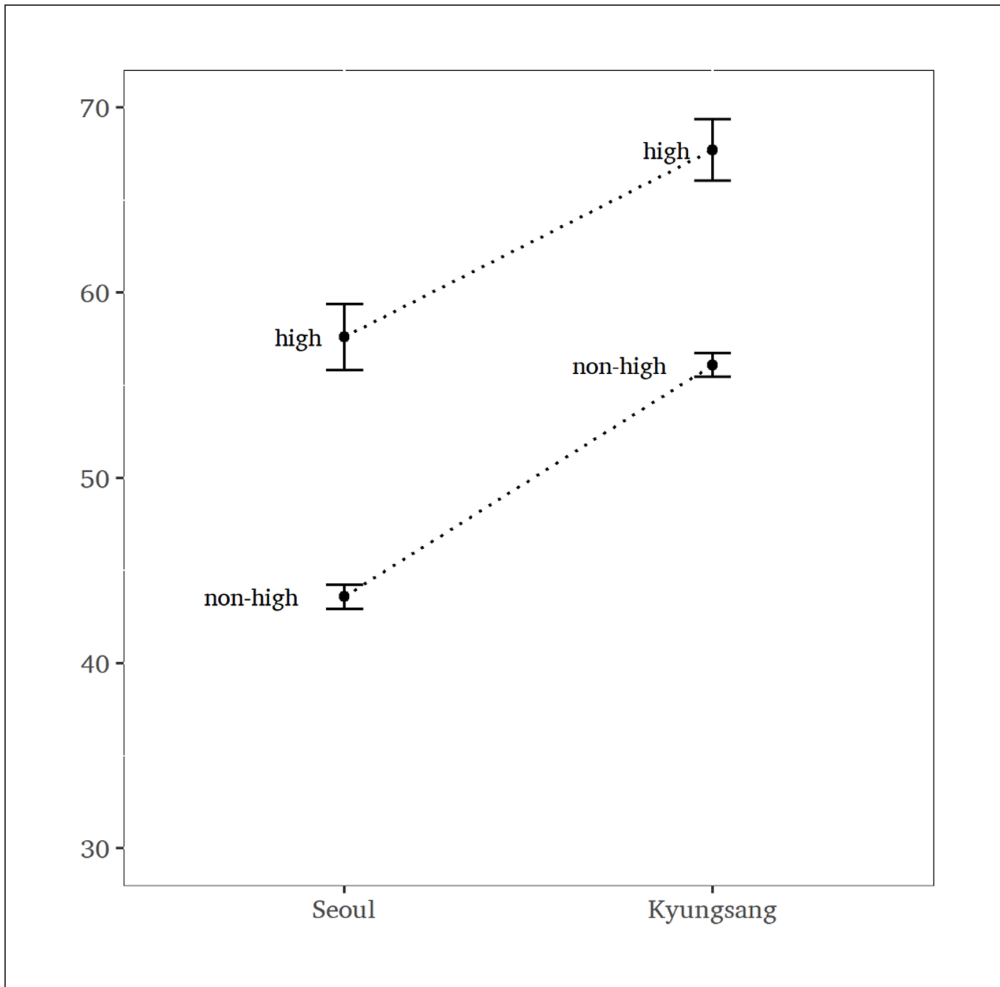
In summary, n-insertion was more likely before a high vowel than before non-high vowels. This observed tendency was true for both Seoul and Kyungsang participants.

### 3.2.7 Frequency

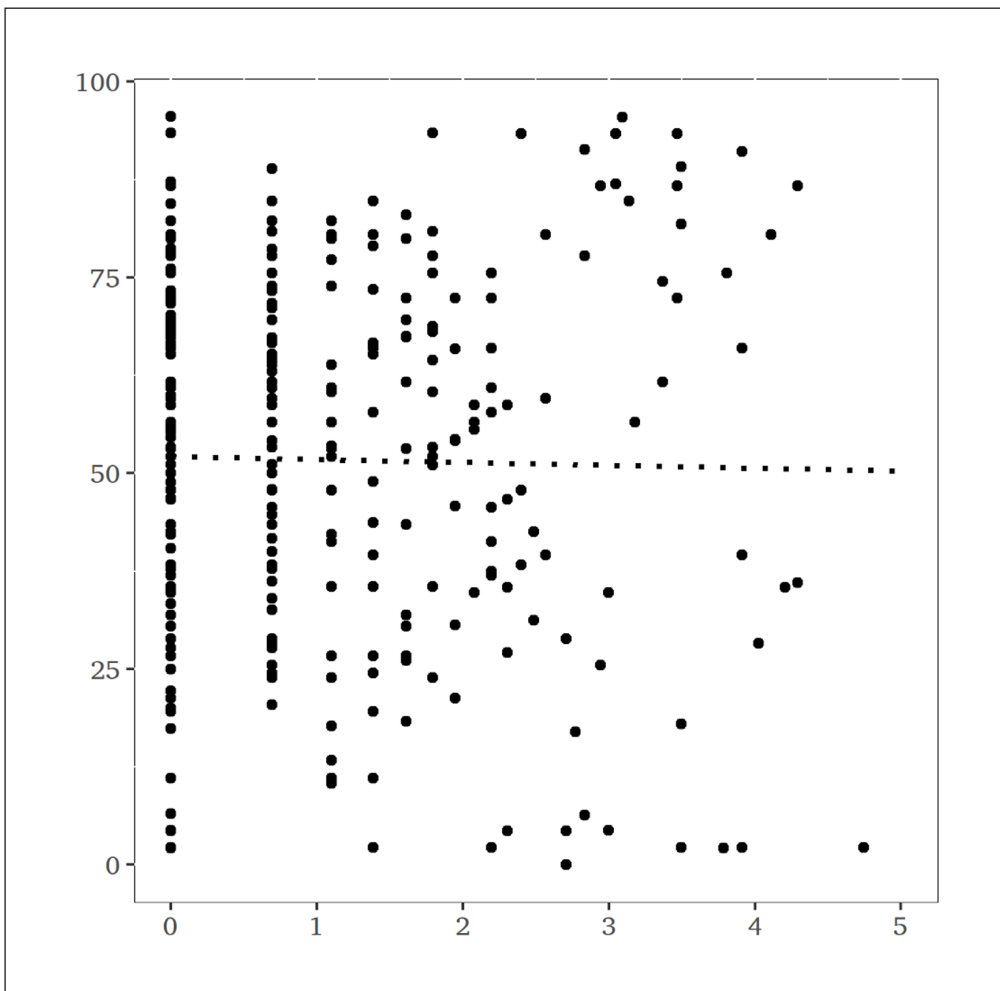
To establish how n-insertion varies depending on the frequency of words, insertion rates were calculated for different (log-transformed) frequencies, and plotted in [Figure 12](#).

Here we use word frequencies in the Sejong corpus, reported in Kang and Kim (2004). It seems that frequencies played no role in determining the rate of n-insertion, as suggested by an almost flat regression line in [Figure 12](#), and by a statistically insignificant low correlation between insertion rates and (log-transformed) frequencies ( $r(291) = -0.040$ ,  $p = 0.487$ ).

The observed lack of frequency effect did not differ depending on the dialect of the participants, as can be seen in (7).



**Figure 11** Insertion rate (%) by  $V_2$  height and dialect.



**Figure 12** Insertion rate (%) by (log-transformed) frequency (n = 293).

dialect	Correlations
Seoul	$r(291) = -0.035, p = 0.552$
Kyungsang	$r(291) = -0.047, p = 0.419$

### 3.3 Mixed effects analysis

In the previous section, we have considered only mean insertion rates, abstracting away from differences of individual participants and words. Here, we provide a more stringent test by taking into consideration individual participant and word differences. A mixed effects logistic regression model was fitted with the `glmer` function from the `lmerTest` package (Kuznetsova et al. 2017) in R (R Core Team 2020). Dependent variable is binary, i.e. n-inserted or not. Each subject and each test word were included as random intercepts.<sup>10</sup>

In the model on the data of existing words, the following independent variables were adopted.<sup>11</sup> (sel = Seoul, ks = Kyungsang)

- dialect (sel, ks)
- $M_1$  morphology (root, stem, prefix)
- $M_2$  morphology (root, stem, suffix)
- $M_1$  origin (sino, non-sino (native and loan))
- $M_2$  origin (sino, native)
- $M_1$  length (mono, poly)
- $M_2$  length (mono, poly)
- $C_1$  type (obs, son, η)
- $V_2$  height (high, non-high)
- log-transformed token frequency (frequency)

The interactions of dialect with each of all the remaining variables were included to address the question of whether the effect of each variable varies across dialects. The following interactions were additionally included in the model based on the results of a CART (Classification and Regression Tree) analysis of each of Seoul and Kyungsang Korean data using the `rpart` package (Therneau & Atkinson 2019).<sup>12</sup>

- 2-way interaction between  $M_1$  origin and  $M_1$  length
- 3-way interaction between dialect,  $M_1$  length and frequency
- 3-way interaction between dialect,  $M_1$  length and  $C_1$  type
- 3-way interaction between dialect,  $M_1$  origin and  $C_1$  type

Categorical factors with three levels,  $M_{1,2}$  morphology and  $C_1$  type, were forward-difference coded in order to compare not only between the first and second levels (e.g. root vs. stem), but also between the second and third levels (e.g. stem vs. affix). All the remaining categorical factors were sum-coded so that coefficient estimates would represent the main effect.

The random effects of the mixed effects model show large variation depending on the participant (variance = 1.178) and test word (variance = 0.886), meaning that different participants and test words show greater and lesser average rates of insertion, as can be seen in [Figures 13](#) and [14](#), respectively.

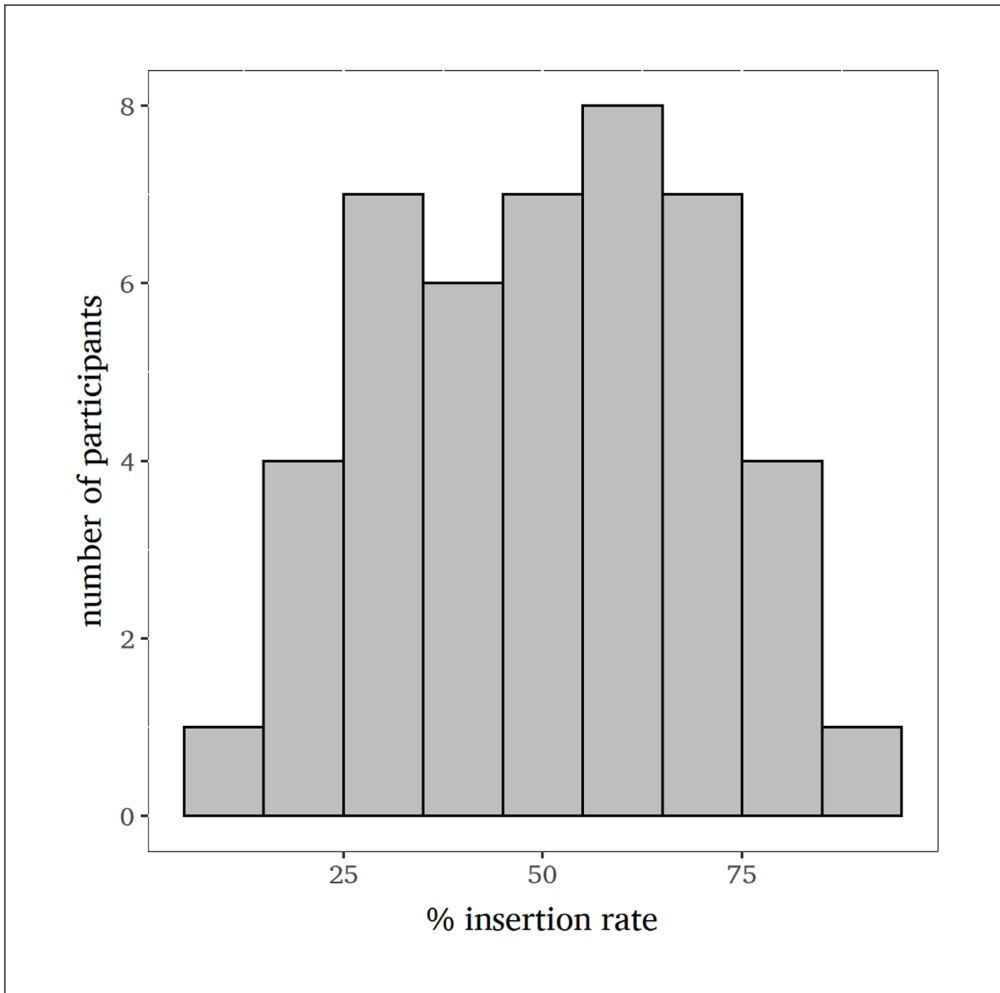
Even once these random factors are taken into account, as can be seen in (8), many fixed factors and their interactions still have a sizable effect on the insertion rate, which holds independently of the specific participant and test word.

<sup>10</sup> More complex models with random slopes failed to converge.

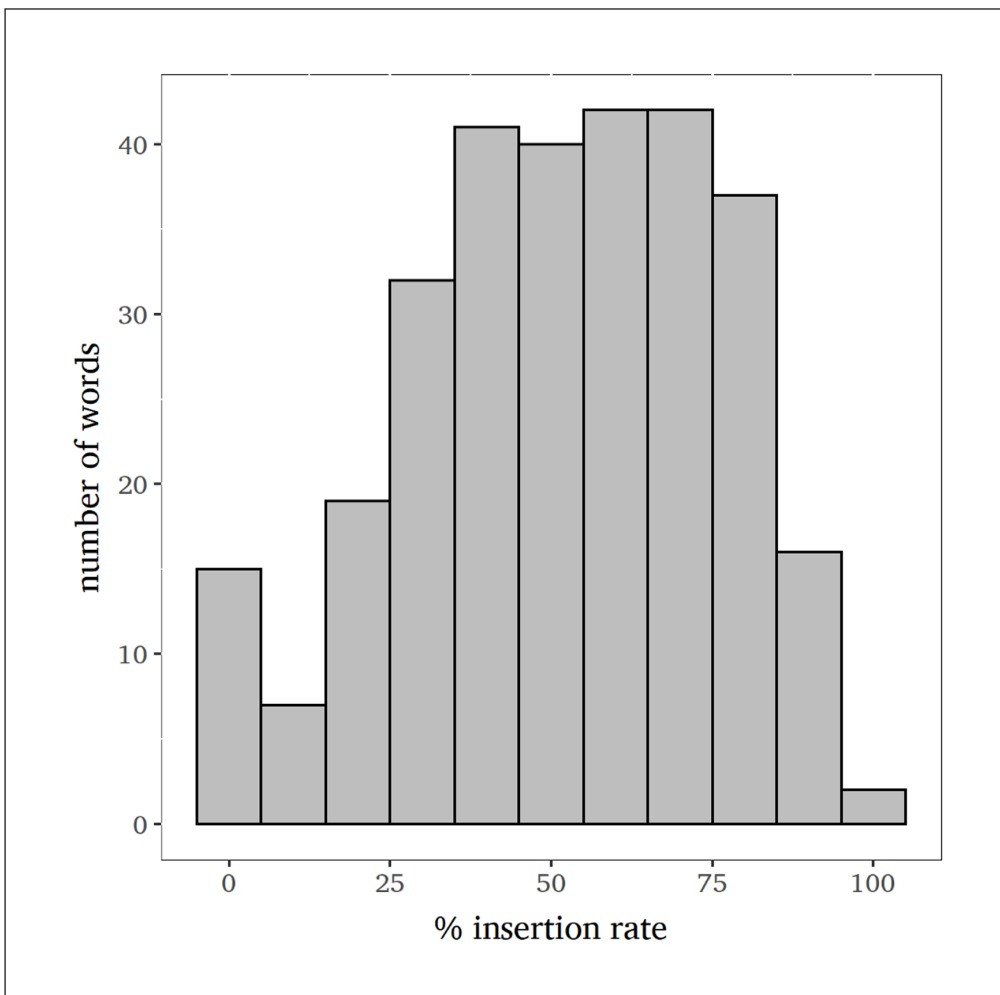
<sup>11</sup> In the morphological category distinction, stem refers to a morpheme which can stand alone whereas root a morpheme which cannot. An exception to this criterion is that a reduplicated adverb (e.g. /jakim-jakim/ [jakimjakim] ~ [jakimnjakim] 'bit by bit'), which is known to undergo n-insertion in the literature on Seoul Korean n-insertion, consists of stems. The distinction between root and affix was based on the SKD.

Recall in section 3.2.3 that there were only four words with loanword  $M_1$  in the test word set, and their insertion rates were more similar to those with native Korean  $M_1$  than to those with Sino-Korean  $M_1$ . Here, I collapse loanword and native Korean  $M_1$  morphemes into a single category.

<sup>12</sup> Thanks to an anonymous reviewer for suggesting this way of selecting potentially significant interactions.



**Figure 13** The numbers of participants for different insertion rates (bin width = 10).



**Figure 14** The numbers of test words for different insertion rates (bin width = 10).

	Estimate	Pr(>  z )
(Intercept)	-0.141	0.584
dialect (ks)	0.209	0.237
M <sub>1</sub> .morphology (root–stem)	-0.6	< 0.001***
M <sub>1</sub> .morphology (stem–prefix)	-0.102	0.709
M <sub>2</sub> .morphology (root–stem)	-0.653	< 0.001***
M <sub>2</sub> .morphology (stem–suffix)	-0.062	0.809
M <sub>1</sub> .origin (sino)	0.009	0.945
M <sub>2</sub> .origin (sino)	-0.219	0.091
M <sub>1</sub> .length (poly)	0.324	0.008**
M <sub>2</sub> .length (poly)	-0.048	0.518
C <sub>1</sub> type (obs–son)	-1.166	< 0.001***
C <sub>1</sub> type (son–ŋ)	1.099	< 0.001***
height (high)	0.283	0.005**
frequency	0.080	0.123
dialect (ks) : M <sub>1</sub> .morphology (root–stem)	0.097	0.143
dialect (ks) : M <sub>1</sub> .morphology (stem–prefix)	0.062	0.534
dialect (ks) : M <sub>2</sub> .morphology (root–stem)	0.091	0.156
dialect (ks) : M <sub>2</sub> .morphology (stem–suffix)	-0.06	0.526
dialect (ks) : M <sub>1</sub> .origin (sino)	0.073	0.109
dialect (ks) : M <sub>2</sub> .origin (sino)	0.089	0.052
dialect (ks) : M <sub>1</sub> .length (poly)	-0.024	0.531
dialect (ks) : M <sub>2</sub> .length (poly)	-0.004	0.899
dialect (ks) : C <sub>1</sub> type (obs–son)	-0.351	< 0.001***
dialect (ks) : C <sub>1</sub> type (son–ŋ)	0.157	0.086
dialect (ks) : height (high)	-0.008	0.834
dialect (ks) : frequency	0.011	0.576
M <sub>1</sub> .origin (sino) : M <sub>1</sub> .length (poly)	0.561	< 0.001***
dialect (ks) : M <sub>1</sub> .length (poly) : frequency	0.035	0.523
dialect (sel) : M <sub>1</sub> .length (poly) : frequency	0.124	0.026*
dialect (ks) : M <sub>1</sub> .length (poly) : C <sub>1</sub> type (obs–son)	0.485	0.007**
dialect (sel) : M <sub>1</sub> .length (poly) : C <sub>1</sub> type (obs–son)	0.320	0.081
dialect (ks) : M <sub>1</sub> .length (poly) : C <sub>1</sub> type (son–ŋ)	0.583	< 0.001***
dialect (sel) : M <sub>1</sub> .length (poly) : C <sub>1</sub> type (son–ŋ)	0.286	0.077
dialect (ks) : M <sub>1</sub> .origin (sino) : C <sub>1</sub> type (obs–son)	-0.202	0.420
dialect (sel) : M <sub>1</sub> .origin (sino) : C <sub>1</sub> type (obs–son)	0.005	0.984
dialect (ks) : M <sub>1</sub> .origin (sino) : C <sub>1</sub> type (son–ŋ)	-0.711	0.015*
dialect (sel) : M <sub>1</sub> .origin (sino) : C <sub>1</sub> type (son–ŋ)	-0.540	0.063

The next section discusses the results of the present survey focusing on the significant effects in the mixed effects model, just presented.

### 3.4 Significant effects

The significant main effects in the mixed effects model, shown in (8), are summarized in (9).

<sup>13</sup> The non-reference level of sum-coded binary factors and the two levels of forward-difference coded ternary factors relevant to the calculation of the corresponding estimate are shown in parentheses. For instance, the positive estimate for dialect (ks), 0.209, means that the average insertion rate for Kyungsang speakers is above the grand mean rate of insertion. The negative estimate for M<sub>1</sub>.morphology (root minus stem), -0.6, means that the average insertion rate is lower with a root M<sub>1</sub> than a stem M<sub>1</sub>.

- (9) Summary: significant main effects
- a.  $M_1$  morphology effect  
 n-insertion is less likely after a bound root than after a free stem (or prefix).
  - b.  $M_2$  morphology effect  
 n-insertion is less likely before a bound root than before a free stem (or suffix).
  - c. Obstruency effect  
 n-insertion is less likely after an obstruent  $C_1$ .
  - d. Velar nasal effect  
 n-insertion is less likely after a velar nasal  $C_1$ .
  - e. Length effect  
 n-insertion is more likely with a polysyllabic  $M_1$  than with a monosyllabic  $M_1$ .
  - f. Height effect  
 n-insertion is more likely with a high vowel following /j/.

All of these main effects can also be seen in the corresponding tables and plots presented in section 3.2. Many of them are in part or probabilistically consistent with previous studies on Korean n-insertion, discussed in section 2 and summarized in (4). Let us consider each of the significant main effects.

First, the  $M_1$  morphology effect (9a) suggests that a bound root  $M_1$  is less likely triggers of n-insertion than a free stem  $M_1$  in both dialects. This and other related findings are only in part consistent with previous studies on Korean n-insertion. As discussed in section 2.1, Han (1994) argues that n-insertion may occur after a free stem, regardless of the dialect, but dialectal variation occurs with words with a bound root  $M_1$ . It is argued that n-insertion is blocked after a bound root in Seoul, not Kyungsang, Korean. The results of the present study agree with Han’s argument in that n-insertion is more frequent after a free stem in Seoul Korean. However, the present study differs from the previous studies in that n-insertion in Seoul Korean was not completely blocked after a bound root  $M_1$ , as can be seen in [Table 3](#) and [Figure 2](#). Some test words with a bound root  $M_1$ , for which Seoul Korean participants applied n-insertion frequently, are shown in (10).

- (10) Some words with a bound root  $M_1$  with high insertion rates in Seoul Korean data

	word	rates (%)
a.	/cik-jukmjʌnc <sup>h</sup> e/ [cikjuŋmjʌnc <sup>h</sup> e] ‘rectangular cuboid’ ~ [ciŋjuŋmjʌnc <sup>h</sup> e]	90.5
b.	/ʌpsin-jʌkita/ [ʌpsinjʌkita] ‘despise’ ~ [ʌpsinnjʌkita]	85.7
c.	/namcon-jʌpi/ [namconjʌpi] ‘male dominance’ ~ [namconnjʌpi]	81.8
d.	/kunsin-juui/ [kunsinjuui] ‘one moral discipline in Confucianism’ ~ /kunsinnjuui/	60.8
e.	/moŋtaŋ-jʌnp <sup>h</sup> il/ [moŋtaŋjʌnp <sup>h</sup> il] ‘a pencil stub’ ~ [moŋtaŋjʌnp <sup>h</sup> il]	54.2

Consequently, the observed dialectal difference with respect to  $M_1$  morphology is consistent with a probabilistic version of Han’s argument on the difference between Seoul and Kyungsang Korean n-insertion.

Second, the  $M_2$  morphology effect (9b) is similar to the  $M_1$  morphology effect, just discussed, in that roots are less likely to trigger n-insertion than stems and affixes. Note that n-insertion occurred when  $M_2$  is a bound root (and affix), as can be seen in [Tables 4, 5](#) and [Figure 3](#). This observation is in conflict with the free  $M_2$  requirement proposed in many previous studies on Seoul Korean n-insertion, mentioned in section 2.1.2. Since all  $M_2$  roots and suffixes in the present data are Sino-Korean, this study provides no counter-examples to the native Korean specific version of the free  $M_2$  requirement. But, recall, in section 2.4, that cases of n-insertion before certain native Korean suffixes have already been reported in some previous studies. Consequently, the free  $M_2$  requirement cannot be an absolute condition for n-insertion in Korean words, whether native or Sino-Korean. But, it seems true that n-insertion is blocked categorically before certain native Korean bound morphemes, as suggested by the results of the present survey involving the causative-converb suffix sequence /jʌ/, discussed in section 3.1.

Third, insertion rates differed depending on the type of  $C_1$ . Two observed relevant effects, obstruency (9c) and velar nasal (9d), suggest that sonorant consonants other than /ŋ/ are frequent triggers of n-insertion. This is only in part consistent with the previous studies on Seoul Korean n-insertion (Cho 1995, Cho & Iverson 1997) arguing that n-insertion applies obligatorily after a sonorant consonant, and optionally after an obstruent. One difference of the present study from the previous studies is that n-insertion did not always apply in words with a sonorant  $C_1$ . As illustrated in (11), Seoul Korean participants rarely applied n-insertion in some words with a sonorant  $C_1$ .

(11) Some words with a sonorant  $C_1$  with low insertion rates in Seoul Korean data

	word			rates (%)
a.	/kan-jaŋm/	[kanjaŋm] ~ [kannjaŋm]	'hepatitis	13.0
b.	/san-ja/	[sanja] ~ [sanna]	'hills and fields'	17.3
c.	/san-jaŋ/	[sanjaŋ] ~ [sannaŋ]	'goat'	20.8
d.	/kil-jaŋsik/	[kiljaŋsik] ~ [killjaŋsik]	'food for travelling'	26.1
e.	/pol-jaŋci/	[poljaŋci] ~ [polljaŋci]	'blusher'	26.1

Thus, the observed obstruency effect is consistent with a probabilistic version of the previous argument on the difference between sonorant and obstruent  $C_1$  consonants. The velar nasal effect (i.e. /ŋ/ is not a likely trigger of n-insertion unlike other sonorant consonants) forms the second difference of the present study from the above-mentioned previous studies. The observed velar nasal effect is consistent with the results of Hwang's (2008) experimental study on Seoul Korean n-insertion.

Fourth, the length effect suggests that a polysyllabic  $M_1$  is more likely triggers of n-insertion than a monosyllabic  $M_1$ , regardless of the dialect of the participants. See below for a discussion of significant interaction effects of the length factor with some other factors.

Finally, insertion rates differed depending on the height of  $V_2$  vowels following the  $M_2$ -initial /j/. The height effect suggests that high  $V_2$  vowels are likely triggers of n-insertion than non-high vowels.

Some of the significant main effects, just discussed, interact with other factors in the present data. Significant interaction effects in the mixed effects model, shown in (8), are summarized in (12).

(12) Summary: significant interaction effects

- a. obstruency and dialect  
The obstruency effect is larger for Kyungsang speakers.
- b. obstruency, length, and dialect  
The obstruency effect in words with a polysyllabic  $M_1$  is smaller for Kyungsang speakers.
- c. velar nasal, length, and dialect  
The velar nasal effect in words with a polysyllabic  $M_1$  is larger for Kyungsang speakers.
- d. velar nasal,  $M_1$  origin, and dialect  
The velar nasal effect in words with a Sino-Korean  $M_1$  is smaller for Kyungsang speakers.
- e. length and  $M_1$  origin  
The length effect is larger with a Sino-Korean  $M_1$ .
- f. frequency, length and dialect  
For Seoul speakers, n-insertion is more likely with frequent words with a polysyllabic  $M_1$ .

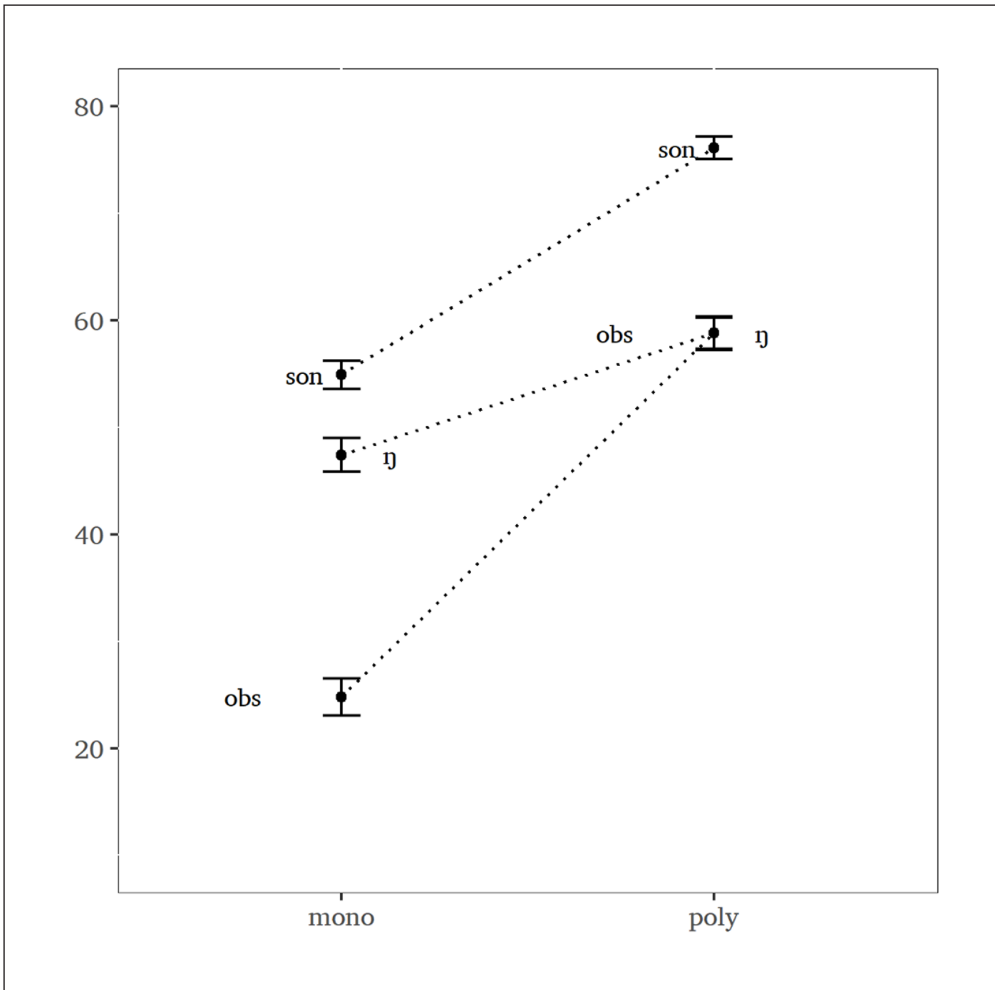
Let us discuss these significant interaction effects.

Four interaction effects in (12a–d) indicate how  $C_1$  type effects, obstruency and velar nasal, varied according to dialect,  $M_1$  length, and/or  $M_1$  origin. The larger obstruency effect for Kyungsang speakers (12a) can be seen in [Table 12](#) and [Figure 7](#). This indicates that obstruent consonants are even weaker triggers of n-insertion for Kyungsang speakers than for Seoul speakers. But, the interaction effect in (12b) suggests that this is mainly true for words with monosyllabic  $M_1$  since the obstruency effect for Kyungsang speakers is larger with monosyllabic  $M_1$  than with polysyllabic  $M_1$ , as can be seen in [Table 22](#) and [Figure 15](#).

The observed stronger obstruency effect with monosyllabic  $M_1$  in the present Kyungsang data is probabilistically consistent with Lee's (2006) description of Kyungsang Korean n-insertion, discussed in section 2.3, which says that Kyungsang Korean n-insertion occurs only after

**Table 22** Insertion rate (%) in Kyungsang Korean data by C<sub>1</sub> type and M<sub>1</sub> length.

C <sub>1</sub> type	mono	poly
obs	24.8	58.8
son	54.9	76.1
ŋ	47.4	58.8



**Figure 15** Insertion rate (%) in Kyungsang Korean data by C<sub>1</sub> type and M<sub>1</sub> length.

sonorants, not obstruents, in Sino-Korean words consisting of monosyllabic root morphemes.<sup>14</sup> As stated in (12c), the velar nasal effect also interacts with dialect and M<sub>1</sub> length, but in the opposite direction. As can also be seen in [Table 22](#) and [Figure 15](#), the velar nasal effect is larger with polysyllabic M<sub>1</sub> morphemes in Kyungsang Korean data. As stated in (12d), the velar nasal effect also interacts with M<sub>1</sub> origin. The velar nasal effect is smaller with Sino-Korean M<sub>1</sub> in Kyungsang Korean data, as can be seen [Table 23](#) and [Figure 16](#).

In addition, the interaction between M<sub>1</sub> length and M<sub>1</sub> origin (12e) indicates that the length effect was larger with a Sino-Korean M<sub>1</sub> than with a native Korean M<sub>1</sub>, as shown in [Table 24](#) and [Figure 17](#).

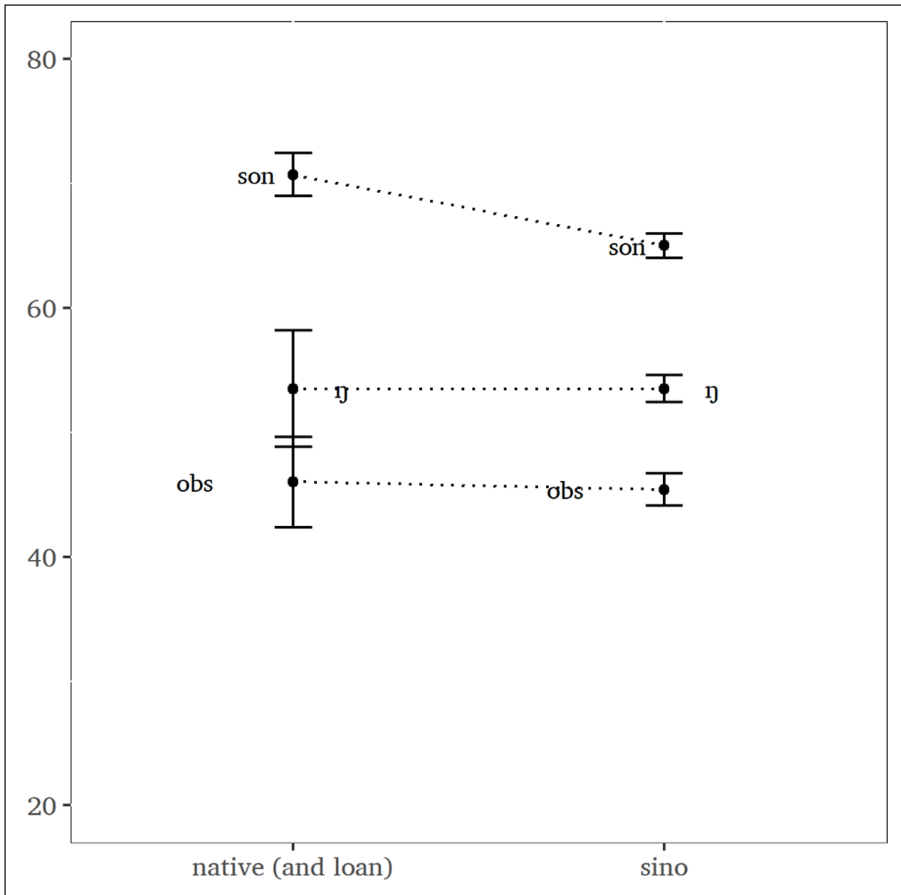
C <sub>1</sub> type	native (and loan)	sino
obs	46.0	45.4
son	70.7	65.0
ŋ	53.5	53.5

**Table 23** Insertion rate (%) in Kyungsang data by C<sub>1</sub> type and M<sub>1</sub> origin.

Note in fact that the M<sub>1</sub> length effect can be seen only among words with Sino-Korean M<sub>1</sub> which formed the majority of the test words.

<sup>14</sup> Note that the majority of root M<sub>1</sub> morphemes adopted in the current survey, 55 out of 58, are Sino-Korean.

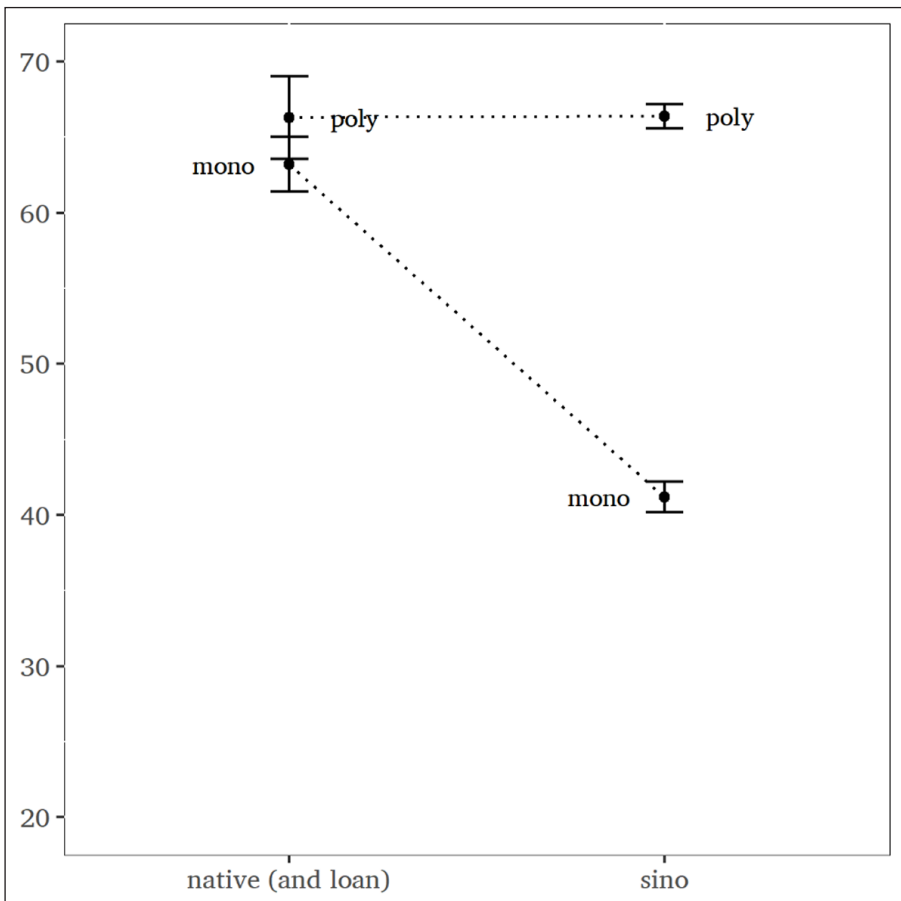




**Figure 16** Insertion rate (%) in Kyungsang Korean data by C<sub>1</sub> type and M<sub>1</sub> origin.

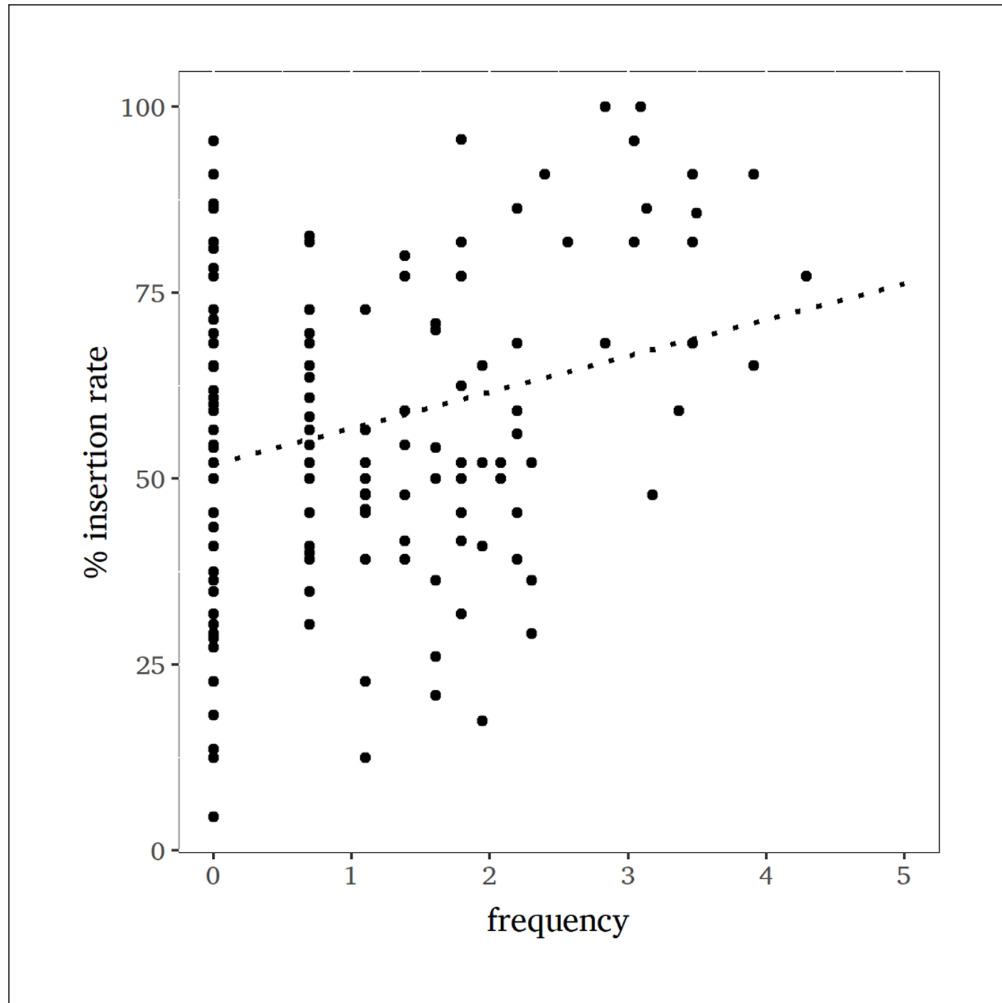
	native (39) and loan (4)	Sino-Korean (250)
monosyllabic	61.1 (30)	32.3 (100)
polysyllabic	60.2 (13)	61.6 (150)

**Table 24** Insertion rate (%) by M<sub>1</sub> origin and M<sub>1</sub> length. (The number of corresponding words is shown in parentheses.)



**Figure 17** Insertion rate (%) by M<sub>1</sub> origin and M<sub>1</sub> length (sino = Sino-Korean).

Finally, although frequency plays no role in predicting rates of Korean n-insertion in the entire data set, as shown in section 3.2.7, it is a significant predictor in a subset of the Seoul Korean data consisting of words with a polysyllabic  $M_1$  (12f). For Seoul Korean participants, words with a polysyllabic  $M_1$  tend to undergo n-insertion more frequently as their token frequencies increase, as indicated by the correlation presented in (13) and a regression line in *Figure 18*.



**Figure 18** Insertion rate (%) by (log-transformed) frequency (dialect = Seoul,  $M_1$  length = polysyllabic,  $n = 163$ ).

(13) Correlations between insertion rates and (log-transformed) frequency (dialect = Seoul Korean)

$M_1$ length	Correlations
monosyllabic (130)	$r(128) = -0.079, p = 0.373$
polysyllabic (163)	$r(161) = 0.262, p < 0.001$

### 3.5 Summary

All the significant effects, main or interaction, discussed above, are summarized in (14)

(14) Observations about n-insertion in existing words (A < B means “n-insertion is more frequent under condition B than condition A”)

Factor		significant effect
a. morphology	$M_1$	root < stem, affix
	$M_2$	root < stem, affix
b. etymology		none
c. $C_1$ type: obs vs. son		i. obs < son
		ii. larger difference (Kyungsang)
		iii. larger difference after mono $M_1$ (Kyungsang)
d. $C_1$ type: /ŋ/ vs. son		i. ŋ < son
		ii. larger difference after poly $M_1$ (Kyungsang)
		iii. smaller difference after sino $M_1$ (Kyungsang)

e. $M_1$ length	i. mono < poly ii. larger difference with sino $M_1$
f. $V_2$ height	non-high < high
g. frequency	less frequent < more frequent after poly $M_1$ (Seoul)

The probability of n-insertion in existing Korean words is significantly affected by various factors, as summarized above. Most of the factors do not completely determine the occurrence or absence of n-insertion, but they have gradient effect, contributing to the overall probability. The next section deals with whether, and which of, the gradient effects found in this section Korean speakers are aware of. We will focus on the effects of dialect,  $C_1$  type,  $M_1$  length, and  $V_2$  height, which can readily be tested with novel words. The rest of the factors tested in the survey on existing words were excluded from the novel word survey mainly due to the difficulty in creating appropriate test items. For instance, it was unclear how to construct words with novel bound roots and novel Sino-Korean, as opposed to native Korean, words.

## 4. n-insertion in novel Korean words

### 4.1 Data and experimental procedure

This section explores n-insertion in novel Korean words by investigating the results of the surveys on native speakers of two dialects of Korean, Seoul and Kyungsang. As in section 3, for novel word data of Seoul Korean speakers, I use the results of a novel word survey conducted by Jun (2015). The results are from responses of thirty-seven Seoul Korean speakers. For Kyungsang Korean data, I performed a survey on Kyungsang Korean speakers, using the same method and the same word set as adopted by Jun.

In total, thirty-two paid Northern Kyungsang Korean speakers participated in the test (mean age = 25.1 years, 17 females and 15 males). None of them participated in the survey with existing words. Most of the participants ( $n = 24$ ) were raised in Daegu, and the rest ( $n = 8$ ) in other parts of Northern Kyungsang province. All participants were recruited from the community at Seoul National University through public advertising and personal referrals. Twenty-eight participants took the survey in a quiet room (for about 25–35 minutes). The rest of the participants ( $n = 4$ ) completed and returned the survey via email.

All test words consist of loanword  $M_1$  and wug stem  $M_2$ .  $M_1$ , which is either mono or disyllabic, ends with one of seven consonants, /m, n, ŋ, l, p, s, k/.  $M_2$  begins with one of /i, ju, ja/. The total number of test items is 84 (2 syllable counts  $\times$  7 codas  $\times$  3 vocoid types  $\times$  2 repeating blocks). The same number of control items (vowel-final  $M_1$  or /a, e/-initial  $M_2$ ) were adopted. The items were pseudorandomized in order, such that test items never followed each other and were always separated by a control item. Two sets of items were prepared: one set was the reversed version of the other.

In the survey form, two parts of a word,  $M_1$  and  $M_2$ , along with its inserted and non-inserted forms, were presented in standard Korean orthography. The experimenter told the participants that the combination of the two parts is a made-up compound noun for a new chemical product. The participants were instructed to choose their pronunciation of each of the given compounds from the following three (or two) options.

(15) Options in the test form

	(i)	(ii)	(iii)
e.g.	s'ʌm + jucenol 'some' + wug	t <sup>h</sup> ap + jucenol 'top' + wug	k <sup>h</sup> iŋ + jucenol 'king' + wug
a. insertion	s'ʌm.nju.ce.nol	t <sup>h</sup> am.nju.ce.nol	k <sup>h</sup> iŋ.nju.ce.nol
b. non-insertion (resyllabified)	s'ʌ.mju.ce.nol	t <sup>h</sup> a.pju.ce.nol	—
c. non-insertion (aligned)	s'ʌm.ju.ce.nol	t <sup>h</sup> ap.ju.ce.nol	k <sup>h</sup> iŋ.ju.ce.nol

Like in the survey of existing words, when  $C_1$  is /ŋ/ as in (15iii), only two options (15a,c) were given. When  $C_1$  was underlyingly an obstruent as in (15ii), the option (15a) with insertion included nasalization of  $C_1$  due to obstruent nasalization. The participants were allowed to choose more than one, each of which was counted as one data point in the analysis. Also, when their pronunciation was not given as an option, they were asked to write it in the blank space on the test form.

In order to construct the database for analysis of n-insertion in novel Korean words, I first combined results of Jun’s (2015) survey on Seoul Korean speakers and the results of the present survey on Kyungsang speakers. Two Seoul Korean speakers inserted more frequently in control tokens than target tokens, which I think is beyond the permissible range of errors. Their responses were considered not reliable, and thus excluded from analysis. After further excluding responses to fillers with vowel-final  $M_1$ , the resulting data consist of 9,814 responses.

An initial investigation of the novel word data shows clear syllabicity effect of  $M_2$ -initial vocoids (i.e. n-insertion is less likely before /i/ than /j/), as can be seen in [Table 25](#).

j	i	a	e
31.4	6.5	6.7	8.4

**Table 25** Insertion rate (%) in novel words by  $M_2$ -initial vocoid.

Insertion rate is obviously higher before a glide /j/ than before vowels. But, /i/ is not higher in insertion rate than control vowels /a, e/, suggesting that insertion before /i/ is no more productive than insertion before other vowels.

Given that only pre-/j/ insertion is productive, the remainder of the analysis will focus on the data for test words with /j/-initial  $M_2$ , which consist of 3,957 responses. This will facilitate a direct comparison between the patterns of novel and existing Korean words. As in section 3, the novel word data will be analyzed, first using mean insertion rates and then using a mixed effects logistic regression model.

## 4.2 Raw data analysis

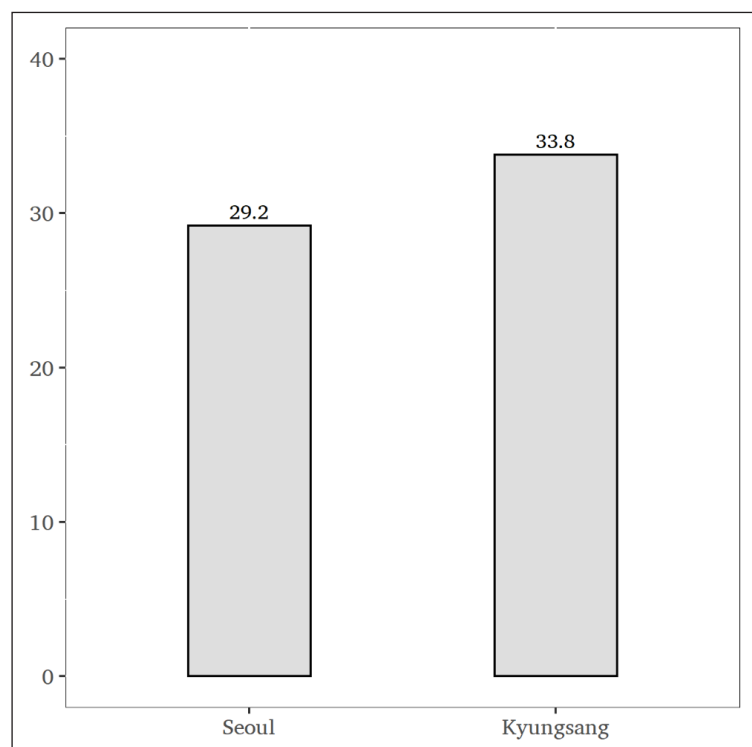
This section discusses only overall patterns of the data by calculating the mean insertion rates for each of the factors adopted in the experiment. The next section provides a statistical analysis of the same data.

### 4.2.1 Dialect

The mean rate of insertion for the entire data set is 31.4%, and the rate is higher for Kyungsang speakers than for Seoul speakers, as shown in [Table 26](#) and [Figure 19](#).

dialect	rates
Seoul	29.2
Kyungsang	33.8

**Table 26** Insertion rate (%) by dialect.



**Figure 19** Insertion rate (%) by dialect.

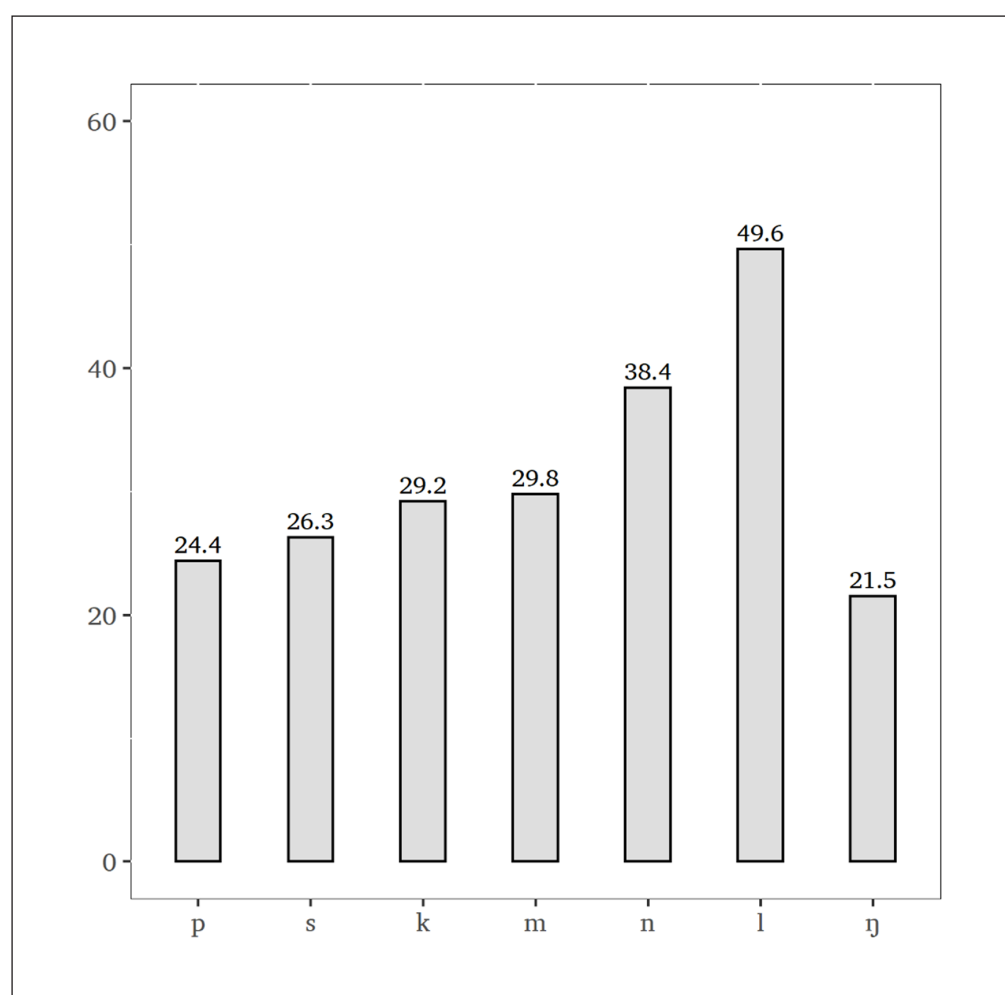
It will be shown that the observed difference is not statistically significant, as in the results of the existing word survey.

#### 4.2.2 C<sub>1</sub> type

Insertion rates were calculated for different C<sub>1</sub> consonants, as shown in [Table 27](#) and [Figure 20](#).

C <sub>1</sub> type	C <sub>1</sub>	rates
obstruent	p	24.4
	s	26.3
	k	29.2
sonorant	m	29.8
	n	38.4
	l	49.6
	ŋ	21.5

**Table 27** Insertion rate (%) by C<sub>1</sub> consonants.



**Figure 20** Insertion rate (%) by C<sub>1</sub> consonants.

Insertion rates were higher with all three sonorant consonants, /m, n, l/, than with obstruents and /ŋ/, although the difference between /m/ and /k/ is very small. To establish the effects of C<sub>1</sub> type, i.e. obstruency and velar nasal effects, which were significant in existing words, insertion rates were calculated for obstruents, /ŋ/ and other sonorant consonants, as shown in [Table 28](#).

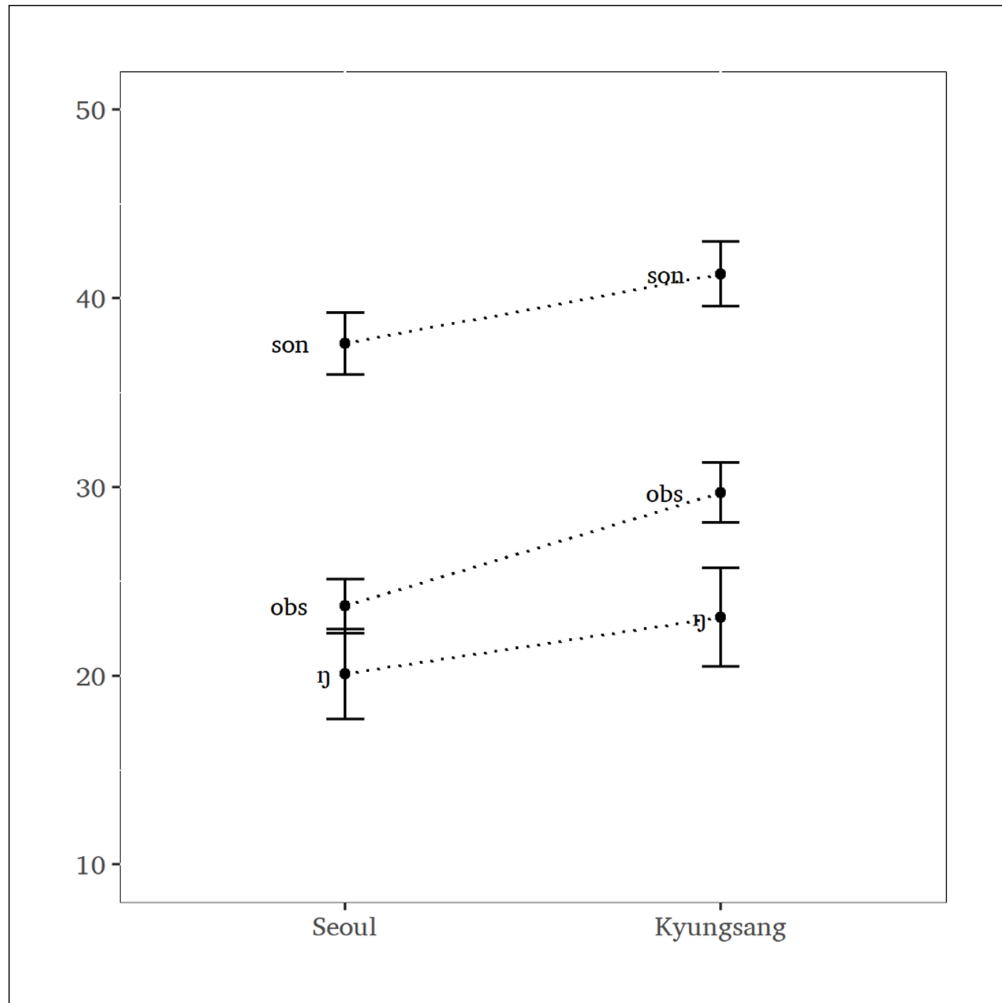
C <sub>1</sub> type	rates
obs	26.6
son	39.4
ŋ	21.5

**Table 28** Insertion rate (%) by C<sub>1</sub> type.

Note that n-insertion in novel words is less likely after obstruents and /ŋ/ than after other sonorant consonants, just like in existing words. The observed lower insertion rates with an obstruent or /ŋ/ C<sub>1</sub> are true for both Seoul and Kyungsang speakers, as can be seen in [Table 29](#) and [Figure 21](#).

C <sub>1</sub> type	Seoul	Kyungsang
obs	23.7	29.7
son	37.6	41.3
ŋ	20.1	23.1

**Table 29** Insertion rate (%) by C<sub>1</sub> type and dialect.



**Figure 21** Insertion rate (%) by C<sub>1</sub> type and dialect.

The ranking by C<sub>1</sub> type, son > obs > /ŋ/, in novel words did not differ by dialect, which is not consistent with the corresponding results of the existing word survey where insertion rate in Kyungsang Korean data was higher with /ŋ/ C<sub>1</sub> than with an obstruent C<sub>1</sub>, unlike in Seoul Korean data.

In summary, n-insertion was less likely after an obstruent or velar nasal than after other sonorant consonants, regardless of the dialect of the participants. These tendencies in novel words are consistent with the tendencies, called obstruency and velar nasal effects, found in existing words.

#### 4.2.3 Length

To establish the effect of M<sub>1</sub> length, insertion rates were calculated for different lengths of M<sub>1</sub>, as shown in [Table 30](#).

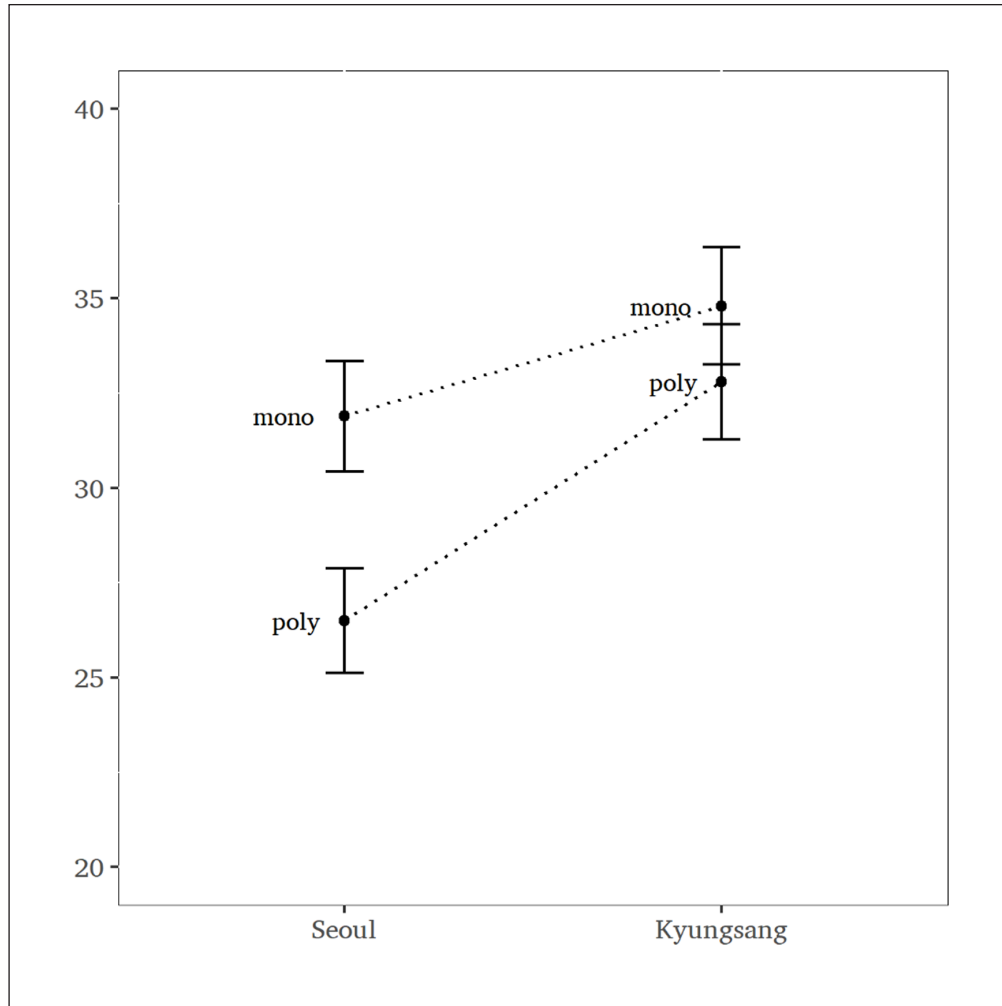
M <sub>1</sub> length	rates
monosyllabic	33.3
polysyllabic	29.5

**Table 30** Insertion rate (%) by M<sub>1</sub> length.

Insertion rate in novel words is higher with a monosyllabic  $M_1$ , which is not consistent with the corresponding results of the existing word survey. The observed higher rate of insertion with monosyllabic  $M_1$  is true for both Seoul and Kyungsang participants, as shown in [Table 31](#) and [Figure 22](#).

	Seoul	Kyungsang
polysyllabic	26.5	32.8
monosyllabic	31.9	34.8

**Table 31** Insertion rate (%) by  $M_1$  length and dialect.



**Figure 22** Insertion rate (%) by  $M_1$  length and dialect.

In summary, n-insertion was more likely with monosyllabic  $M_1$  than with polysyllabic  $M_1$ , regardless of the dialect of the participants. This is the opposite of the tendency observed in existing words, that is, higher rates with a polysyllabic  $M_1$ . Consequently, the results of the existing and novel word surveys are not consistent with each other with respect to the effect of  $M_1$  length. (See section 5.2 for a further discussion.)

#### 4.2.4 $V_2$ height

To establish the effect of the height of a vowel following /j/, called  $V_2$ , insertion rates were calculated for high and non-high  $V_2$  vowels, as shown in [Table 32](#).

$V_2$ height	rates
non-high	27.3
high	35.3

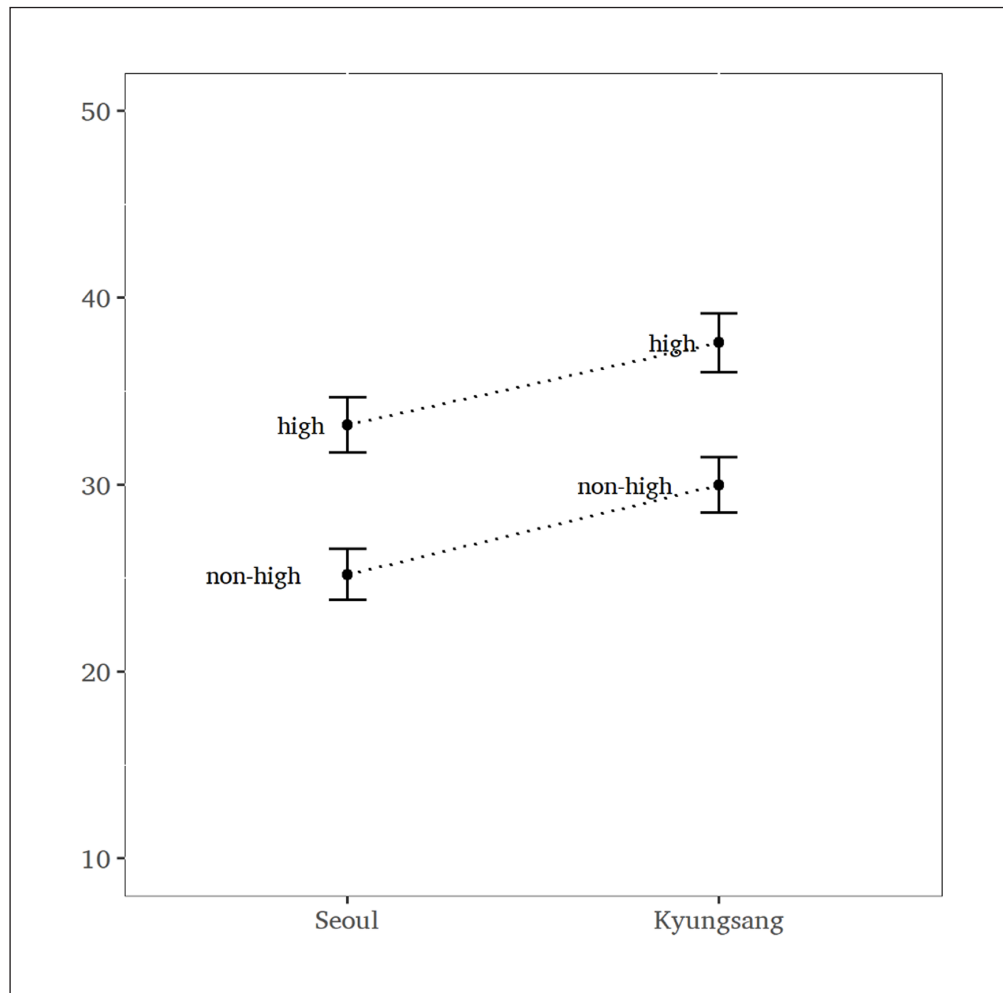
**Table 32** Insertion rate (%) by  $V_2$  height.

Insertion rate in novel words was higher before a high  $V_2$  than before a non-high  $V_2$ , which is consistent with the corresponding results of the existing word survey. This observed higher rate

of n-insertion with a high  $V_2$  is true for both Seoul and Kyungsang participants, as can be seen in [Table 33](#) and [Figure 23](#).

$V_2$ height	Seoul	Kyungsang
non-high	25.2	30.0
high	33.2	37.6

**Table 33** Insertion rate (%) by  $V_2$  height and dialect.



**Figure 23** Insertion rate (%) by  $V_2$  height and dialect.

In summary, n-insertion was more likely with a high  $V_2$  than with a non-high  $V_2$ , regardless of the dialect of the participants. This tendency in novel words is consistent with the tendency, called height effect, found in existing words.

### 4.3 Mixed effects analysis

As in the analysis of existing Korean words, a mixed effects logistic regression model was fitted with the `glmer` function from the `lmerTest` package (Kuznetsova et al. 2017) in R (R Core Team 2020). Dependent variable is binary, i.e. n-inserted or not. Each subject and each test word were included as random intercepts.

In the model on the data of novel words, the following independent variables were adopted.

- dialect (sel, ks)
- $M_1$  length (mono, poly)
- $C_1$  type (obs,  $\eta$ , son)
- $V_2$  height (high, non-high)

The interactions of dialect with all the remaining factors were included to address the question of whether the effect of each variable varies across dialects. As in the analysis of existing words, I have conducted a CART (Classification and Regression Tree) analysis of each of Seoul and Kyungsang Korean data using the `rpart` package (Therneau & Atkinson 2019), in order



to determine which interactions are active in the novel word data. Although no interactions turned out to be active in the novel word data, I have added a following interaction which was significant in existing words and can be tested in novel words: a 3-way interaction between dialect,  $M_1$  length and  $C_1$  type. As in the analysis of existing word data,  $C_1$  type with three levels was forward-difference coded, and all the remaining categorical factors were sum-coded.

As in the analysis of existing word data, the random effects of the mixed effects model show variation depending on the participant (variance = 2.814) and test word (variance = 0.124), meaning that different participants and test words show greater and lesser average rates of insertion. Even after these random factors are taken into account, as can be seen in (16), some fixed factors still have a sizable effect on the insertion rate, which holds independently of the specific participant and test word.

(16) Results of a mixed effects logistic regression model (novel words)<sup>15</sup>

	Estimate	Pr(>  z )
(Intercept)	-1.368	< 0.001***
dialect (ks)	0.102	0.634
$C_1$ type (obs-son)	-0.812	< 0.001***
$C_1$ type (son-ŋ)	1.233	< 0.001***
$V_2$ height (high)	0.269	0.001***
$M_1$ length (poly)	-0.151	0.098
dialect (ks) : $C_1$ type (obs-son)	0.085	0.34
dialect (ks) : $C_1$ type (son-ŋ)	0.041	0.765
dialect (ks) : $V_2$ height (high)	-0.008	0.849
dialect (ks) : $M_1$ length (poly)	0.044	0.378
dialect (ks) : $M_1$ .length (poly) : $C_1$ type (obs-son)	-0.017	0.93
dialect (sel) : $M_1$ .length (poly) : $C_1$ type (obs-son)	0.018	0.924
dialect (ks) : $M_1$ .length (poly) : $C_1$ type (son-ŋ)	0.173	0.536
dialect (sel) : $M_1$ .length (poly) : $C_1$ type (son-ŋ)	0.008	0.977

The next section discusses the results of the novel word survey focusing on the significant effects in the mixed effects model, just presented.

#### 4.4 Significant effects

Unlike in existing word data in which several main and interaction effects were significant, only the main effects involving  $C_1$  type and  $V_2$  height turned out to be significant ( $\alpha = 0.05$ ) in novel word data, as summarized in (17)

- (17) Summary: significant main effects (novel words)
- a. Obstruency effect  
n-insertion is less likely after an obstruent  $C_1$ .
  - b. Velar nasal effect  
n-insertion is less likely after a velar nasal  $C_1$ .
  - c. Height effect  
n-insertion is more likely with a high vowel following /j/.

All of these main effects can also be seen in the corresponding tables and plots presented in section 4.2. Let us discuss these significant effects.

Insertion rates differed depending on the type of  $C_1$ . Two observed relevant effects, obstruency (17a) and velar nasal (17b), suggest that sonorant consonants other than /ŋ/ are frequent triggers of n-insertion in novel Korean words. Recall in section 3 that the two observed effects involving  $C_1$  type, obstruency and velar nasal, were also found significant in the existing word data. This suggests that both obstruency and velar nasal effects were extended to novel words.

<sup>15</sup> As in the model on existing words, shown in (8), the non-reference level of sum-coded binary factors and the two levels of forward-difference coded ternary factors relevant to the calculation of the corresponding estimate are shown in parentheses.

In addition, the height effect (17c) suggests that a high  $V_2$  vowel is a more likely trigger of n-insertion in novel words than non-high  $V_2$  vowels. Recall in section 3 that this  $V_2$  height effect was also significant among existing words, indicating that the  $V_2$  height effect was extended to novel words.

To summarize the results of the survey with novel Korean words, three main effects, obstruency, velar nasal and  $V_2$  height, were confirmed. None of the remaining effects, main or interaction, were significant. As will be discussed in section 5.2, these results suggest that Korean speakers can learn prominent patterns among existing words through phonological generalizations.

## 5. Discussion

This section discusses how to analyze the occurrence of n-insertion and the observed tendencies, the data vs. learning (mis)match, dialectal variation and remaining problems.

### 5.1 Motivation, constraint and P-map effects

To illustrate how to analyze the occurrence of n-insertion, consider inserted and non-inserted forms shown in (18).

- (18) Forms with and without n-insertion (A close bracket (]) and a dot (.) indicate the right edge of  $M_1$  and syllable boundary, respectively.)

	non-inserted		c. inserted
	a. resyllabification	b. alignment	
/som-ɪpul/	so.m]i.pul	som.]i.pul	som.]ni.pul
/com-jak/	co.m]jak	com.]jak	com.]njak

If /n/ is not inserted,  $C_1$  consonants at the end of  $M_1$  would be either resyllabified into the onset of  $M_2$ -initial syllable as in (18a) or be aligned with the end of the  $M_1$ -final syllable as in (18b). The non-inserted forms with resyllabification in (18a) involves a misalignment between morpheme and syllable boundaries. Many previous analyses of Korean n-insertion have argued, or assumed, that n-insertion is motivated by the requirement to align morpheme-syllable boundaries (Jun 2015 and references therein). In the non-inserted forms with alignment in (18b), vocoids /i, j/ initiate a syllable, leaving the preceding  $C_1$  consonants at the end of  $M_1$ -final syllable. This syllable structure is marked in Korean where an intervocalic consonant (with the possible exception of /ŋ/) is syllabified as an onset. In terms of Optimality Theory (Prince & Smolensky 1993/2004), the forms with resyllabification and alignment violate the alignment constraint (“The right edge of a morpheme coincides with the right edge of a syllable.”) and syllable structure constraints, respectively. The relevant syllable structure constraints here include ONSET (“Syllables have an onset.”) and VOCOID-NUCLEUS (“Every [–consonantal] segment must be in the nucleus.”). The former penalizes aligned forms with a vowel-initial syllable (e.g. [som.i.pul]) whereas the latter penalizes those with a glide-initial syllable (e.g. [com.jak]). In contrast, the forms with n-insertion in (18c) satisfy both the alignment and syllable structure constraints at the cost of violating the constraint militating against insertion of a segment, namely DEP. The epenthetic /n/ initiates  $M_2$ -initial syllables, not only aligning between morpheme and syllable boundaries, but also avoiding syllables with no (or a bad, i.e. [–consonantal]) onset. To summarize, a consonant needs to be inserted at the beginning of  $M_2$  to obey the alignment and syllable well-formedness constraints.

Let us now consider the questions of why /n/, not any other consonant, is inserted, and why insertion takes place only before high front vocoids. Jun (2015) answers these questions by arguing that /n/ in Korean is perceptually weak in the context of n-insertion, i.e. before high front vocoids /i, j/. Specifically, Jun’s analysis of Korean n-insertion is couched within the framework of 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. In Korean, if /n/ occurs before high front vocoids /i, j/, it undergoes allophonic palatalization, failing to induce coarticulatory changes on the following vocoids. Thus, the input–output pairs i–ni (or, more precisely, i–ɲi) and j–ɲj

(or  $j-nj$ ) are perceptually more similar pairs, compared to other pairs such as  $i-mi$  and  $a-na$ . This suggests not only that insertion of  $/n/$  before high front vocoids is perceptually less salient than insertion of other consonants in the same environment, but also that insertion of  $/n/$  is less salient before high front vocoids than before other vocoids. Consequently,  $/n/$  is inserted before high front vocoids because it is the segment most confusable with zero there.

A similar P-map-based account can be given for some of the observed effects, syllabicity and  $V_2$  height. In Korean, unlike  $/ni/$ , the sequence  $/nj/$  is phonetically realized almost as a single segment  $[n]$ , as stated in footnote 2. Thus the input–output pair  $j-nj$  is perceptually almost identical to  $j-n$ , which must be more similar than  $i-ni$ , since the difference is less than a single segment in the former, but the entire segment  $[n]$  in the latter. Also, the degree of perceptual modification involved in pre- $/j/$  n-insertion 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.  $[n]$ . But when it is followed by a non-high vowel,  $/j/$ 's low-frequency resonance is not maintained for as long, making it less nasal-like, so that the relevant pair is less similar. The height effect can be attributed to this difference in perceptual similarity between the input–output pairs with high and non-high vowels. To summarize, Korean n-insertion is more likely before a glide  $/j/$  (cf.  $/i/$ ) and a high  $V_2$  (cf. non-high  $V_2$ ) since  $/n/$  is more confusable with zero in those environments.

Let us now consider the rest of the observed tendencies. Hwang (2008) and Jun (2015) have attributed the obstruency effect to automatic obstruent nasalization in Korean. If n-insertion applied to words with obstruent  $C_1$ ,  $C_1$  would become a nasal, violating ID(sonorant) (“Do not change the value of [sonorant].”). Since inserted forms of words with  $M_1$ -final sonorants would not violate ID(sonorant), other things being equal, n-insertion would be less likely after an obstruent than after a sonorant.

Jun (2015) has explained the velar nasal effect by arguing that an intervocalic  $/ŋ/$  in Korean can be ambisyllabic. Given that non-inserted forms with an ambisyllabic  $C_1$  may satisfy the alignment constraint, the demand for n-insertion would be less strong for words with  $/ŋ/$   $C_1$ , which is consistent with the velar nasal effect.<sup>16</sup>

The  $M_{1,2}$  morphology effects (“n-insertion is more likely with stems or affixes than with roots.”) and lexical effects found in existing Korean words may be explained by indexing some of the constraints, discussed above, to morphological categories and lexical items (Archangeli & Pulleyblank 2002, Pater 2007). For instance, the prefix-specific version of alignment constraint would encourage n-insertion to occur after a prefix.

Thus far, it has been shown that not only the occurrence of n-insertion but also most of the observed effects (except the length effect) can be explained based on universal phonological constraints and assumptions specific to Korean phonology. Finally, note that a complete formal analysis of the current n-insertion data with much variation need to and can be couched within a probabilistic constraint grammar, as suggested by Jun’s (2015) analysis of Seoul Korean n-insertion based on the maximum entropy harmonic grammar (Hayes & Wilson 2008).

## 5.2 Does learning match data?

This section discusses whether the tendencies observed in existing Korean words were mirrored in novel words, in order to establish whether learning matches data. Specifically, it compares the patterns of existing words, presented in section 3, with those of novel words with  $/j/$ -initial  $M_2$ , presented in section 4.

The overall insertion rate of novel words, 31.4%, is much lower than that of the existing words, 51.4%. It is then expected that the size of an effect in existing words would be reduced in novel words, and only strong tendencies in the lexicon can be extended to novel words with statistical significance. In fact, three main effects in novel word data were significant ( $\alpha = 0.05$ ), as summarized in (17), and no other main and interaction effects were significant.

---

<sup>16</sup> Note that the non-inserted forms with an ambisyllabic  $/ŋ/$   $C_1$  would violate (low-ranked/weighted) constraints including \*CODA (“Syllables do not have codas.”) and the constraint prohibiting an ambisyllabic  $/ŋ/$ .

Recall in section 3 that all the three significant effects in novel word data were also significant in existing word data. It suggests that Korean speakers are aware of each of the corresponding tendencies in the lexicon, and they used this knowledge in the novel word survey.

What about the rest of the tendencies tested in the novel word survey? All of them, which turned out to be insignificant in novel words, involve  $M_1$  length and dialect. In the remainder of this subsection, I will consider the  $M_1$  length effect, postponing the discussion of dialect effects to the next section.

As shown in section 4.2.3, insertion rate in novel words was higher with monosyllabic  $M_1$ . The relevant difference was not significant in the mixed effect model shown in section 4.3 ( $M_1$  length (poly):  $b = -0.151$ ,  $p = 0.098$ ). This statistically insignificant tendency in novel words is the opposite of the tendency in existing words. Thus, this cannot be a case where the size of a lexical tendency is simply reduced in novel words. This suggests that Korean speakers failed to learn the tendency involving the  $M_1$  length of existing words. This mismatch has been reported by Jun (2015), from which the Seoul Korean data of the present study are drawn. The insignificant interaction between  $M_1$  length and dialect shown in (16), along with the insertion rates calculated in novel words for different lengths of  $M_1$ , shown in [Table 31](#) and [Figure 22](#), suggest that the failure of extending the length effect to novel words was not confined to Seoul Korean speakers. Jun attributes the observed mismatch between data and learning to the lack of a phonetic/phonological motivation for the tendency related to the length of  $M_1$ . Recall in the previous section that not only the occurrence of n-insertion but also most of the observed effects, except the length effect, can be explained in a phonetically and/or phonologically plausible way. It has been argued in the literature (Hayes et al. 2009, Becker et al. 2011, Hayes & White 2013) that lexical patterns lacking phonetic/phonological naturalness either cannot be learned or can be learned with much difficulty. Thus, the lack of phonetic/phonological naturalness seems to be a potentially plausible account of why the length effect in the lexicon failed to extend to novel words. To explain the failure of extending the length effect to novel words, one might consider a possibility that Korean speakers applied n-insertion in the novel word survey using their knowledge of patterns in existing native, not Sino-Korean, words. Recall, as shown in [Table 24](#) and [Figure 17](#) that the  $M_1$  length effect was effective mainly in existing words with Sino-Korean, not native,  $M_1$ . However, it is unclear not only why Korean speakers ignored distributional patterns in Sino-Korean words which form the majority of the Korean lexicon, but how they successfully distinguished native Korean morphemes from Sino-Korean ones.

To summarize, three tendencies, i.e. obstruency, velar nasal and  $V_2$  height effects, observed from the novel words mirror those from existing words, indicating a match between learning and data. This suggests that speakers can generalize gradient statistical patterns in the lexicon, as has been confirmed in many previous studies (Jun & Albright 2017 and references therein). An obvious mismatch involves the length effect, which may possibly lack phonetic/phonological naturalness.

### 5.3 Dialectal variation and remaining problems

One of the aims of the present study is to explore dialectal variation in Korean n-insertion. In this section, I will first discuss the interaction effects of dialect with other factors, and then turn to the main effect of dialect.

In the analysis of the results of the survey on existing words, five out of six significant interaction effects involve the dialect of the participants as summarized in (12). All the significant interaction effects of dialect and other factors in existing word data shown in (12) were either not tested or insignificant in the novel word data. Frequency and  $M_1$  origin factors including their interactions with dialect were excluded from the novel word survey mainly due to the difficulty in creating appropriate test items. The remaining three interaction effects (12a–c) were tested in the novel word survey, but none of them turned out to be significant in the mixed effects logistic regression model presented in (16). It seems that the lack of statistical significance of the relevant dialectal differences in novel words might be due to the general reduction of average insertion rate in novel words, mentioned at the beginning of section 5.2.

Let us now compare existing and novel words with respect to the main effect of dialect. Recall that the average insertion rate was higher for Kyungsang speakers than Seoul speakers,

regardless of whether test items were existing or novel Korean words. This might lead one to think that n-insertion is more likely to occur for Kyungsang speakers than Seoul speakers, and that the relevant lexical patterns were extended to novel words. However, given that the main effect of dialect was not significant in both statistical models for existing and novel word data, as presented in (8) and (16) respectively, neither Kyungsang speakers' higher rate of insertion nor the consistency between existing and novel words can be ensured.

Note that the method of the present study has some limitations in detecting the dialectal difference even if the insertion rate is truly different between Kyungsang and Seoul Korean. First, a majority of Kyungsang participants employed in this study were relatively young, in their 20s (mean age = 25.9 years in existing word survey, 25.1 years in novel word survey). It is well-known in the literature on Korean linguistics that younger Kyungsang Korean speakers are more familiar with Seoul Korean than older Kyungsang Korean speakers. It is thus possible that the Kyungsang Korean participants in the current survey might, at least sometimes, have relied on their knowledge of Seoul Korean during the survey. Second, as pointed out by an anonymous reviewer, the modality of the task, i.e. text, not audio/production, might reduce the likelihood of finding dialectal differences. Given that non-standard dialects are usually used in spoken, as opposed to written, forms, it is possible that the modality of the present task prevented the Kyungsang participants from relying on their Kyungsang Korean grammar. Finally, as mentioned in section 3, all 304 test words used in the existing word survey were standard, thus Seoul, Korean words. The same set of test words was used for Kyungsang and Seoul Korean speakers for the purpose of making a direct comparison between the dialects. It might be then possible that Kyungsang Korean participants used their knowledge of Seoul Korean during the survey, at least when the test words were not in their native word lexicon. A more reliable detection of the dialectal difference under consideration can be made by employing older speakers of Kyungsang Korean as participants and spoken Kyungsang Korean words as test stimuli. This is left for future research.

To summarize, the present study found several statistically significant dialectal differences in the patterns of n-insertion in existing Korean words. However, it is still not known what other dialectal differences are active in existing words, and whether the dialectal differences in the lexicon can generally be extended to novel words.

## 6. Conclusion

In this study, I have explored variation in Korean n-insertion for the purpose of finding out whether phonological and morphological factors may have gradient effects in application of a morphophonological process, how they interact, and which of the gradient effects speakers are aware of. From the results of the survey on n-insertion in existing Korean words, I have found several gradient tendencies involving a variety of factors including phonological and morphological ones and their interactions. Such factors include morphological category, etymology and length of component morphemes, sonorancy and place of articulation of the consonant preceding the insertion site, height of the initial vowel of the morpheme following the insertion site and dialects of speakers. None of these factors are absolute conditions for the occurrence of n-insertion, contrary to the previous studies on Korean n-insertion, and they have gradient effect, contributing to the overall probability. Thus, Korean n-insertion provides a clear case where previous categorical proposals do not match with gradient reality, lending support to quantitative theories of morphophonology.

Some, not all, of the observed tendencies were tested in the survey with novel Korean words. The tendencies involving the morphological category and etymology of component morphemes were excluded from the novel word investigation, mainly due to the difficulty to construct appropriate test items. Three phonological tendencies, obstruency, velar nasal and height effects, were mirrored in the results of a survey involving novel words, suggesting that Korean speakers are aware of the differential influence of such phonological factors on the probability of the application of n-insertion. No other effects, including the length and dialect effects, turned out to be significant in the novel word investigation. The statistical insignificance of some of these effects might be due to the general reduction of the average insertion rate in novel words. However, it seems clear from the relevant result of statistical analysis and insertion rates that

the length effect was not extended to novel words, and thus Korean speakers failed to learn the length effect. This failure has been attributed to the lack of phonological naturalness involved.

## Abbreviations

I use the Leipzig glossing rules with the following additions:

ks = Kyungsang, loan = loanword, mono = monosyllabic, obs = obstruent, poly = polysyllabic, SE = Sentence Ender, sel = Seoul, sino = Sino-Korean, SKD = Standard Korean dictionary, son = sonorant except /ŋ/

## Acknowledgements

This paper has greatly benefited from the advice and comments from many people. I am grateful to Adam Albright, Hanyoung Byun, Edward Flemming, Bruce Hayes, Sarang Jeong, Jinyoung Jo, Hyoju Kim, Yeong-Joon Kim, Seon Park, Donca Steriade, participants of my graduate Morphology class at Seoul National University, audiences at the 27th Manchester Phonology Meeting, University of Manchester (May 2019) and at the 6th Annual meeting on phonology, UC San Diego (October 2018). I also thank Sarang Jeong and Nayoung Park for their valuable comments and assistance in survey data collection, the three anonymous reviewers for their suggestions and detailed comments, and my experimental subjects for their time. This work was supported by the Seoul National University Research Grant in 2018.

## Competing interests

The author has no competing interests to declare.

## Author affiliation

Jongho Jun  [orcid.org/0000-0002-0463-5551](https://orcid.org/0000-0002-0463-5551)  
Seoul National University, KR

## References

- Ahn, Mee-Jin. 2008. The Korean /n/-insertion before [j] as [nasal] insertion. *Studies in Phonetics, Phonology and Morphology* 14(3). 371–388. DOI: <https://doi.org/10.17959/sppm.2008.14.3.371>
- Ahn, Mee-Jin. 2009. /n/-insertion and onset simplification in Kyungsang Korean. *Studies in Phonetics, Phonology and Morphology* 15(2). 263–282. DOI: <https://doi.org/10.17959/sppm.2009.15.2.263>
- Archangeli, Diana & Douglas Pulleyblank. 2002. Kinande vowel harmony: Domains, grounded conditions and one-sided alignment. *Phonology* 19(2). 139–188. DOI: <https://doi.org/10.1017/S095267570200430X>
- Bae, Ju-Chae. 2003. *Hankwukeuy palum* [The pronunciation of Korean]. Seoul: Samkyong Munhwasa.
- Becker, Michael, Nihan Ketrez & Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87(1). 84–125. DOI: <https://doi.org/10.1353/lan.2011.0016>
- Cho, Young-mee Yu. 1995. In defense of juncture rules/constraints. *Language Research* 31(4). 589–614.
- Cho, Young-mee Yu. 2016. Korean phonetics and phonology. In Mark Aronoff (ed.), *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. Published online.
- Cho, Young-mee Yu & Gregory K. Iverson. 1997. Korean phonology in the late twentieth century. *Language Research* 33(4). 687–735.
- Choi, Hyewon. 2002. *Phyocunpalum silthaycosa* [A survey of standard pronunciation]. Seoul: The National Institute of the Korean Language.
- Coetzee, Andries & Joe Pater. 2011. The place of variation in phonological theory. In J. Goldsmith, J. Riggle & A. Yu (eds.), *The Handbook of Phonological Theory*, 401–434. 2nd edition. Malden, MA and Oxford, UK: Blackwell. DOI: <https://doi.org/10.1002/9781444343069.ch13>
- Han, Eunjoo. 1993. The phonological word in Korean. In Patricia M. Clancy (ed.), *Japanese/Korean Linguistics* 2. 117–129. CSLI, Stanford University.
- Han, Eunjoo. 1994. *Prosodic structure in compounds*. PhD dissertation, Stanford University.
- Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39. 379–440. DOI: <https://doi.org/10.1162/ling.2008.39.3.379>

- Hayes, Bruce & James White. 2013. Phonological naturalness and phonotactic learning. *Linguistic Inquiry* 44. 45–75. DOI: [https://doi.org/10.1162/LING\\_q\\_00119](https://doi.org/10.1162/LING_q_00119)
- Hayes, Bruce, Kie Zuraw, Peter Siptar & Zsuzsa Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85. 822–863. DOI: <https://doi.org/10.1353/lan.0.0169>
- Hong, Soonhyun. 2006. /n/-insertion in native Korean and Sino-Korean revisited. *Studies in Phonetics, Phonology and Morphology* 12(2). 391–413.
- Huh, Woong. 1984. *Kwukeyumwunhak* [Korean phonology]. Seoul: Cengumsa.
- Hwang, Sangjin. 2008. *Korean speakers' knowledge of /n/-insertion: P-map approach*. MA thesis, Seoul National University.
- Jun, Jongho. 2015. Korean n-insertion: A mismatch between data and learning. *Phonology* 32(3). 417–458. DOI: <https://doi.org/10.1017/S0952675715000275>
- Jun, Jongho. 2018. Morpho-phonological processes in Korean. In Mark Aronoff (ed.), *Oxford Research Encyclopedia of Linguistics*. New York: Oxford University Press. Published online. DOI: <https://doi.org/10.1093/acrefore/9780199384655.013.241>
- Jun, Jongho & Adam Albright. 2017. Speakers' knowledge of alternation is asymmetrical: Evidence from Seoul Korean verb paradigms. In *Journal of Linguistics* 53(3). 567–611. DOI: <https://doi.org/10.1017/S0022226716000293>
- Jurjec, Peter. 2016. Velar palatalization in Slovenian: Local and long-distance interactions in a derived environment effect. *Glossa* 1(1). 24. DOI: <https://doi.org/10.5334/gjgl.129>
- Kang, Beom-mo & Kim Hung-gyu. 2004. *Hankwuke hyengtayso mich ehwi sayong pintouy pwunsek 2* [Frequency analysis of Korean morpheme and word usage 2]. Seoul: Institute of Korean culture, Korea University.
- Kang, Ongmi. 2003. *Hankwuke umwunlon* [Korean Phonology]. Seoul: Taehaksa.
- Kim, Seoncheol. 2003. *Phyocwunpalum silthaycosa II* [A survey of standard pronunciation II]. Seoul: The National Institute of the Korean Language.
- Kim, Soo-Jung. 2000. *Accentual effects on phonological rules in Korean*. PhD dissertation, University of North Carolina. Chapel Hill.
- Kim, Yu-Pum, Sunwoo Park, Byoung-Sup Ahn & Bongwon Lee. 2002. Niun sapiphyensanguy yenkwusacek kemtho [An overview on the studies of Korean /n/-insertion]. *Emwunnoncip* 46. 43–71.
- Kim-Renaud, Young-Key. 1974/1991. *Korean consonantal phonology*. PhD dissertation, University of Hawaii, Honolulu. Published 1991, Seoul: Hanshin.
- Ko, Kwang-mo. 1992. Niun chemkawa saioseytayhan yenkwu [The n-epenthesis and sai-sios in Korean]. *Eoneohak* 14. 31–51.
- Kook, Kyung-A, Ju-Won Kim & Ho-Young Lee. 2005. Senhoto cosaluthonghan niun chemkahyensanguy silhyenyangsang yenkwu [A study of n-insertion preferences in Korean]. *Malsori* 53. 37–60.
- Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 8213. 1–26.
- Kwuklip kwuke yenkwuwen [National Institute of the Korean Language]. ed. 1999. *Phyocwun kwuke taysacen* [Standard Korean Dictionary]. Seoul: Doosan Dong-A.
- Lee, Hoyoung. 1996. *Kukeumsenghak* [Korean phonetics]. Seoul: Thayhaksa.
- 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(3). 623–641.
- Lee, Yongsung & Minkyung Lee. 2006. n-insertion as y-devocalization in Korean. *Korean Journal of Linguistics* 31(3). 413–440.
- McPherson, Laura & Bruce Hayes. 2016. Relating application frequency to morphological structure: the case of Tommo So vowel harmony. *Phonology* 33. 125–167. DOI: <https://doi.org/10.1017/S0952675716000051>
- Moore-Cantwell, Claire & Joe Pater. 2016. Gradient exceptionality in Maximum Entropy grammar with lexically specific constraints. *Catalan Journal of Linguistics* 15. 53–66. DOI: <https://doi.org/10.5565/rev/catjl.183>
- Oh, Mira. 2006. Niun-sapip hwankyenguy caykemtho [Reexamination of environments for /n/-insertion]. *The Linguistic Association of Korea Journal* 14(3). 117–135.
- Pater, Joe. 2007. The Locus of Exceptionality: Morpheme-Specific Phonology as Constraint Indexation. *Linguistics Department Faculty Publication Series* 172.
- Park, Jong-Hee. 2001. A status of onglide 'j' in Korean. *Korean Journal of Linguistics* 26(4). 715–733.
- Prince, Alan & Paul Smolensky. 1993/2004. *Optimality Theory: constraint interaction in generative grammar*. Oxford: Blackwell. DOI: <https://doi.org/10.1002/9780470759400>
- R Core Team. 2020. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume & Keith Johnson (eds.), *The role of speech perception in phonology*, 219–50. New York: Academic Press.

- Steriade, Donca. 2009. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In Sharon Inkelas & Kristin Hanson (eds.), *On the nature of the word*, 151–179. MIT Press.
- Therneau, Terry & Beth Atkinson. 2019. rpart: Recursive Partitioning and Regression Trees. R package version 4.1–15.
- Zuraw, Kie. 2000. *Patterned exceptions in phonology*. Ph.D. dissertation, UCLA.
- Zuraw, Kie. 2002. Aggressive reduplication. *Phonology* 19(3). 395–539. DOI: <https://doi.org/10.1017/S095267570300441X>
- Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28(2). 417–472. DOI: <https://doi.org/10.1007/s11049-010-9095-z>
- Zuraw, Kie & Bruce Hayes. 2017. Intersecting constraint families: an argument for Harmonic Grammar. *Language* 93. 497–548. DOI: <https://doi.org/10.1353/lan.2017.0035>
- Zymet, Jesse. 2018. *Lexical propensities in phonology: corpus and experimental evidence, grammar, and learning*. Doctoral dissertation, UCLA.

TO CITE THIS ARTICLE:

Jun, Jongho. 2021. Morphophonological gradience in Korean n-insertion. *Glossa: a journal of general linguistics* 6(1): 40. 1–40. DOI: <https://doi.org/10.5334/gjgl.1401>

Submitted: 29 July 2020

Accepted: 06 March 2021

Published: 07 April 2021

COPYRIGHT:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

*Glossa: a journal of general linguistics* is a peer-reviewed open access journal published by Ubiquity Press.