Japanese-Korean 24
satellite workshop

October 13, 2016

jongho@snu.ac.kr

Syllable count effects in Korean n-insertion

Jongho Jun
(Seoul National University)

Introduction

- I’m interested in the question of whether speakers learn all the tendencies which can be observed in existing words of their native language.
- To answer this question, I investigated variable n-insertion in Korean.

Overview

- I first conducted a survey on speakers of two dialects of Korean, Seoul and Kyungsang with existing Korean words.
- I found several tendencies including **syllable count** effects.
- I then conducted a wug test on Korean speakers of the same dialects.

Overview

- Results show that most of the tendencies observed from the existing words are mirrored in Korean speakers' responses in novel word experiments, suggesting that they are aware of these tendencies.
- However, syllable count effects were not extended in novel Korean words, indicating that Korean speakers failed to learn these effects.
- I argue that syllable count in Korean n-insertion is not phonologically natural unlike word length which have been considered a natural factor in some previous studies, notably Becker et al.'s (2011).

Overview

- This talk is an extension of my previous work on Seoul Korean n-insertion.

  Jun, Jongho (2015) Korean n-insertion: a mismatch between data and learning. In *Phonology* 32: 417-58.
- Today, in addition to Seoul Korean data, I discuss comparable data from Kyungsang Korean.
- I will analyze the data from the two dialects of Korean, using an expanded set of factors, some of which were not considered in my previous work.

Roadmap

I. **Basic patterns of Korean n-insertion**

II. Existing words

III. Novel words

IV. Syllable count effect

Basic patterns of Korean n-insertion

- $\varnothing \rightarrow n / \mathbf{C_1]_{M1}\_\__{M2}[i/j}$

  ($M_{1,2}$ = morpheme; C_1 = M_1-final consonant)

- /n/ is *optionally* inserted at the juncture of two morphemes when M_1ends with a consonant, C_1, and M_2 begins with a high front vocoid /i j/.
- M_1 and M_2 may form an affixed word, compound or syntactic phrase.

Examples

prefix-stem: /təs-jaŋmal/	[tən**n**jaŋmal] ‘anklet socks’
compound: /com-jak/	[com**n**jak] ‘mothball’
phrase: /mək-ɨn # jəs/	[məkɨn**n**jət] ‘taffy that (someone) ate’
English phrase	[k^{h}æn**n**ju …] ‘Can you …?’

Focus

- Compounds
- Pre-/j/ insertion
  - Pre-/i/ n-insertion is unproductive (Hwang 2008; Jun 2015).

Automatic phonological processes and n-insertion

- When C_1 is an obstruent:
  - Obstruent nasalization: an obstruent becomes a nasal before a nasal.

  /kiə**p**-jesan/ [ki.ə**m**.**n**je.san] ‘corporation budget’
- When the C_1 is a liquid:

  /al-jak/ → [al.**n**jak] → [a**l**.**l**jak] ‘tablet’

n-insertion

lateralization

Phonetic realization of inserted /n/

- The epenthetic consonant /n/ is phonetically realized as a palatalized coronal sonorant, [ɲ] or, after a lateral, [ʎ], due to **allophonic palatalization**.

/com-jak/	[comɲ(j)ak]	‘mothball’
/al-jak/	[aʎ.ʎ(j)ak]	‘tablet’

- This palatalization is not reflected in the transcriptions in the rest of this talk.

Variation

- It has been argued in the literature that n-insertion occurs only when M_2 is a stem or root which can be an **independent** word.
- This is not true.
- n-insertion with M_2 suffixes

/pisaŋ-joŋ/ [pisaŋ**n**joŋ] ‘for emergency’ (/-joŋ/ ‘for use’)

/kɬləm-jo/ [kɬləm**n**jo] ‘Of course’ (/-jo/ sentence ender)

➢ Some **dependent** M_2 elements may trigger n-insertion.

Variation

- n-insertion is optional.

  /com-jak/ [com**n**jak] ~ [comjak] ‘mothball’

- The probability of n-insertion may vary across speakers and words.

➢ The present study examines what factors determine the rate of this optional n-insertion.

Factors investigated

$$\varnothing \rightarrow n \,/\, \mathbf{C_1]_{M1}} \_\_ \mathbf{{}_{M2}[j\ V_2}$$

- C_1 type: sonorant (other than /ŋ/), obstruent, ŋ
- V_2 height: non-high, high
- M_2 morpheme type: independent, dependent
- M_1, M_2 origin: native-Korean, Sino-Korean, loanword
- M_1, M_2 length: monosyllabic, polysyllabic
- Word frequency

Roadmap

I. Basic patterns of Korean n-insertion

II. Existing words

III. Novel words

IV. Syllable count effect

Survey on existing words

- Test words: 303 polymorphemic Korean words
  - ✓ Words with orthographic sequences of a syllable-final consonant followed by /j/.
  - ✓ Words with frequency of at least 1 in the Sejong corpus (http://www.sejong.or.kr/) and, at the same time, listed as standard Korean words in the Standard Korean dictionary (Kwuklip kwuke yenkwuwen 1999).
- Participants
  - ✓ 22 Seoul Korean speakers
  - ✓ 23 Northern Kyungsang Korean speakers

Survey on existing words

- Survey form
  - ✓ Both inserted and non-inserted forms of each test word were presented in standard Korean orthography. (N.B. Basic syllable divisions can be seen in the written words.)
  - ✓ The participants were asked to choose what they think is their pronunciation, among three (or two) choices.

Existing words: Response choices

	e.g. com+jak	tʰaŋ+jak
(i) **inserted**	com.**n**jak	tʰaŋ.**n**jak
(ii) non-inserted (**resyllabified**)	co.mjak	--
(iii) non-inserted (**aligned**)	com.jak	tʰaŋ.jak

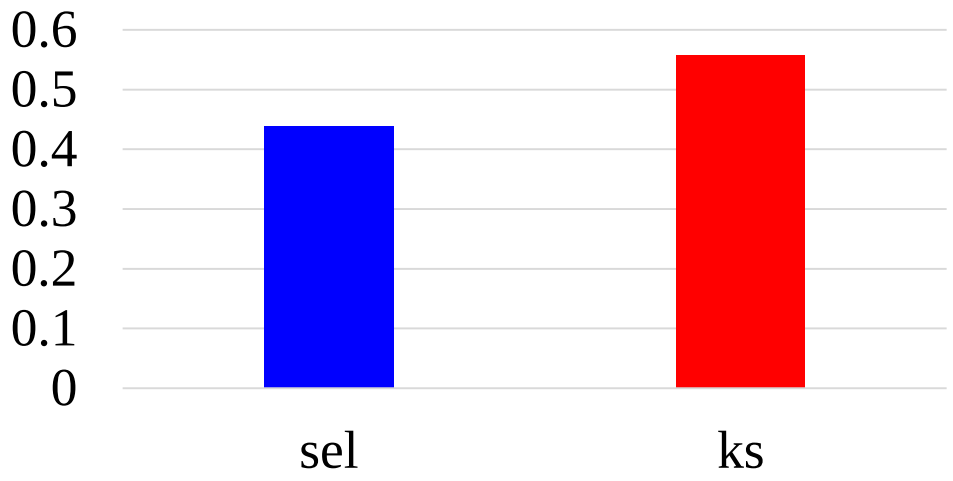
- When C_1 is the velar nasal, only two options (i) and (iii) were given since there is no way to represent the second option in the standard Korean orthography, and it is generally assumed that [ŋ] in onset position is prohibited in Korean phonology.

Results of the survey on existing words

- Results show much variation across speakers and words.
- Insertion rates are calculated according to the factors mentioned above.
- Insertion rate = $\frac{\#\text{ of inserted responses}}{\#\text{ of both inserted and noninserted responses}}$

Overall insertion rate (existing words)

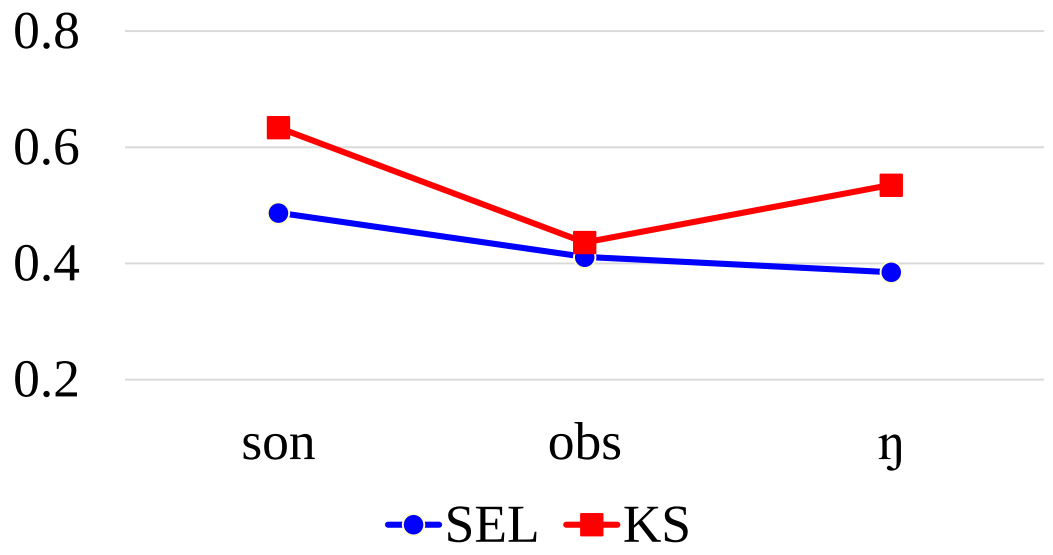
Seoul (sel)	Kyungsang (ks)
0.438	0.557



Insertion rate by **C_1 type** (existing words)

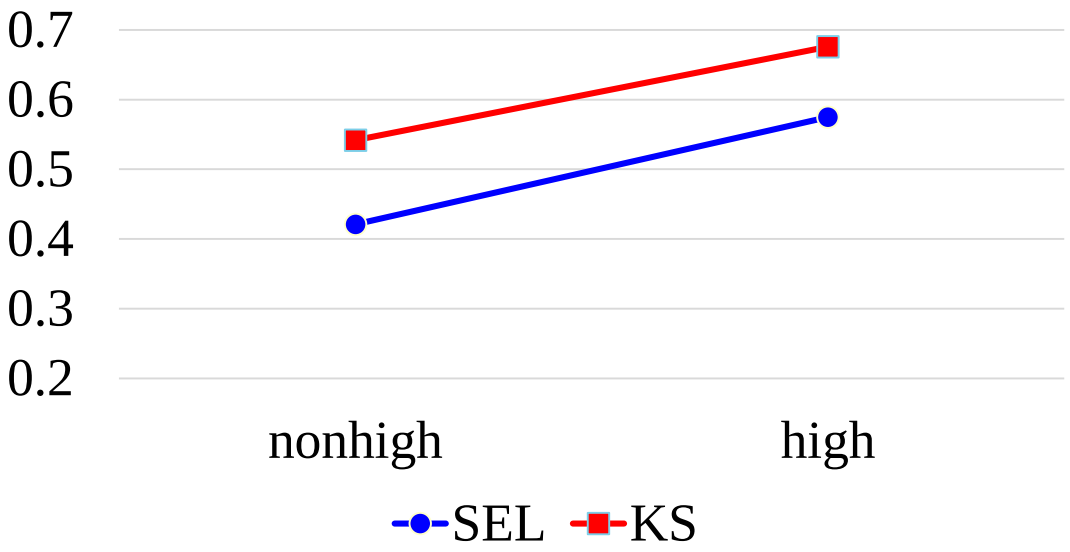
	son	obs	ŋ
SEL	0.487	0.411	0.385
KS	0.634	0.436	0.535
# word	138	72	93

(son = sonorants other than /ŋ/, obs = obstruents)



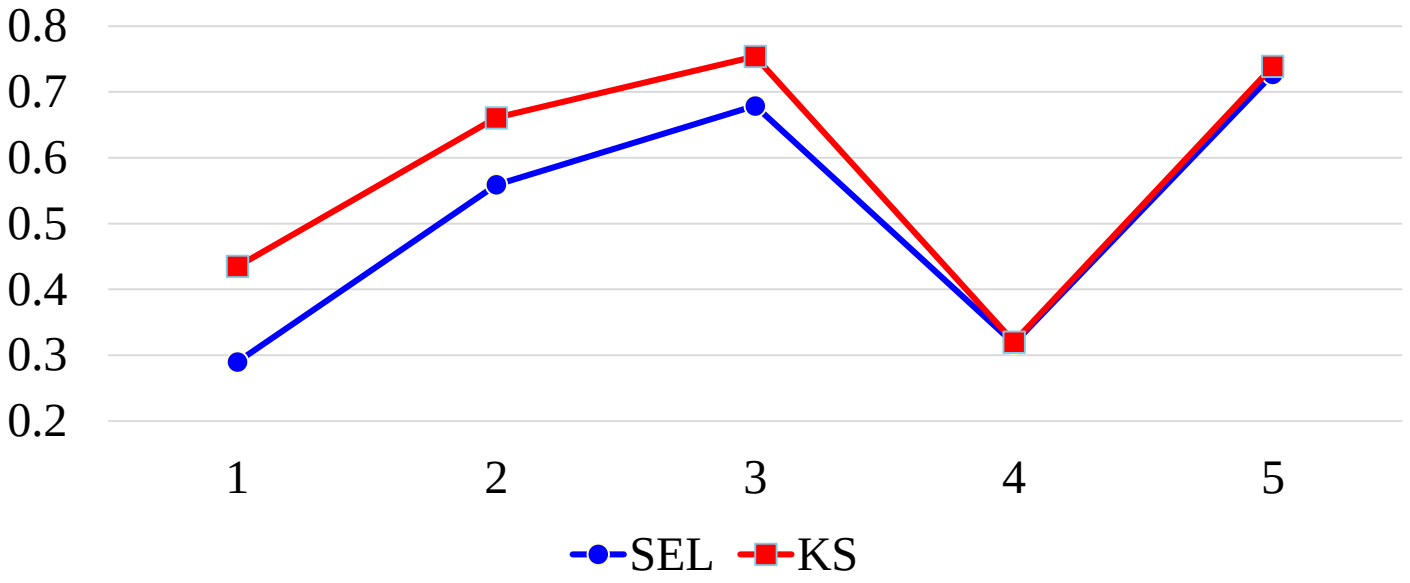
Insertion rate by **height** (existing words)

	nonhigh	high
SEL	0.421	0.575
KS	0.542	0.676
# word	269	34



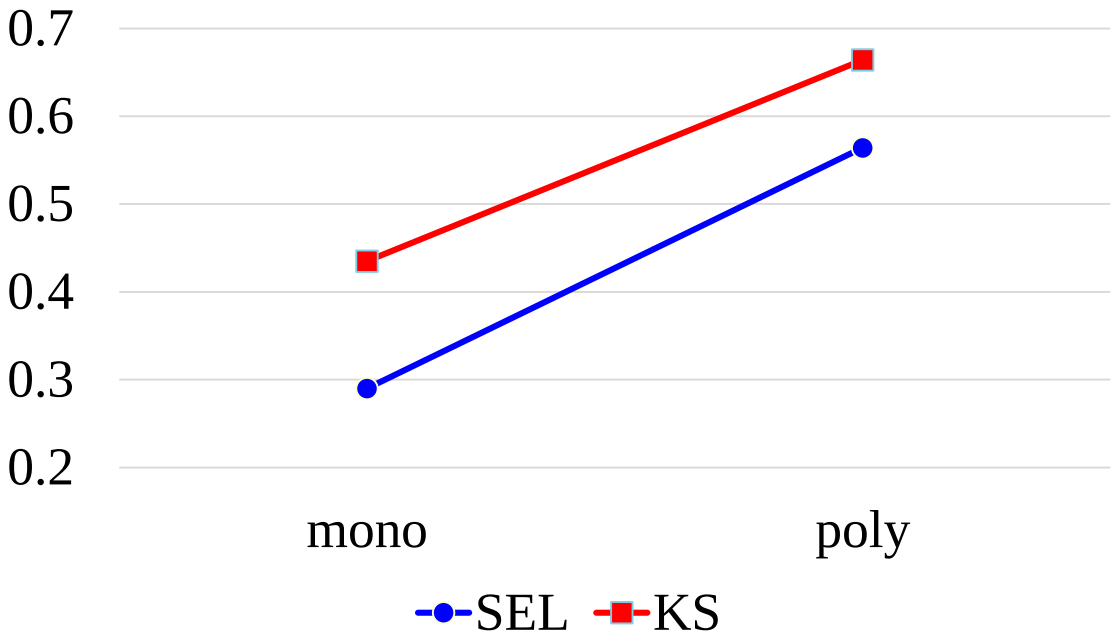
Insertion rate by $\mathbf{M_1}$ **syllable count** (existing words)

	1σ	2σ	3σ	4σ	5σ
SEL	0.29	0.559	0.679	0.318	0.727
KS	0.435	0.661	0.754	0.32	0.739
# word	140	154	**7**	**1**	**1**



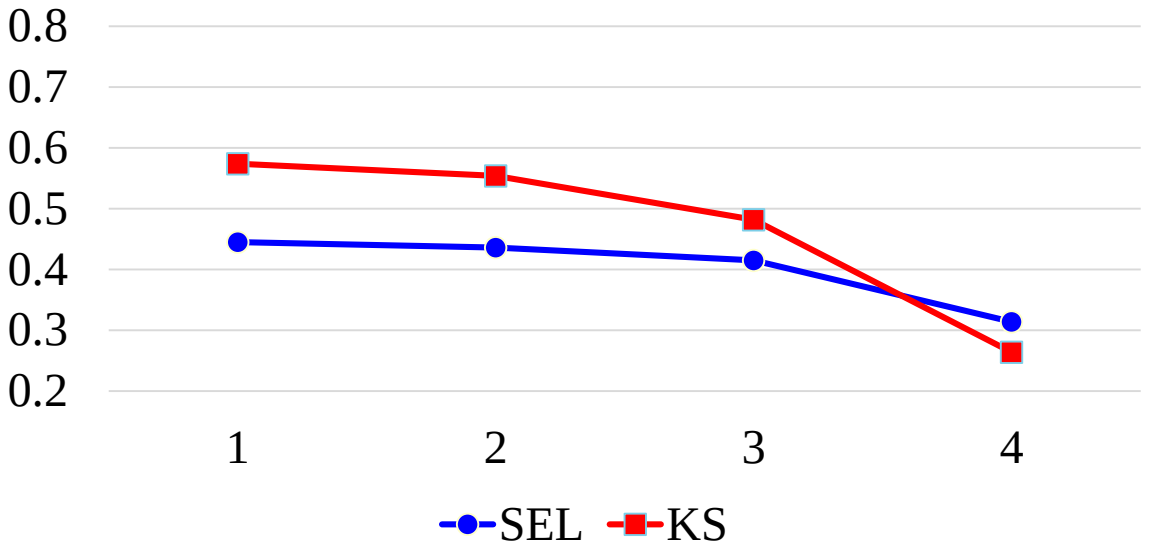
Insertion rate by $\mathbf{M_1}$ **syllable count** (existing words)

	mono σ	poly σ
SEL	0.29	0.564
KS	0.435	0.664
# word	140	163



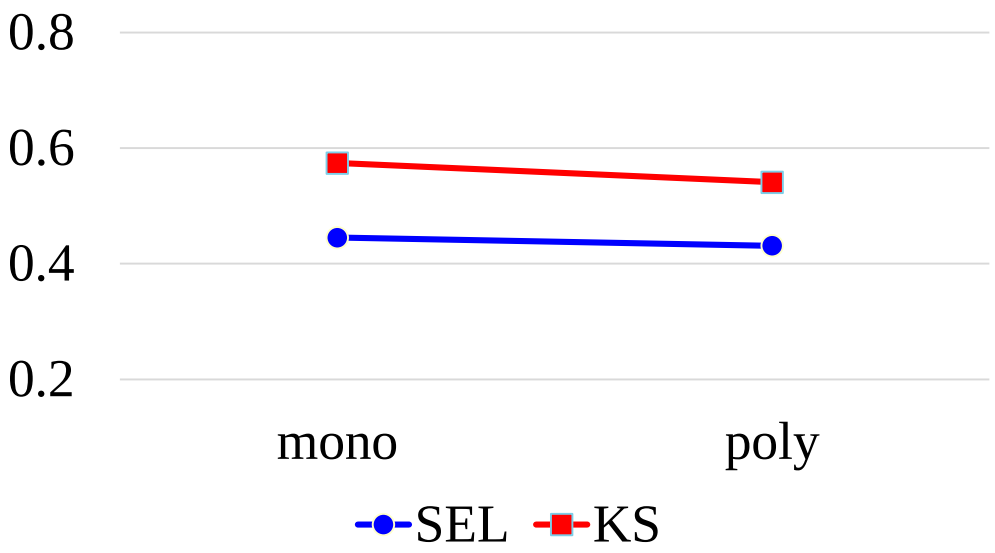
Insertion rate by **M_2 syllable count** (existing words)

	1σ	2σ	3σ	4σ
SEL	0.445	0.436	0.415	0.314
KS	0.574	0.554	0.482	0.264
# word	148	136	**16**	**3**



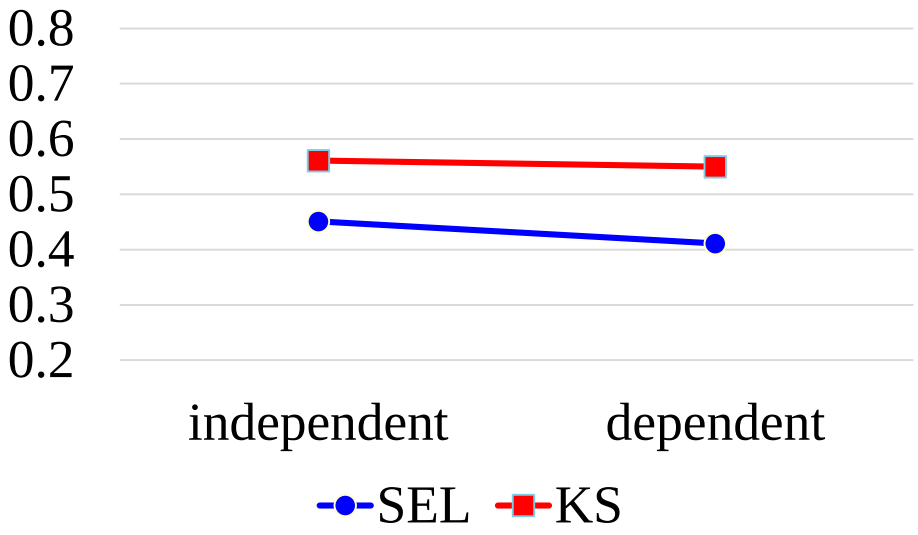
Insertion rate by **M_2 syllable count** (existing words)

	mono σ	poly σ
SEL	0.445	0.431
KS	0.574	0.541
# word	148	155



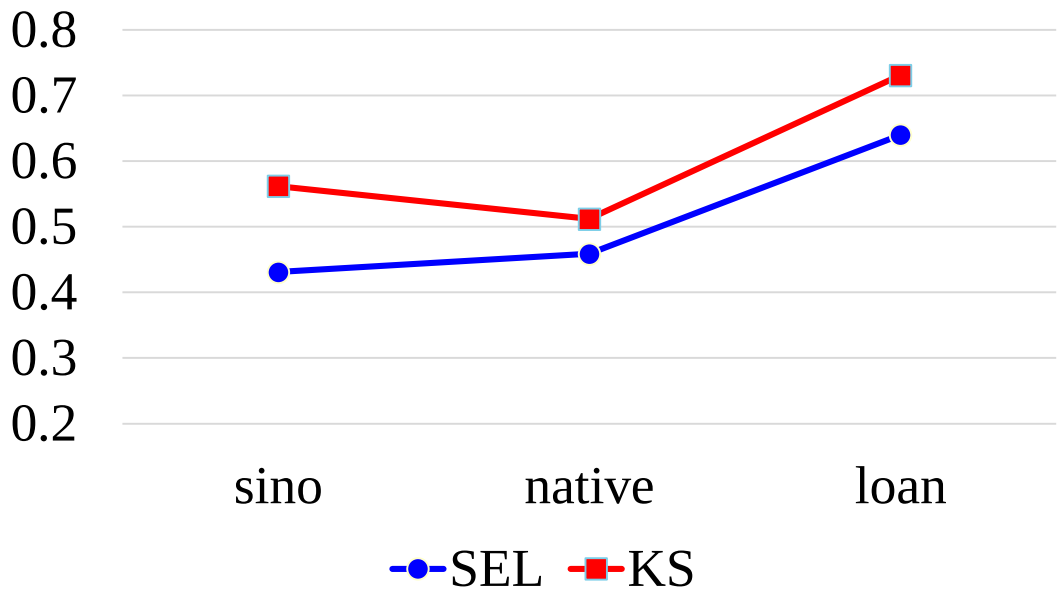
Insertion rate by **M_2 morpheme type** (existing words)

	independent	dependent
SEL	0.451	0.411
KS	0.561	0.55
# word	204	99



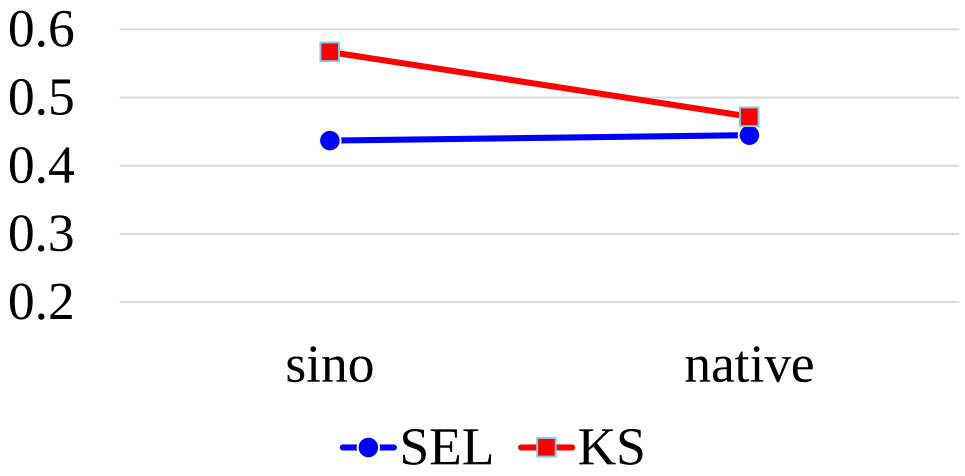
Insertion rate by **M_1 origin** (existing words)

	sino	native	loan
SEL	0.431	0.459	0.64
KS	0.562	0.512	0.731
# word	253	46	4



Insertion rate by **M_2 origin** (existing words)

	sino	native
SEL	0.437	0.445
KS	0.567	0.472
# word	270	33



A mixed effect logistic regression model

- The results of the present survey were fitted with the lmer function from the lme4 package (Bates et al. 2011) in R (R Development Core Team 2014).
- Dependent variable is binary, i.e., n-inserted or not.
- Each subject and each test word were included as random intercepts.

Fixed factors (A mixed effect logistic regression model)

- Dialect (Seoul, Kyungsang)
- C_1 type (son, obs, ŋ)
- Height (nonhigh, high)
- M_2 morpheme type (independent, dependent)
- M_1, M_2 syllable count (mono, poly)
- M_1, M_2 origin (native, sino, loan)
- Token frequency in Sejong corpus (log(sejong.freq + 1))

➢ C_1 type (backward difference coding: obs < ŋ < son), all other categorical variables (binary, sum coding)

Fixed effects: A logistic regression model (existing words)

	Estimate	Pr(>|z|)
(Intercept)	-.712	.001 **
Dialect (kyungsang)	.226	.171
C1type (ŋ-obs)	.717	<.001***
C1type (son-ŋ)	.538	<.001***
Height (high)	.360	<.001***
M2 stem (dep)	-1.181	<.001***
M1 syllable count (mono)	-.432	<.001***
M1 origin (other)	.166	.150
M2.origin (native)	-.701	<.001***
Dialect (Kyungsang):C1type(ŋ-obs)	.278	<.001***
Dialect (Kyungsang):C1type(son-ŋ)	.096	.058 .
M1 syllable count (mono):C1type(ŋ-obs)	.855	<.001***
M1 syllable count (mono):C1type(son-ŋ)	-.405	.009**
plus some other interactions		

- Negative numbers under estimate indicate a factor discourages n-insertion.

Main effects (existing words)

	estimate	Pr(>|z|)	
(Intercept)	-.712	.001 **	
Dialect (kyungsang)	.226	.171	→ more likely insertion for Kyungsang
C_1 type (ŋ-obs)	.717	<.001***	→ more likely insertion after /ŋ/ than obs
C_1 type (son-ŋ)	.538	<.001***	→ more likely after son than /ŋ/
Height (high)	.360	<.001***	→ more likely when /j/ precedes high V
M_2 type (dep)	-1.181	<.001***	→less likely before dependent M_2
M_1 syllable count (mono)	-.432	<.001***	→ less likely after mono σ M_1
M_1 origin (other)	.166	.150	
M_2 origin (native)	-.701	<.001***	→ less likely before native M_2

Summary: n-insertion in **existing** Korean words

n-insertion is **less likely** …

- after obstruent C_1 consonant. (Obstruency effect)
- after /ŋ/. (Velar nasal effect)
- after monosyllabic M_1. (**M_1 syllable count** effect)
- before a glide /j/ followed by a nonhigh vowel. (Height effect)
- before dependent M_2 (M_2 morpheme type effect)
- before native M_2 (M_2 origin effect)

➢ Dialect, M_1 origin, and word frequency were not statistically significant factors.

- For the purpose of finding out which of these effects can be extended to novel words, I conducted wug tests.
- Specifically, I investigated the following effects:
  - ✓ obstruency
  - ✓ velar nasal
  - ✓ height
  - ✓ M_1 syllable count

  plus

  - ✓ syllabicity (which will not be discussed here)

Roadmap

I. Basic patterns of Korean n-insertion

II. Existing words

III. Novel words

IV. Syllable count effect

Wug test: experimental token

- **loan word M_1** + wug stem M_2
- M_1 ends with one of seven consonants /m n ŋ l p s k/.

M_1					
test				control	
mono-σ		**di-σ**			
/t^{h}ap/	‘top’	/hiphap/	‘hiphop’	/hɛp^{h}i/	‘happy’
/has/	‘hot’	/siwis/	‘sweet’	/jellou/	‘yellow’
/k^{h}ik/	‘kick’	/pillæk/	‘black’	/pillu/	‘blue’
/s’əm/	‘some’	/sillim/	‘slim’	/silpə/	‘silver’
/t^{h}en/	‘ten’	/k^{h}illin/	‘clean’	/lɛti/	‘red’
/k^{h}iŋ/	‘king’	/wək^{h}iŋ/	‘working’		
/t^{h}ol/	‘tall’	/simol/	‘small’		

Wug test: Experimental token

- loan word M_1 + **wug stem M_2**

M_2	
test	control
/jucenol/ /jacenal/ /icɨpa/	/acepa/ /ekenol/

Wug test: Experimental token

- # of test items: 84 (2 syllable count x 7 coda x 2 repeating block).

	Example tokens	
	M_1	M_2
test (84)	t^hap pɬlæk hiphap	jucenol jacenol jucenal
control (84)	hɛp^{h}i t^hap pɬlæk	jucenol acepa ekenol

Wug test: participants

- 37 Seoul Korean speakers
- 32 Northern Kyungsang Korean speakers

➢ None of them participated in the existing word survey.

Wug test: procedure

- The experimenter told the participants that the given words are made-up compound nouns for new chemical products.
- They were instructed to choose their pronunciation of each of the given compounds, from the following three (or two) options written in Korean orthography.

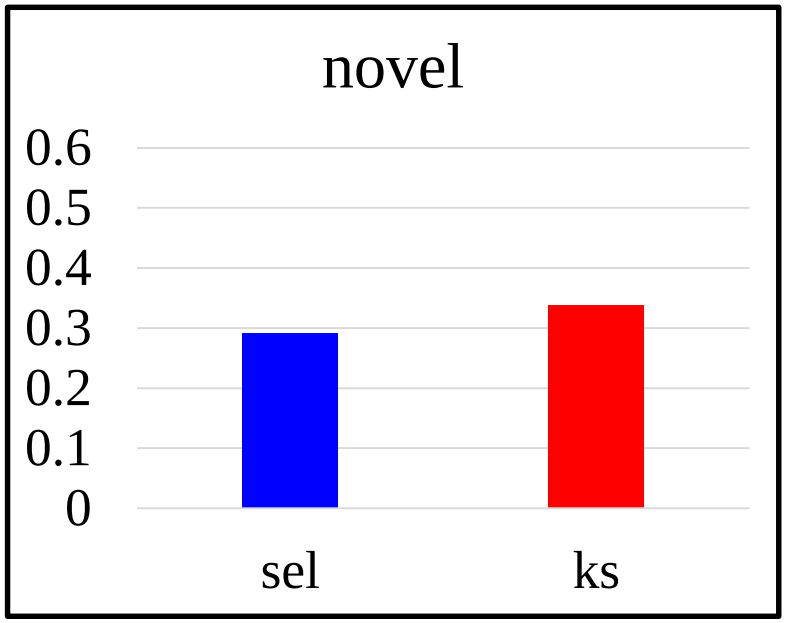
	e.g. s’əm+jucenol	k^hiŋ+jucenol
(i) **inserted** form:	s’əm.**n**ju.ce.nol	k^hiŋ.**n**ju.ce.nol
(ii) non-inserted (**resyllabified**)	s’ə.mju.ce.nol	--
(iii) non-inserted (**aligned**)	s’əm.ju.ce.nol	k^hiŋ.ju.ce.nol

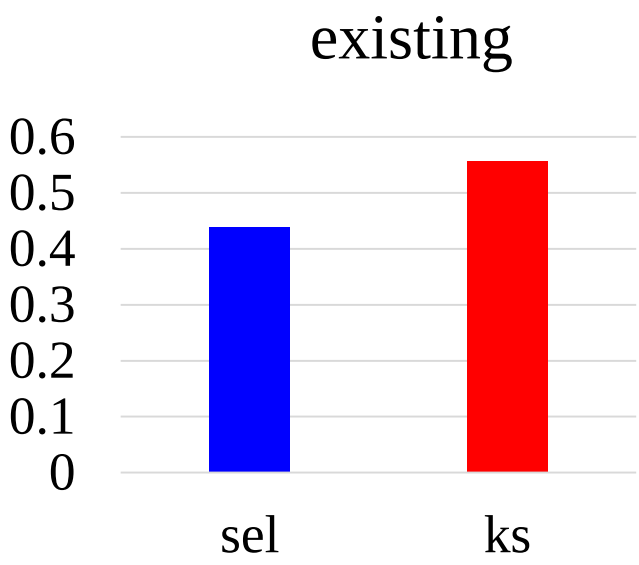
Wug test: Results

- Two Seoul Korean speakers inserted more frequently in control tokens than target tokens.
- Their responses were excluded from analysis.

Overall insertion rates (novel words)

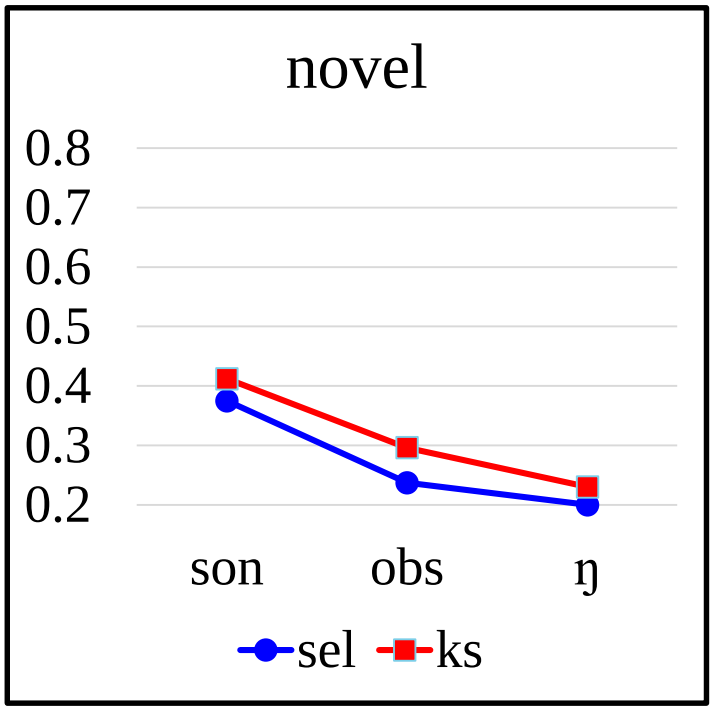
Seoul	Kyungsang
.291	.337

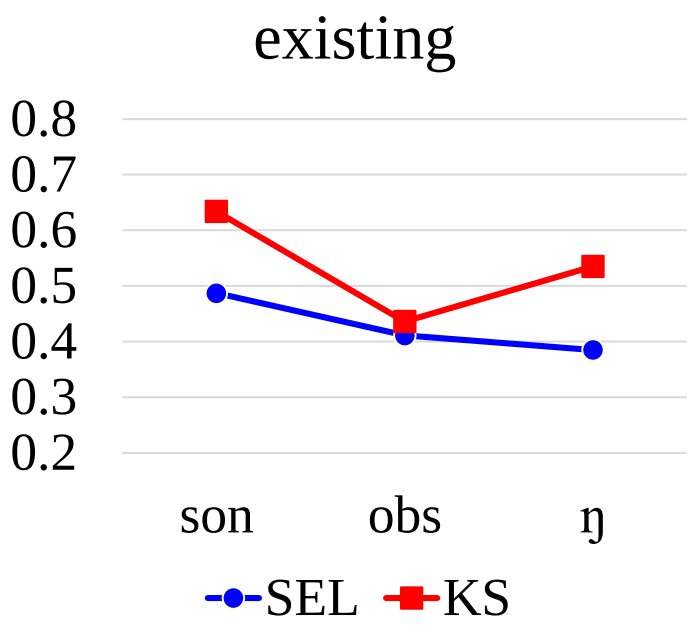




Insertion rate by C_1 **type** (novel words)

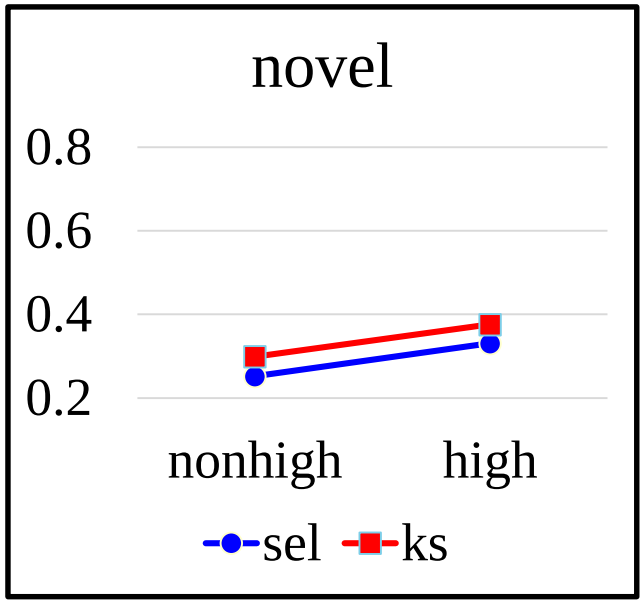
	son	obs	ŋ
SEL	0.375	0.237	0.2
KS	0.412	0.296	0.23

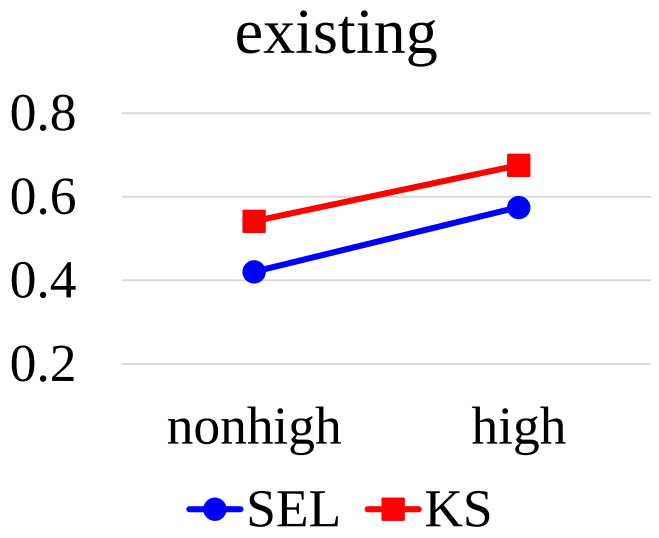




Insertion rate by **height** (novel words)

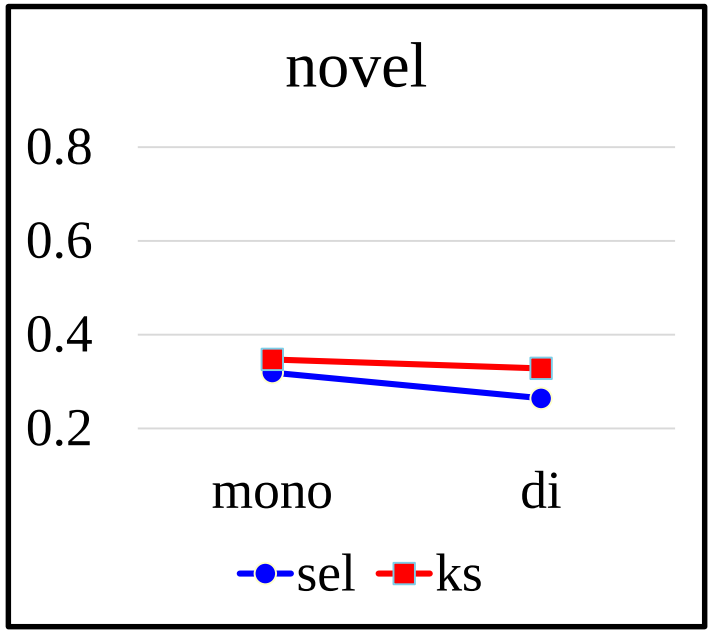
	nonhigh	high
SEL	0.252	0.331
KS	0.299	0.376

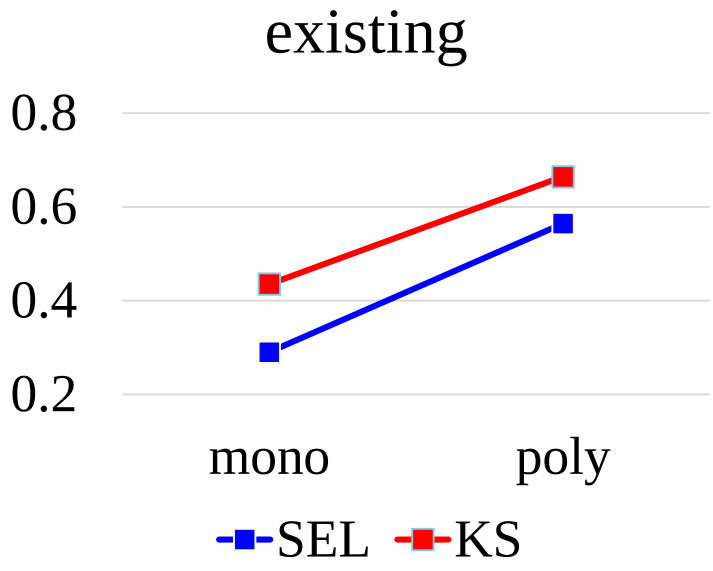




Insertion rate by **M_1 syllable count** (novel words)

	mono-σ	di-σ
SEL	**0.319**	**0.264**
KS	**0.347**	**0.328**





A mixed effect logistic regression model

- The results of the present survey were fitted with the lmer function from the lme4 package (Bates et al. 2011) in R (R Development Core Team 2014).
- Dependent variable is binary, i.e., n-inserted or not.
- Each subject and each test word were included as random intercepts.

Fixed factors (A mixed effect logistic regression model)

- Dialect (Seoul, Kyungsang)
- C_1 type (son, obs, ŋ)
- Height (nonhigh, high)
- M_1 syllable count (monosyllabic, disyllabic)

➢ C_1 type (backward difference coding: obs < ŋ < son), all others (binary, sum coding)

Fixed effects: A logistic regression model (novel words)

	estimate	Pr(>|z|)	
(Intercept)	-1.371	<.001***	
C_1type (ŋ-obs)	.422	.101	→ No significant difference between /ŋ/ and obs
C_1type (son-ŋ)	.811	<.001***	→ more likely after son than /ŋ/
Height (high)	.268	.001**	→ more likely when /j/ precedes high V

- Obstruency, velar nasal and height effects were significant.
- There were no dialectal and M_1 syllable count effects.
- A model with M_1 syllable count as a fixed factor is not significantly different from the above model according to model comparison.

Mixed effect model with M_1 syllable count as fixed factor

	Estimate	Pr(>|z|)
(Intercept)	-1.372	<.001***
C_1 type (ŋ-obs)	.422	.101
C_1 type (son-ŋ)	.812	<.001***
Height (high)	.268	.001**
Syllable count (mono)	.134	0.088

- There is no M_1 syllable count effect.

Trends in Korean n-insertion: existing vs. **novel** words

- (O = confirmed; X = not confirmed)

Effect	Existing	Novel
dialect	X	X
obstruency	O	O
velar nasal	O	O
height	O	O
M_1 syllable count	**O**	**X**

Roadmap

I. Basic patterns of Korean n-insertion

II. Existing words

III. Novel words

IV. Syllable count effect

Syllable count effect

- The results of the present study suggest that Korean speakers know most of the trends in the distribution of existing Korean words.
- They use this knowledge when they apply n-insertion to novel words.
- However, the results of the present study also suggest that Korean speakers do not know all statistically prominent patterns in existing words.

Syllable count effect

- When M_1 is polysyllabic (cf. monosyllabic), insertion rate is higher in existing Korean words.
- But such higher rate of insertion in words with polysyllabic M_1 was not mirrored in novel Korean words.
- This means that Korean speakers are not aware of the syllable count effect in existing words.

Syllable count effect

- Why did Korean speakers fail to learn the trend about the length of M_1?
- An answer from my previous study (Jun 2015):

  The length of M_1 is not a phonologically natural factor which can condition the application of n-insertion or affect the probability that n-insertion applies.

- Then, failure to learn the syllable count effect can be understood under the hypothesis that only phonologically natural patterns can be learned at least with ease (Becker et al. 2011; Hayes et al. 2009; Hayes & White 2013).

Syllable count effect

- Is it always the case that syllable count (or word length) effect is accidental and not learnable?
- Is syllable count generally considered a phonologically unnatural factor?

Syllable count effect

- Relevant previous studies show mixed results.

- Were word length effects found in existing words productively extended in novel words?
  - ✓ **No**: Ito's (2014) study on Korean compound tensing.
  - ✓ **Yes**: Becker et al.'s (2011) study on laryngeal alternation in Turkish lexicon and references therein.

Syllable count effect: Ito (2014)

- Korean compound tensing

W_1	W_2	phonetic form
/hɛ/	/pic^h/	[hɛp'it]
'sun'	'light'	'sun light'
/kaɨl/	/pi/	[kaɨlp'i]
'autumn'	'rain'	'autumn rain'
/pom/	/pi/	[pomp'i]
'spring'	'rain'	'spring rain'
/pok/	/cuməni/	[pokc'uməni]
'luck'	'pocket'	'lucky bag'

- Syllable count effect found in existing Korean compounds: tensing rate is lower for longer words.

Syllable count effect: Ito (2014)

- The syllable count effect was not mirrored in the novel Korean words.

> …a factor that **cannot** be straightforwardly explained as a **universal preference (a syllable-count effect)** is not productively generalized in the novel words. (Ito 2014: 355, emphasis added)

Syllable count effect: Becker et al. (2011)

- Laryngeal alternations in Turkish are observed at the right edges of nouns.

	bare stem	possessive
alternating	gy**tʃʰ**	gy**dʒ**-y
nonalternating	a**tʃʰ**	a**tʃʰ**-~~ɨ~~

- Several phonological properties including the noun's size correlate with stem-final alternations.
- While above **60% of polysyllables alternate**, most **monosyllables do not**.

Syllable count effect: Becker et al. (2011)

- Becker et al. (2011) reports that word length plays a role in explaining the rate of laryngeal alternation in Turkish lexicon, arguing that **word length is a phonologically natural factor**.
- This word length effect was mirrored in novel words, and thus Turkish speakers know this word length effect.

Syllable count effect

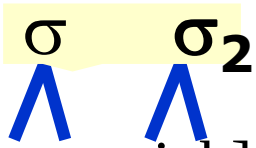
- Then, the question is why word length can condition laryngeal alternation rate in Turkish but it cannot n-insertion (and compound tensing) in Korean.
- Is word length phonologically natural or unnatural?

Syllable count effect: Turkish laryngeal alternation

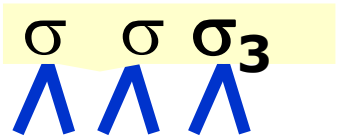
- Notice that Becker et al. do not argue that word length per se is a phonologically natural factor. They interpret the word length effect as word-initial syllable faithfulness.
- The reason why final stops in monosyllabic words are less likely to alternate in Turkish is because they are in the initial syllables which can be protected by **positional faithfulness for initial syllables** (in Optimality-Theoretic terms).

Syllable count effect: Korean n-insertion

- Let us now consider whether syllable count effect in Korean n-insertion is also positional faithfulness effect.
- Notice that /n/ is inserted at the beginning of M_2. Thus even when M_1 is shortest, n-insertion occurs in the second syllable, and thus initial syllable faithfulness should not be at work, regardless of whether M_1 is monosyllabic or polysyllabic.



i. **mono-**σ M_1 /co**m**-jak/ [com.**n**jak] ‘mothball’



ii. **poly-**σ M_1 /p^hii**m**-jak/ [p^hi.im.**n**jak] ‘birth control pill’

Syllable count effect: Korean n-insertion

- But, if M_1 ends with an obstruent, n-insertion would cause its nasalization. Then, initial syllable faithfulness may prevent monosyllabic M_1-final obstruents from undergoing nasalization, resisting n-insertion.
- In contrast, C_1 obstruents of polysyllabic M_1 cannot be protected by initial syllable faithfulness.
- **C_1 = obs**

  σ_1 σ σ

  i. **mono**-σ M_1 /hɛ**k**-jəlljo/ [hɛ**ŋ**.**n**jəl.ljo] ‘nuclear fuel’

  σ σ_2 σ

  ii. **poly**-σ M_1 /soto**k**-jak/ [so.to**ŋ**.**n**jak] ‘disinfectant’

Syllable count effect: Korean n-insertion

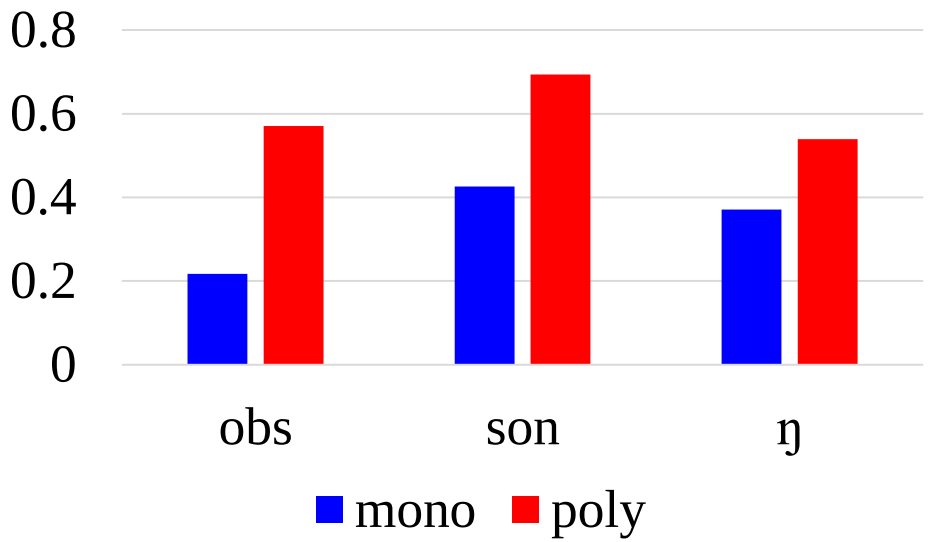
- When C_1 = sonorant, initial syllable faithfulness cannot be activated, regardless of whether M_1 is mono-σ or poly-σ.
- **C_1 = sonorant**

i. mono-σ M_1	/co**m**-jak/ [co**m.n**jak]	‘mothball’
ii. poly-σ M_1	/p^{h}ii**m**-jak/ [p^{h}i.i**m.n**jak]	‘birth control pill’

➢ Accordingly, higher rate of n-insertion in words with polysyllabic M_1 is expected only when C_1 = obstruent.

Syllable count effect: Korean n-insertion

- Insertion rate by C_1 type and M_1 syllable count (existing words)



- Insertion rates are higher in polysyllabic M_1 across different C_1 types.

Syllable count effect: Korean n-insertion

- Insertion rate by C_1 type and M_1 syllable count (existing words)

	obs	son	ŋ
mono-σ	.217	.426	.371
poly-σ	.571	.694	.539
difference	**-.354**	**-.268**	**-.168**

- Since the syllable count effect in Korean n-insertion is not confined to words with obstruent-final M_1, it cannot be attributed to the initial syllable faithfulness.
- This may explain why syllable count effect was extended to novel words in Turkish but wasn't in Korean.

Conclusion

- Most of the tendencies in existing Korean words were mirrored in novel words, suggesting that Korean speakers know these tendencies.
- However, syllable count effects were not mirrored in the results of the experiments with novel words, suggesting that Korean speakers failed to generalize the tendency involving syllable count to novel words.

Conclusion

- The syllable count effect in Korean n-insertion can hardly be attributed to initial syllable faithfulness.
- Thus, unlike word length in Turkish laryngeal alternation explored by Becker et al. (2011), the syllable count in Korean n-insertion may not be a natural factor.
- Accordingly, the syllable count effect in Korean n-insertion can be considered accidental and Korean speakers' failure to learn it can be understood under the hypothesis that only phonologically natural patterns can be learned at least with ease.

References

Bates, Douglas; Martin Maechler & Ben Bolker (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-39. http://CRAN.R-project.org/package=lme4.

Becker, Michael; Nihan Ketrez & Andrew Nevins (2011). The surfeit of the stimulus: analytic biases filter lexical statistics in Turkish laryngeal alternations. In *Language* 87.1: 84-125.

Hayes, Bruce & James White (2013). Phonological naturalness and phonotactic learning. In *Linguistic Inquiry* 44.1: 45-75.

Hayes, Bruce; Kie Zuraw; Péter Siptár & Zsuzsa Londe (2009). Natural and Unnatural Constraints in Hungarian Vowel Harmony. In *Language* 85: 822-863.

Hwang, Sangji (2008) *Korean speakers' knowledge of /n/-insertion: P-map approach*. MA thesis, Seoul National University.

Ito, Chiyuki (2014) Compound tensification and laryngeal co-occurrence restrictions in Yanbian Korean. In *Phonology* 31.3: 349-98.

Jun, Jongho (2015) Korean n-insertion: a mismatch between data and learning. In *Phonology* 32.3: 417-58.

Kwuklip kwuke yenkwuwen [National Institute of the Korean Language] (ed.) (1999). *Standard Korean dictionary*. [*Pyocwun kwuke taysacen*.] Seoul: Doosan Dong-A.

R Development Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.