Phonological variation in Seoul Korean

n-insertion

Jongho Jun
(Seoul National University)
Roadmap

I. Introduction
II. Trends in existing words
III. Analysis
IV. Learning simulation
V. Wug test
I. Introduction: n-insertion in Seoul Korean

- $\emptyset \rightarrow n / C_1 [M_1 - [M_2 i/j]$
  
  $(M_{1,2} = \text{morpheme}; C_1 = M_1\text{-final consonant})$

i. prefix-stem

- $/təs-jaŋmal/ [tənjaŋmal] \quad \text{‘anklet socks’}$
- $/hoth-ipul/ [honnipul] \quad \text{‘unlined comforter’}$

ii. compound

- $/com-jak/ [comnjak] \quad \text{‘mothball’}$
- $/som-ipul/ [sominipul] \quad \text{‘cotton sheet’}$

iii. Phrase

- $/mək-il # jəs/ [məkilljət] \quad \text{‘taffy that (someone) will eat’}$
- $/han # il/[hannil] \quad \text{‘thing that (someone) has done’}$
Epenthetic consonant /n/

- phonetically realized as palatalized coronal sonorants, [ɲ] and [ʎ] (after a lateral)
- /som-ipul/ (\(\rightarrow\) [som.ɲi.pul] \(\rightarrow\) [som.ɲi.pul])
- /sol-ip^h/ (\(\rightarrow\) [sol.nip] \(\rightarrow\) [sol.lip] \(\rightarrow\)) [sol.ʎ.ʎip]

- palatalization
- lateralization
C1 is always phonetically realized as a nasal.

- **Obstruent Nasalization**: an obstruent becomes a nasal before a nasal.

  /kiʔ-p jesan/ [ki.ʔm.nje-san] ‘corporation budget’

  /paθʰ-ilanŋ/ [paŋ.nilaŋ] ‘the ridge of a field’

- Except when the C1 is a lateral.
N-insertion is an **optional** process.

/kəmʃəl/  [kəmʃəl] ~ [kəmʃnʃəl] ‘censorship’

/hɔtʰ-ipul/  [hotipul] ~ [honnipul] ‘unlined comforter’


> Is there any phonological tendency within variation involved in n-insertion?
Phonological effects in novel words (Hwang 2008)

Hwang performed a productivity test on Seoul Korean n-insertion, employing loan and wug words in Korean.

N-insertion is less likely …

- before [i] than before [j]. (Syllabicity effect)
- after obstruents. (Obstruency effect)
- after /ŋ/ cf. other sonorants. (Velar nasal effect)
Questions

• Where do these asymmetries come from?
• Why do Seoul Korean speakers apply n-insertion to novel words less frequently before [i] and after obstruents and /ŋ/?

• One possible answer
  ➢ They do so with existing Korean words.
  ➢ In the production of novel words, they use their knowledge about existing Korean words.
• To see whether and, if so, to what extent this is true, I have investigated the n-insertion patterns of existing Seoul Korean words.
II. Trends in existing words

Three sources

• Dictionary
• Previous survey and experimental studies
• My own survey
Summary: dictionary and previous survey/experiments

• (O = confirmed; X = not confirmed)

<table>
<thead>
<tr>
<th>Effect</th>
<th>Dictionary</th>
<th>Previous survey/experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>syllabicity</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>obstruency</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>velar nasal</td>
<td>X</td>
<td>O</td>
</tr>
</tbody>
</table>
Survey: method

- **Test words:** 304 multi-morphemic words with /j/-initial M2 in my dictionary database.
- **Participants:** 22 Seoul Korean speakers
- **Survey form**
  - Both inserted and non-inserted forms for 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 the following three (or two) choices:
### Response choices

<table>
<thead>
<tr>
<th>(i) inserted form:</th>
<th>e.g. tok+jak</th>
<th>noŋ+jak</th>
</tr>
</thead>
<tbody>
<tr>
<td>toŋ.njak</td>
<td>noŋ.njak</td>
<td></td>
</tr>
<tr>
<td>(ii) non-inserted form (resyllabified)</td>
<td>to.kjak</td>
<td>--</td>
</tr>
<tr>
<td>(iii) non-inserted form (aligned)</td>
<td>tok.jak</td>
<td>noŋ.jak</td>
</tr>
</tbody>
</table>
Results of the current survey

- Insertion rate by C1 types (# of words shown in parentheses; son = [m, n, l])

<table>
<thead>
<tr>
<th>son</th>
<th>obs</th>
<th>η</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.49 (138)</td>
<td>0.41 (72)</td>
<td>0.38 (93)</td>
</tr>
</tbody>
</table>

- The results support both **obstruency** and **velar nasal** effects.
A mixed effect logistic regression model

- The results of the present survey are 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)

(underlined= reference level)

- C1 type (son, obs, η)
- M2-initial vocoid (jV: ja, jə, je, jo, ju)
- M1 syllable number (sylnum1)
- M2 syllable number (sylnum2)
- M1 origin (native, sino, loan)
- M2 origin (native, sino)
- Token frequency in Sejong corpus (log(sejong.freq + 1))
Results: A logistic regression model (Fixed effects)

|                          | Estimate  | Std. Error | z value | Pr(>|z|) |
|--------------------------|-----------|------------|---------|---------|
| (Intercept)              | -1.42277  | 0.54856    | -2.594  | 0.00950 ** |
| C1 = η                   | -0.52689  | 0.20386    | -2.585  | 0.00975 ** |
| C1 = obs                 | -0.67327  | 0.21823    | -3.085  | 0.00203 ** |
| jV = jœ                  | 0.05239   | 0.21762    | 0.241   | 0.80975  |
| jV = je                  | 0.97724   | 0.76042    | 1.285   | 0.19875  |
| jV = jo                  | 0.18659   | 0.28376    | 0.658   | 0.51081  |
| jV = ju                  | 0.90350   | 0.30948    | 2.919   | 0.00351 ** |
| sylnum1                  | 1.40994   | 0.15263    | 9.237   | < 2e-16 *** |
| sylnum2                  | -0.18613  | 0.14169    | -1.314  | 0.18898  |
| M1.origin = loan         | -0.36726  | 0.78783    | -0.466  | 0.64110  |
| M1.origin = sino         | -0.60207  | 0.31764    | -1.895  | 0.05803 . |
| M2.origin = sino         | -0.16558  | 0.36607    | -0.452  | 0.65105  |
| log(sejong.freq + 1)     | -0.08942  | 0.08665    | -1.032  | 0.30211  |

- Significant factors: marked in boldface
- Negative numbers under estimate indicate a factor discourages n-insertion.
Height & length effects

• Insertion rate by M2-initial vocoid sequence

<table>
<thead>
<tr>
<th></th>
<th>ja</th>
<th>jə</th>
<th>je</th>
<th>jo</th>
<th>ju</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insertion rate</td>
<td>0.40</td>
<td>0.42</td>
<td>0.47</td>
<td>0.44</td>
<td>0.57</td>
</tr>
<tr>
<td># word</td>
<td>73</td>
<td>149</td>
<td>4</td>
<td>43</td>
<td>34</td>
</tr>
</tbody>
</table>

• Insertion rate by M1 length
  a. Monosyllabic: \textbf{0.29} (140 words)
  b. Disyllabic: \textbf{0.56} (154 words)
  c. Longer: Trisyllabic 0.67 (7 words), Quadrisyllabic 0.31 (1 word), Pentasyllabic 0.72 (1 word)
Summary: n-insertion in existing Korean words

N-insertion is …

• less likely before /i/ than before /j/. (Syllabicity effect)
• less likely after obstruent C1 consonant. (Obstruency effect)
• less likely after /ŋ/. (Velar nasal effect)
• more likely before a glide /j/ followed by a high vowel. (Height effect)
• less likely with monosyllabic M1. (Length effect)
III. Analysis

- I analyze the observed patterns of Seoul Korean n-insertion by adopting Optimality-theoretic (OT) constraints which are assigned numerical weights within the framework of the maxent model.

- I discuss what constraints are responsible for the occurrence of Seoul Korean n-insertion and the observed asymmetric phonological trends.

- The specific weights of the constraints are determined by the results of a learning simulation.
Main variants

<table>
<thead>
<tr>
<th>Variants</th>
<th>som+ipul</th>
<th>com+jak</th>
<th>Constraints violated</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Inserted</td>
<td>somnipul</td>
<td>comnjak</td>
<td>DEP-n, *CODA</td>
</tr>
<tr>
<td>b. Resyllabified</td>
<td>sommipul</td>
<td>commjak</td>
<td>ALIGN-R</td>
</tr>
<tr>
<td>c. Aligned</td>
<td>somipul</td>
<td>i. com[ja]nk</td>
<td>*CODA, Onset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ii. com[ja]nk</td>
<td>*CODA, V-NUC</td>
</tr>
</tbody>
</table>

- **ALIGN-MORPH-R (ALIGN-R)**
  The right edge of a morpheme coincides with the right edge of a syllable.

  - Adopted by most previous studies on Korean n-insertion.
  - Active in Korean, independently of n-insertion.
ALIGN-R

VCV → V.CV

• An intervocalic consonant (except /ŋ/) is always syllabified as an onset: /aka/ [a.ka] ‘baby’.

VC+V → V.CV ~ VC.V

• E.g., compounding.
• The intervocalic consonant is variably syllabified between an onset or a coda.
  i. /sikol+ai/ [si.ko.ra.i] ~ [si.kol.a.i] ‘a country-bred child’
     cf. /sikol/ ‘country’, /ai/ ‘child’
  ii. /mul+ori/ [mu.ro.ri] ~ [mul.o.ri] ‘a wild duck’
     cf. /mul/ ‘water’, /oli/ ‘duck’
• N.B. Allophonic variation of a singleton liquid: [l] in the coda and [r] in the onset.
Main variants

<table>
<thead>
<tr>
<th></th>
<th>som+ipul</th>
<th>com+jak</th>
<th>Constraints violated</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>Inserted</td>
<td>som.nipul</td>
<td>DEP-n, *CODA</td>
</tr>
<tr>
<td>b.</td>
<td>Resyllabified</td>
<td>so.m[j]ipul</td>
<td>co.m[j]ak</td>
</tr>
<tr>
<td>c.</td>
<td>Aligned</td>
<td>som.ipul</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>i. com.[ja]_Nk</td>
<td>*CODA, ONSET</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ii. com.j[a]_Nk</td>
<td>*CODA, V-NUC</td>
</tr>
</tbody>
</table>

i. Glides in Korean are syllabified in the nucleus.


ii. Glides may occupy the onset when it’s empty (J. Park 2001, Y. Yun 2004).
   I assume that (ii), not (i), is the attested variant.

- **V-NUC:** [-consonantal] segments must be in the nucleus.
  (Rubach 2000)
Constraints: phonological effects

- **Obstruency** effect
  - Due to automatic **Obstruent Nasalization** in Korean
  - **IDENT**(sonorant)

<table>
<thead>
<tr>
<th>Example</th>
<th>[ pronunciation ]</th>
<th>Proforma</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. /tok+jak/</td>
<td>[toŋ.njak]</td>
<td>*</td>
</tr>
<tr>
<td>b. /com+jak/</td>
<td>[com.njak]</td>
<td>✓</td>
</tr>
</tbody>
</table>
Velar nasal effect

- In Korean, /ŋ/ is special.
- No Korean words begin with /ŋ/.
  
  *\[\sigma \eta\] “No velar nasal in the (word-initial) onset.”
  

- Does this constraint prevent the **intervocalic** /ŋ/ from occupying the onset position?
Velar nasal between vowels

• Two options
  ➢ syllabified exclusively in the coda
  ➢ Ambisyllabic

(i) \( \sigma \quad \sigma \quad V \, \eta \, V \)

(ii) \( \sigma \quad \sigma \quad V \, \eta \, V \)
Ambisyllabic /ŋ/

- “When the velar nasal occurs between vowels, it is difficult to determine whether it is a coda of a preceding syllable or an onset of the following syllable.” (Huh 1984: 208)

- Nasal consonant duration
  - coda >> onset
  - coda >> VŋV ≈ onset
Nasal consonant duration

- From Y. Hwang (2002) (msec; C = lenis stop)

<table>
<thead>
<tr>
<th></th>
<th>coda (blue)</th>
<th>onset (yellow)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>V_CV</td>
<td>V_V</td>
</tr>
<tr>
<td>m</td>
<td>179.12 (43.01)</td>
<td>66.27 (9.78)</td>
</tr>
<tr>
<td>n</td>
<td>170.77 (36.54)</td>
<td>80.22 (9.08)</td>
</tr>
<tr>
<td>ñ</td>
<td>171.47 (38.10)</td>
<td>87.15 (28.74)</td>
</tr>
</tbody>
</table>

- VñV: much shorter than obvious coda nasals.
Avoidance of ONSET violation in Korean

- E.g. vocative suffix allomorphs
  - a when the noun ends in a consonant: e.g. /tasom-a/
  - ja when the noun ends in a vowel: e.g. /somi-ja/
- /-ja/ is chosen to satisfy ONSET.

- When the noun ends in /ŋ/, -a is chosen.
  e.g. /saranŋ-a/, not */saranŋ-ja/
  - This suggests that VŋV satisfies ONSET.
  - Thus, the velar nasal cannot be exclusively syllabified in the coda.
Velar nasal effect

• The velar nasal in Korean can be ambisyllabic unlike other consonants.
• Under the assumption that an ambisyllabic /ŋ/ may satisfy both ONSET and ALIGN-R, words with /ŋ/ C1 would be more likely to surface as such, resisting n-insertion.
• *AMBI SYLLABIC (*AMBI)
  “The (velar nasal) consonant is not ambisyllabic.”
Remaining effects: parameterized DEP

- **Syllabic**ity effect: $\text{DEP-n/}_i > \text{DEP-n/}_j$
- **Height** effect: $\text{DEP-n/}_j[-\text{high}] > \text{DEP-n/}_j[+\text{high}]$
- Epenthetic /n/: $\text{DEP-t/}_{\{i, j\}}, \text{DEP-m/}_{\{i, j\}} \ldots > \text{DEP-n/}_{\{i, j\}}$
- Before high front vocoids:
  
  $\text{DEP-n/}_a, \text{DEP-n/}_u \ldots > \text{DEP-n/}_{\{i, j\}}$

- These ranking/weight differences may be understood under P-map (Steriade 2001, 2009)
P-map (Steriade 2001, 2009)

- A segment can be inserted or deleted only when its insertion or deletion is not so obtrusive in perception.
- The segment most confusable with zero is predicted to be chosen as the epenthetic segment.
- Its choice is determined based on a context-dependent hierarchy of perceptual similarity between the input and output.
- Common epenthetic $C = [?]$, [h], [j, w] (next to homorganic V’s)
  ✓ No coarticulatory influence on the neighboring segments
  ✓ $V-?V$, $V-hV$, or $V-GV$, are the most similar input-output pairs.

Internal ranking of DEP constraints:

$$\text{DEP-}p,t,k/_V > \text{DEP-}?/_V$$
P-map in Korean

- /n/ $\rightarrow$ [ŋ] / __ {i, j}  (Allophonic palatalization)
- Input-output pairs involving n-insertion are similar just like V-GV
  - i-ni  (or, more precisely, i-ñi)
  - j-nj  (or j-ñj)

Cf. i-ti, i-mi …; a-na, u-nu …

\[ \text{DEP-t/\_}{i, j}, \text{DEP-m/\_}{i, j} .. \]
\[ > \text{DEP-n/\_}{i, j} \]
\[ \text{DEP-n/_a, DEP-n/_u} .. \]
More on epenthetic /n/ before /i, j/

• /i, j/ are high vocoids with a **low frequency resonance**, and they are **acoustically like nasals** (Borden & Harris 1984).

• Camouflage: placing a new segment right next to an existing segment similar to it must not be prominent.

• Inserting a nasal (i.e. /n/) next to a nasal-like segment (i.e. /i, j/) would be less prominent than …
  ✓ inserting a **non-nasal** segment to a nasal-like segment
  ✓ inserting a nasal **to a true oral segment**.
P-map: Syllabicity effect

- /nj/ is phonetically realized as almost (if not completely) a single segment [ɲ] (O. Kang 2003; Lee & Lee 2006)

/nj/ → [ɲ(j)]

• Input-output pairs

<table>
<thead>
<tr>
<th></th>
<th>j-nj</th>
<th>i-ni</th>
</tr>
</thead>
<tbody>
<tr>
<td>palatalization</td>
<td>j-ɲj</td>
<td>i-ɲi</td>
</tr>
<tr>
<td>j-reduction</td>
<td>j-ɲ</td>
<td>i-ɲi</td>
</tr>
<tr>
<td>similarity</td>
<td></td>
<td>&gt;&gt;</td>
</tr>
</tbody>
</table>

• DEP-n/_i > DEP-n/_j
P-map: Height effect

- N-insertion would not be prominent before /i, j/.
- One reason: they are high vocoids with a low frequency resonance, and they are acoustically like nasals.
- If /j/ is followed by a high vowel, its low frequency resonance can be fully maintained (at least longer than when it is followed by a non-high vowel), thus being more similar in perceptual nasality to its corresponding output with n-insertion, i.e. [ɲ].
- Similarity of Input-output pairs:
  j-ɲ/_[+high] > j-ɲ/_[-high]

  \[
  \text{DEP-n/}_j[-\text{high}] > \text{DEP-n/}_j[+\text{high}]
  \]
IV. Learning simulation

- Maxent (Maximum entropy) learner implemented in OTSoft (Hayes, Tesar & Zuraw 2013)
- Input learning/training data: the results of my own survey
- The current survey data have responses only to words with /j/-initial M2, not words with /i/-initial M2. Only the relevant constraints were employed in the simulation. For instance, DEP-n/_i was excluded.
- The frequencies of inserted and non-inserted forms (resyllabified or aligned) are taken from the relevant distribution in the results of the current survey, as shown below.
<table>
<thead>
<tr>
<th>C1 type</th>
<th>Response type</th>
<th>Frequency</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>η</td>
<td>Inserted</td>
<td>715</td>
<td>0.381</td>
</tr>
<tr>
<td></td>
<td>Resyllabified</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>aligned</td>
<td>1160</td>
<td>0.619</td>
</tr>
<tr>
<td>obs</td>
<td>Inserted</td>
<td>538</td>
<td>0.385</td>
</tr>
<tr>
<td></td>
<td>Resyllabified</td>
<td>588</td>
<td>0.421</td>
</tr>
<tr>
<td></td>
<td>aligned</td>
<td>271</td>
<td>0.194</td>
</tr>
<tr>
<td>son</td>
<td>Inserted</td>
<td>1304</td>
<td>0.472</td>
</tr>
<tr>
<td></td>
<td>Resyllabified</td>
<td>926</td>
<td>0.335</td>
</tr>
<tr>
<td></td>
<td>aligned</td>
<td>533</td>
<td>0.193</td>
</tr>
<tr>
<td>C1 type</td>
<td>Response type</td>
<td>Frequency</td>
<td>Proportion</td>
</tr>
<tr>
<td>---------</td>
<td>---------------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>η</td>
<td>Inserted</td>
<td>95</td>
<td>0.424</td>
</tr>
<tr>
<td></td>
<td>Resyllabified</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>aligned</td>
<td>129</td>
<td>0.576</td>
</tr>
<tr>
<td>obs</td>
<td>Inserted</td>
<td>127</td>
<td>0.635</td>
</tr>
<tr>
<td></td>
<td>Resyllabified</td>
<td>27</td>
<td>0.135</td>
</tr>
<tr>
<td></td>
<td>aligned</td>
<td>46</td>
<td>0.23</td>
</tr>
<tr>
<td>son</td>
<td>Inserted</td>
<td>217</td>
<td>0.644</td>
</tr>
<tr>
<td></td>
<td>Resyllabified</td>
<td>31</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>aligned</td>
<td>89</td>
<td>0.264</td>
</tr>
</tbody>
</table>
Additional training data

• 25,110 words with /VC₁+V₂/ (V₂ ≠ i/j)
  ➢ /sikol+ai/ → [si.ko.ra.i] (resyllabified: 63%)
  ➢ [si.kol.a.i] (aligned: 37%)
  ➢ These are added for the purpose of capturing the fact that ALIGN-R and ONSET are in competition in Korean phonology.
  ➢ Otherwise, all optimal outputs would satisfy ONSET.
• Non-inserted forms (from the results of my own survey)
  ➢ Resyllabified 1,572 (63%)
  ➢ Aligned 939 (37%)
  Total 2,511
• I need a large number of words where M2 begins with vocoids other than /j/: 25,110 = 10 x 2,511.
## Constraint weights learned

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Weight</th>
<th>Rule</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>*[ŋ]</td>
<td>50</td>
<td>ID(son)</td>
<td>0.479</td>
</tr>
<tr>
<td>ONSET</td>
<td>2.321</td>
<td>DEP-n/_j[+high]</td>
<td>0.069</td>
</tr>
<tr>
<td>ALIGN-R</td>
<td>1.762</td>
<td>*CODA</td>
<td>0.065</td>
</tr>
<tr>
<td>*V-NUC</td>
<td>1.699</td>
<td>*AMBI</td>
<td>0.0</td>
</tr>
<tr>
<td>DEP-n/_j[-high]</td>
<td>0.708</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The obstruency, velar nasal and height effects in the training data were reproduced.
V. Wug test

• I performed a novel word experiment on n-insertion in Seoul Korean.

• Purpose: to find out whether Seoul Korean speakers are aware of all the trends observed in existing words.
Experimental tokens

- All test words consist of two morphemes, loanword M1 and wug stem M2: e.g. /tʰap + jucenol/ cf. / tʰap / ‘top’.
- M1, which is either mono or di-syllabic, ends with one of seven consonants, /m, n, ŋ, l, p, s, k/.
- M2 begins with one of /i, ju, ja/.
- The total number of test items is 84 (2 (syllable count) x 7 (coda) x 3 (vocoid type) x 2 repeating blocks).
- The same number of control items (vowel-final M1 or /a, e/-initial M2) were adopted.
Method

• Participants: 37 Seoul Korean speakers
• In the test form, both inserted and non-inserted forms for each word were presented in standard Korean orthography.
• The experimenter told the participants that the words in the test form are made-up compound nouns for new chemical products.
• The participants were instructed to choose their pronunciation of each of the given compounds, from three choices, inserted, resyllabified and aligned. (When C1 = [ŋ], no “resyllabified” choice was given.)
Results (wug test)

- Insertion rate by C1 type (# of the relevant responses shown in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>son</th>
<th>obs</th>
<th>η</th>
</tr>
</thead>
<tbody>
<tr>
<td>son</td>
<td>0.28 (197)</td>
<td>0.22 (187)</td>
<td>0.10 (61)</td>
</tr>
</tbody>
</table>

- It seems both obstruency and velar nasal effects are confirmed.
A mixed effect logistic regression model

- The results of the present survey are 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.
- Random intercepts were set for item and subject. Random slopes were set only for subject.
- Fixed factors (underlined=reference level)
  - **C1 type** (son, obs, ƞ)
  - **M2-initial vocoid** (jV: ja, ju, i)
  - **M1 syllable number** (sylnum1)
### Results: A mixed effect logistic regression model (wug test: fixed effects)

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (Intercept) | 0.01263 | 0.63853 | 0.020 | 0.9842 |
| **C1 = η** | **-2.40162** | 0.49805 | -4.822 | 1.42e-06 *** |
| **C1 = obs** | **-0.65958** | 0.29942 | -2.203 | 0.0276 * |
| **V = i** | **-2.97401** | 0.43785 | -6.792 | 1.10e-11 *** |
| **V = ju** | 0.58794 | 0.30093 | 1.954 | 0.0507 . |
| sylnum1 | -0.46375 | 0.28532 | -1.625 | 0.1041 |

- All three main effects, syllabicinity, obstruency and velar nasal: confirmed
- Height effect: (almost) confirmed
- Length effect: **NOT** confirmed
Wug test results (blue)
vs. MaxEnt model prediction (yellow)

• Let us consider how well these wug test results are matched with the predictions of the Maxent model learning simulations.

• Insertion rate by C1 type and M2-initial vocoids

<table>
<thead>
<tr>
<th></th>
<th>j[-high]</th>
<th></th>
<th>j[+high]</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>son</td>
<td>obs</td>
<td>η</td>
<td>son</td>
</tr>
<tr>
<td>Maxent model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prediction</td>
<td>0.515</td>
<td>0.397</td>
<td>0.31</td>
<td>0.668</td>
</tr>
<tr>
<td>Wug test result</td>
<td>0.355</td>
<td>0.271</td>
<td>0.146</td>
<td>0.445</td>
</tr>
</tbody>
</table>
Plot of insertion rate by C1 type and M2-initial vocoids

- Prediction = predictions of Maxent model; Wug = wug test results

• Prediction = predictions of Maxent model; Wug = wug test results
Conclusion

- Most of the trends observed from the existing words were mirrored in Seoul Korean speakers’ production and intuition on novel words.
- Their internalized grammar has been formalized by adopting OT constraints.
- Its learning has been implemented in a maxent learner.
- Some crucial effects were analyzed under P-map.
- Length effect: only in existing words, not novel words.
  - This suggests that speakers do not learn all statistically prominent patterns in existing words.
  - This might mean that the length of M1 is a phonologically unnatural factor for the application of n-insertion (Becker et al. 2011; Hayes & White 2014).
References (selected)


