Frequency effects and regularization in Korean nouns

Yoonjung Kang
University of Toronto Scarborough
kang@utsc.utoronto.ca

1. Introduction

(1) Overview


- In many Western dialects of Korean, noun-final coronal obstruents are optionally realized as [s]. For /tʰ/-final nouns, there is an additional variant with final [cʰ].

- These novel variants are not equally available in all suffixal contexts, often creating a "mixed paradigm".

(2) Proposal

I. Reflection of the frequency distribution in the lexicon (Jun 2007)
II. Phonological reanalysis/overgeneralization
III. Morphological generalization

2. Background

(3) Korean phoneme inventory (cf. H. Kim 1999)

<table>
<thead>
<tr>
<th>p, pʰ, p’</th>
<th>t, tʰ, t’</th>
<th>k, kʰ, k’</th>
</tr>
</thead>
<tbody>
<tr>
<td>c, cʰ, c’</td>
<td>s, s’</td>
<td>h</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>i</th>
<th>i</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>o</td>
<td>(æ)</td>
</tr>
<tr>
<td>a</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(4) Coda neutralization: /tʰ, t, c, cʰ, s, s’/ → [t]_/coda

STEM-DECL.      STEM-and
[katʰ-α]        [kat-k’o]    ‘same’
[pat-α]          [pat-k’o]    ‘to receive’
[c’ocʰ-α]        [c’ot-k’o]   ‘to chase’
[cac-α]          [cat-k’o]    ‘frequent’
[is’-ə]          [it-k’o]     ‘to exist’
[pəs-ə]          [pət-k’o]    ‘to take off’

(5) Affrication: /t, tʰ/ → [c, cʰ]/_/stem(h){i, j} (Derived environments only)

/i/:  /tʰɒk-pat-i/-  [tʰɒk’ɒci]    ‘bib’
      /katʰ-i/-      [kacʰ-i]      ‘together’
/i/:  /pat-ini/-    [patʰini]*[pacʰini] ‘receive-therefore’
      /katʰ-ini/-    [katʰini]*[kacʰini] ‘same-therefore’
/a/:  /pat-α/-      [patα]        *[pacα] ‘receive-IMPERATIVE’
      /katʰ-α/-      [katʰα]        *[kacʰα] ‘same-IMPERATIVE’
/a/:  /k atol-α/-   [kətα]        *[kəcα] ‘gather-IMPERATIVE’
      /putʰ-α/-      [putʰα]        *[pucʰα] ‘stick-IMPERATIVE’

3. Neutralization and variation


- The unsuffixed form is the most “informative” form of the Korean noun paradigm, for the language in general, from which other forms in the paradigm can be projected (Albright 2005).

- Neutralization of underlying contrast in noun paradigm

<table>
<thead>
<tr>
<th>Unsuffixed</th>
<th>NOM. (-i)</th>
<th>ACC. (-il)</th>
<th>DIR. (-ilo)</th>
<th>LOC. (-e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/-s/</td>
<td>-t</td>
<td>-s</td>
<td>-s</td>
<td>-s</td>
</tr>
<tr>
<td>/-c/</td>
<td>-t</td>
<td>-c</td>
<td>-c</td>
<td>-c</td>
</tr>
<tr>
<td>/-cʰ/</td>
<td>-t</td>
<td>-cʰ</td>
<td>-cʰ</td>
<td>-cʰ</td>
</tr>
<tr>
<td>/-tʰ/</td>
<td>-t</td>
<td>-tʰ</td>
<td>-tʰ</td>
<td>-tʰ</td>
</tr>
</tbody>
</table>

:Coda Neutralization
:Affrication Neutralization
• In the unsuffixed form, all coronal obstruents neutralize. Therefore, the learners are presented with the unsuffixed form and have to “guess” what the underlying consonant should be.

```
-\_V
\_V t
-\_V
\_V ts
\_V t
\_V ts
\_V t
-\_V
\_V
```

• They opt for the most common /s/-final nouns, which gives them the best chance at being correct.

```
-\_V
\_V s t
\_V \_V
\_V\_V s
\_V\_V t
\_V\_V t
\_V\_V s
\_V\_V t
```

(7) \([t^h] \sim [ch]\) variation: projecting from the nominative (Kang 2005)

• The nominative form is the second most informative form of the Korean noun paradigm, for the language in general (Albright 2005).

• Given the \([c^h]\)-final form in the nominative, learners cannot be sure which consonant should appear before other vowel-initial suffixes.

```
NOM. (-i) other V-initial suffix
-\_V
\_V ch
\_V \_V
\_V\_V - ch
\_V\_V \_V
```

I. Western dialects (Seoul, Kyenggi, Chunchung, Cenla): the change is mainly in the direction of original /t^h/ \(\rightarrow [ch]\) and the change of original /c^h/ \(\rightarrow [t^h]\) is marginal.

```
NOM. (-i) other V-initial suffix
-\_V
\_V ch
\_V \_V
\_V\_V - ch
\_V\_V \_V
```

3
II. Eastern dialects (Kangwon, Kyengsang): the change is mainly in the direction of original /cʰ/ → [tʰ] and the change of original /tʰ/ → [cʰ] is marginal.


\[
\text{Nom. (-i)} \\
\begin{array}{c}
\text{-cʰ/} \\
\text{tʰ/}
\end{array}
\begin{array}{c}
\text{other V-initial suffix} \\
\text{cʰ/} \\
\text{tʰ/}
\end{array}
\]

III. Hampuk dialect: The dialect does not have affrication rule and therefore, there is no neutralization of /tʰ/ and /cʰ/ nouns in the nominative form. [tʰ] ~ [cʰ] variation is not found in either direction (B. Kim 2005).
4. Distribution of [ch] variants

- Hye-Won Choi, Min-Kyeng Suh, Yen-Sin Hwang, Mi-Yeng Kwen
- 1174 Seoul-Incheon-Keynggi speakers
- Multiple choice questionnaire
- 14 /tʰ/-final nouns
  - pyolh ‘sunlight’, mit h ‘bottom’, pat h ‘field’, k’it h ‘end’, p’at h ‘red bean’,
    solh ‘pot’, kət h ‘outside’ kyolh ‘side’ nat h ‘a unit’, twik’jol h ‘backyard’,
    molimat h ‘bedside’, mut h ‘land’, pak’at h ‘outside’, and sut h ‘thickness
    (of hair)’

(9) Proportion of [tʰ], [ch] and [s] responses for /tʰ/-final nouns in each suffix context
(Based on NAKL 2004)

- The proportion of [s] responses is roughly constant across suffixes.
- The ratio between [tʰ] and [ch] responses differs significantly by the suffix.
5. Lexical diffusion and frequency of use

(10) Certain sound changes affect low- and high-frequency words differently

I. Phonetically motivated changes
• Phonetically motivated changes (typically ones arising from lenition) affect high-frequency words first.
• Phonetically motivated changes progress with each use of the word. Therefore, the more frequent a word is put to use, the more advanced the change is.
  ex. Schwa deletion in English: every vs. mammry

II. Analogically motivated changes
• Analogically motivated changes affect low-frequency words first.
• High-frequency words form strong mental representations and resist change motivated by analogy to other forms.
  ex. Regularization of English irregular past: weeped (< wept) vs. *keeped (<kept)

(11) The frequency of use in coronal obstruent nouns
• Nouns with meaning of location or time—which include most /tʰ/-final nouns—are used in the locative form frequently and the locative form of these nouns tends to resist the changes in noun-final coronals.

<table>
<thead>
<tr>
<th>‘place’</th>
<th>acc.</th>
<th>loc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>15C:</td>
<td>kot-ɔl</td>
<td>kot-ɔy</td>
</tr>
<tr>
<td>19C Cenla Dialect:</td>
<td>koc-ɪl~kos-ɪl</td>
<td>kot-e</td>
</tr>
<tr>
<td>20C:</td>
<td>kos-ɪl</td>
<td>kos-e</td>
</tr>
</tbody>
</table>

  • 8 Seoul speakers; well-formedness rating; Written questionnaire format
  • Rating: between 1 (bad) and 4 (good).
  • Ratings for [s] variant negatively correlate with the frequency of use.\(^1\)
    
    ▪ /c/-final nouns (3 nouns)

\[
\begin{array}{c|c|c|c|c|c|c}
\text{Frequency} & 1 & 10 & 100 & 1000 & 10000 \\
\hline
\text{Wellformedness rating} & \text{nom} & \text{cop} & \text{acc} & \text{dir} & \text{top} & \text{loc} \\
\hline
\end{array}
\]

○ /c/-final nouns (5 nouns)

\[
\begin{array}{c|c|c|c|c|c|c}
\text{Frequency} & 1 & 10 & 100 & 1000 & 10000 \\
\hline
\text{Wellformedness rating} & \text{nom} & \text{cop} & \text{acc} & \text{dir} & \text{top} & \text{loc} \\
\hline
\end{array}
\]

○ /t/-final nouns (8 nouns)

\[
\begin{array}{c|c|c|c|c|c|c}
\text{Frequency} & 1 & 10 & 100 & 1000 & 10000 \\
\hline
\text{Wellformedness rating} & \text{nom} & \text{cop} & \text{acc} & \text{dir} & \text{top} & \text{loc} \\
\hline
\end{array}
\]

\(^1\) Frequency counts from KAIST Concordance program (KCPMSTAT) containing 13.6 million words (http://morph.kaist.ac.kr/kcp/).
The frequency of use and the proportion of [çʰ] responses for /tʰ/-final nouns (Based on NAKL 2004)

- Frequency counts: Google search (June 2007)
- The frequency of use effect does not explain away the suffix effect: words of similar frequency of use show different rates of [çʰ] responses depending on the suffix it contains.
6. Frequency distribution in the lexicon

(14) Jun (2007)


- Innovative [cʰ] variants are more likely to spread to those contexts where /cʰ/-final nouns are already abundantly attested.

(15) Google token frequency and NAKL data²

<table>
<thead>
<tr>
<th></th>
<th>Nom (i)</th>
<th>Acc. (il)</th>
<th>Dir. (ilo)</th>
<th>Loc. (e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google token frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of [cʰ] out of</td>
<td>100%</td>
<td>70.9%</td>
<td>22.6%</td>
<td>6.2%</td>
</tr>
<tr>
<td>[cʰ] and [tʰ] combined in</td>
<td>by rule</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hits for /cʰ/-final nouns</td>
<td>9,263,510</td>
<td>2,059,463</td>
<td>1,313,608</td>
<td></td>
</tr>
<tr>
<td>Hits for /tʰ/-final nouns</td>
<td>3,810,939</td>
<td>7,070,496</td>
<td>19,772,809</td>
<td></td>
</tr>
<tr>
<td>NAKL (2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of [cʰ]-final</td>
<td>96.8%</td>
<td>49.1%</td>
<td>27.2%</td>
<td>5.5%</td>
</tr>
<tr>
<td>responses out of [cʰ] and</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[tʰ] final responses</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>combined</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(16) Change in the other direction also honours the hierarchy of relative frequency distribution of [cʰ] vs. [tʰ] in different suffix context. (Kang 2005)

- Innovative [tʰ] variants are more likely to spread to the suffix context where /tʰ/-final nouns are already abundantly attested.

---
² Jun (2007) used type frequency of nouns in different suffix contexts in 1.5 million word Sejong corpus. However, the inflectional suffixes productively combine with nouns and given a large enough corpus (such as Google), it is not unlikely that the type frequency of [cʰ] vs. [tʰ] nouns come out similar across different suffix contexts. For this reason, I am using token frequency counts, rather than type frequency counts here. In this table, the relative frequency counts are based on a Google search (June 2007) of all monomorphemic /tʰ/ and /cʰ/-final nouns attested in the Sejong corpus: 15 monomorphemic /tʰ/-final nouns and 11 monomorphemic /cʰ/-final nouns (cf. NAKL 2003 survey of word frequency).
• Kang (2003, 2005): Ratings for /cʰ/ nouns pronounced as [tʰ]-final. (Average of 8 speakers)

<table>
<thead>
<tr>
<th>Google token Frequency</th>
<th>0 % (by rule)</th>
<th>29.1 %</th>
<th>77.4 %</th>
<th>93.8 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of [tʰ]</td>
<td>0 %</td>
<td>29.1 %</td>
<td>77.4 %</td>
<td>93.8 %</td>
</tr>
<tr>
<td>Out of [tʰ] and [cʰ]</td>
<td>0 %</td>
<td>29.1 %</td>
<td>77.4 %</td>
<td>93.8 %</td>
</tr>
</tbody>
</table>

• /cʰ/ nouns: Kyenggi dialects (AKS 1995)

<table>
<thead>
<tr>
<th></th>
<th>Nom./i/</th>
<th>Acc. /i/ (~/ol/)</th>
<th>Loc. /e/ (N=19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/sucʰ/ ‘charcoal’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cʰ</td>
<td>cʰ</td>
<td>tʰ</td>
<td>5</td>
</tr>
<tr>
<td>cʰ</td>
<td>cʰ</td>
<td>tʰ</td>
<td>1</td>
</tr>
<tr>
<td>cʰ</td>
<td>cʰ</td>
<td>cʰ</td>
<td>3</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>s</td>
<td>6</td>
</tr>
<tr>
<td>cʰ</td>
<td>s</td>
<td>s</td>
<td>1</td>
</tr>
<tr>
<td>cʰ</td>
<td>s</td>
<td>cʰ</td>
<td>1</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>tʰ</td>
<td>1</td>
</tr>
<tr>
<td>/k’ocʰ/ ‘flower’</td>
<td>(cʰ)</td>
<td>cʰ</td>
<td>tʰ</td>
</tr>
<tr>
<td>------------------</td>
<td>------</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td></td>
<td>cʰ</td>
<td>cʰ</td>
<td>cʰ</td>
</tr>
<tr>
<td></td>
<td>cʰ</td>
<td>cʰ</td>
<td>cʰˌs</td>
</tr>
<tr>
<td>(s)</td>
<td>s</td>
<td>s</td>
<td>tʰ</td>
</tr>
</tbody>
</table>

- /cʰ/-final nouns: Southern Kyengsang dialect (AKS 1993)

<table>
<thead>
<tr>
<th>/sucʰ/ ‘charcoal’</th>
<th>Nom./i/</th>
<th>Acc. /s</th>
<th>Loc. /e/</th>
<th>(N=15)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cʰ</td>
<td>tʰ</td>
<td>tʰ</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>cʰ</td>
<td>cʰ</td>
<td>tʰ</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>cʰ</td>
<td>tʰ</td>
<td>cʰ</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/k’ocʰ/ ‘flower’</th>
<th>Nom./i/</th>
<th>Acc. /s</th>
<th>Loc. /e/</th>
<th>(N=15)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cʰ</td>
<td>tʰ</td>
<td>tʰ</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>cʰ</td>
<td>cʰ</td>
<td>tʰ</td>
<td>5</td>
</tr>
</tbody>
</table>

- /cʰ/-final nouns: Kangwon dialect (AKS 1990)

<table>
<thead>
<tr>
<th>/sucʰ/ ‘charcoal’</th>
<th>Nom./i/</th>
<th>Acc. /il</th>
<th>Loc. /e/</th>
<th>(N=15)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cʰ</td>
<td>tʰ</td>
<td>tʰ</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>cʰ</td>
<td>cʰ</td>
<td>cʰ</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>cʰ</td>
<td>cʰ</td>
<td>tʰ</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>cʰ</td>
<td>tʰ</td>
<td>cʰ</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/k’ocʰ/ ‘flower’</th>
<th>Nom./i/</th>
<th>Acc. /il</th>
<th>Loc. /e/</th>
<th>(N=15)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cʰ</td>
<td>tʰ</td>
<td>tʰ</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>cʰ</td>
<td>cʰ</td>
<td>tʰ</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>cʰ</td>
<td>cʰ</td>
<td>cʰ</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>cʰ</td>
<td>cʰ</td>
<td>tʰ</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>cʰˌs</td>
<td>tʰ</td>
<td>s</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/tœcʰ/ ‘trap’</th>
<th>(cʰ)</th>
<th>(cʰ)</th>
<th>(cʰ)</th>
<th>(N=15)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cʰ</td>
<td>cʰ</td>
<td>cʰ</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>cʰ</td>
<td>cʰ</td>
<td>cʰ</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>cʰ</td>
<td>cʰ</td>
<td>tʰ</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>cʰ</td>
<td>tʰ</td>
<td>tʰ</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>cʰ</td>
<td>tʰ</td>
<td>tʰ</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>1</td>
</tr>
</tbody>
</table>
7. Wellformedness rating


- Wellformedness rating for [tʰ] pronunciation of /tʰ/-final nouns
- Mirrors the frequency distribution of [tʰ].
• Wellformedness rating for [cʰ] pronunciation of /tʰ/-final nouns
• Some speakers seem to have reanalyzed the suffixal asymmetry as one conditioned by the vowel height.

Speaker 6 does not accept innovative [cʰ] variants at all.
• Speakers 5 and 7 show the suffix asymmetry in line with the lexical frequency.
• Four speakers (1, 2, 3 and 4) rate [cʰ] variants in all high vowel contexts (Nom, Acc, Dir) comparably good.
• These speakers seem to have reanalyzed the distribution of [cʰ] as conditioned by the vowel height and extended the context of affrication rule to all high vowel contexts.
7. Morphonological generalization

(18) Acc.-Dir. Split in Southern Cenla dialects

- 32 speakers each

- Kwangewu

- Kwangyang

- Dir. (ilo) patterns with Loc. (e) and shows a very low rate of [cʰ] responses, while Acc (il) shows a rate of [cʰ] response comparable to that of Nom. (i).
(19) Why no phonological regularization in these dialects?

- Variation in suffixal vowels in many Western dialects (including Cenla) (J. J. Choi 2001, K. Lee 2000, Han and Kim 200x)
  - The locative suffix (-e) is variably realized as /i/ (< ij 19th Century).
  - But, ij > i change occurred after the affrication (t, tʰ > c, cʰ/__ i) ceases to be productive and [tʰ] fails to affricate in the locative.
    - [mitʰ-i] ‘bottom, loc.’
    - [patʰ-ida] ‘field, loc.’
  - Affrication does not reliably apply before any vowel context. And the learners do not seek a phonological generalization for the distribution of [cʰ]. Rather, they resort to the morphological contexts for generalization.

(20) Implications

- Sound changes can be sensitive to the frequency distribution of the lexicon.
- Faced with a complex pattern of variation, learners/speakers seem to seek a simpler generalization (phonological or morphological).
- Phonology remains manageable.

Selected References


Albright, A. 2005. Explaining universal tendencies and language particulars in analogical change. Manuscript, MIT.
최 전승 (Choi, J.) 1986. 19세기 후기 전라방언의 음운현상과 그 역사상. [Phonology of late 18th century Cenla dialect]. 전북대학교 박사학위 논문.


한영목, 김래영 (Han and Kim) 200x. 충남방언 조사의 형태변이와 통사 의미의 몇 양상 [On Inflectional suffixes in Chungnam dialects]. 인문학 연구 30. 2.


