Onset voicing and tonal distribution in the Taiwanese lexicon

Sheng-Fu Wang (王聖富)

New York University

NACCL-29
June 18, 2017
Overview

- Background: onset voicing and tonogenesis
- Research question: relationship between phonetic motivations and synchronic lexicon?
- Method: lexical statistics from dictionary
- Results and Discussion: a mixture of diachronic and synchronous phonetic effects
Onset voicing and Tonogenesis

- Voiced/voiceless stops cause F0 perturbations in vowel onset (e.g., Ohde, 1984)
  - voiced → low F0
  - voiceless → high F0

- Tonogenesis: when such F0 differences are reanalyzed as distinctive phonological feature (e.g., Hombert et al., 1979)
Onset voicing and Tonogenesis in Min

Min: Tonogenesis and voicing change (e.g., Norman, 1973, 1974; Handel, 2003)

- *p-, *pʰ- → p-, pʰ- with high tones
- *b- → p-, pʰ- with low tones
- *m-, *mʰ- → m-/b- with low tones

Research question: does the current tone-onset association reflect the phonetics of tonogenesis?

- high tones (44, 52, 44q) → more likely to occur with voiceless onsets (or null effect)?
- low tones (24, 11, 32, 21q) → more likely to occur with voiced onsets?
The twist: flip of tonal register

- The low-level tone (11) had high tonal register (yin-qu)
- The checked tones (44q, 21q) used to have the opposite tonal registers
- Are they more likely to occur with voiceless or voiced onsets?
  - If there’s an association, is it more likely to follow the tonal value before or after the flip?

<table>
<thead>
<tr>
<th>diachronic class &amp; register</th>
<th>tonal value in TW</th>
</tr>
</thead>
<tbody>
<tr>
<td>*A, high (yin-ping)</td>
<td>44</td>
</tr>
<tr>
<td>*A, low (yang-ping)</td>
<td>24</td>
</tr>
<tr>
<td>*B, high (yin-shang)</td>
<td>42</td>
</tr>
<tr>
<td>*B, low (yang-shang)</td>
<td>∅</td>
</tr>
<tr>
<td>*C, high (yin-qu)</td>
<td>11</td>
</tr>
<tr>
<td>*C, low (yang-qu)</td>
<td>32</td>
</tr>
<tr>
<td>*D, high (yin-ru)</td>
<td>21q</td>
</tr>
<tr>
<td>*D, low (yang-ru)</td>
<td>44q</td>
</tr>
</tbody>
</table>
Why looking at this? Synchronic vs. diachronic explanation

- What is the nature of the association (if any) between onset voicing and tonal register in the synchronic grammar?
  - results of ‘blind’ language change that are phonetically motivated (e.g., Blevins, 2004; Mielke, 2008)?
    - Laryngeal-tone systems are independent of the tonal development (other than remnants of sound change)
    - It is possible for low tones to associate more with voiceless stops
  - somehow encoded in Universal Grammar that constrains language change (see de Lacy and Kingston, 2013)?
    - Constraints on laryngeal-tone systems should affect how languages change/develop
    - High tones should be associated more with voiceless stops in the contemporary lexicon no matter what
Potential Results on tone-onset voicing association

- The question: $p \leftrightarrow H$ & $b/m \leftrightarrow L$?
- For tones that have not undergone flip of tonal register:
  - Yes: Phonetic explanation makes sense both diachronically and synchronically
  - No: Motivation for sound change is not reflected in synchronic patterns (or anywhere at all?)
Potential Results on tone-onset voicing association (cont.)

- The question: p ↔ H & b/m ↔ L?
- For tones that have undergone flip of tonal register:
  - consistent with diachronic tonal register:
    - The onset-tone distribution has not changed because of tonal change
    - No phonetic effects on the likelihood of tone-onset combination in syllable types developed/introduced later
  - consistent with synchronic tonal register:
    - The distribution has changed because of tonal change?
    - after the tonal flip, phonetic effects on the likelihood of tone-onset combination in newly introduced syllable types
Method: Dictionary

- The Taiwan Southern Min online dictionary by the Ministry of Education
  - the version open-sourced by the MoeDict (萌典) team of g0v.tw (零時政府)
- Dictionary processing:
  - Tallying attested tones for each syllable type
  - Mapping syllable-level phonological features to syllable types
    - e.g., Onset-voicing (+, −, n/a), vowel nasality (+, −, n/a), vowel length (mono, diph, triph)
Method: Co-occurrence statistics

- Co-occurrence statistics: Observed over Expected
- Testing whether the distribution of onset-voicing and attestedness of tones is independent
  - Cross-table and Chi-square tests
  - Attested vs. unattested
  - Past studies: OCP (Coetzee and Pater, 2008; Frisch and Zawaydeh, 2001), palatalization and derived environment effect (Chong, 2016)

- One cross-table for each tone
  - e.g., Is the occurrence of onset voicing independent from the attestedness of High-Level tone across syllable types?

<table>
<thead>
<tr>
<th></th>
<th>with tone X</th>
<th>without tone X</th>
</tr>
</thead>
<tbody>
<tr>
<td>voiced onset</td>
<td></td>
<td></td>
</tr>
<tr>
<td>voiceless onset</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


### Results: example of one analysis (high-level tone)

\[ \chi^2(1, N = 151) = 71.60, \ p < .00001 \]

→ voiceless onsets are more likely to be associated with high-level tone

<table>
<thead>
<tr>
<th></th>
<th>with high-level tone</th>
<th>without high-level tone</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>voiced onset</td>
<td>6 (30.18)</td>
<td>43 (18.82)</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>3.97% / 19.98 %</td>
<td>28.48% / 12.46%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>O/E = 0.20</td>
<td>O/E = 2.28</td>
<td></td>
</tr>
<tr>
<td>voiceless onset</td>
<td>87 (62.82)</td>
<td>15 (39.18)</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>57.61% / 41.60%</td>
<td>9.93% / 25.95%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>O/E = 1.38</td>
<td>O/E = 0.38</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>93</td>
<td>58</td>
<td>151</td>
</tr>
</tbody>
</table>
## Results: Overview

<table>
<thead>
<tr>
<th>diachronic class &amp; register</th>
<th>tonal value in TW</th>
<th>Result (positive association with..)</th>
<th>consistent with…</th>
</tr>
</thead>
<tbody>
<tr>
<td>*A, high</td>
<td>44</td>
<td>voiceless onsets</td>
<td>diachronic/synchronic</td>
</tr>
<tr>
<td>*A, low</td>
<td>24</td>
<td>voiced onsets</td>
<td>diachronic/synchronic</td>
</tr>
<tr>
<td>*B, high</td>
<td>42</td>
<td>voiceless stops</td>
<td>diachronic/synchronic</td>
</tr>
<tr>
<td>*B, low</td>
<td>∅</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*C, high</td>
<td>11</td>
<td>voiced onsets</td>
<td>synchronous</td>
</tr>
<tr>
<td>*C, low</td>
<td>32</td>
<td>voiceless onsets</td>
<td>neither</td>
</tr>
<tr>
<td>*D, high</td>
<td>21q</td>
<td>voiceless onsets</td>
<td>diachronic</td>
</tr>
<tr>
<td>*D, low</td>
<td>44q</td>
<td>voiced onsets</td>
<td>diachronic</td>
</tr>
</tbody>
</table>
Results: summary

- ping-sheng and shang-sheng (44, 24, 42): onset-tone associations are consistent with diachronic and synchronic phonetic contexts
- ru-sheng (21q, 44q): onset-tone associations follow the tonal value before tonal flip
  - Following diachronic phonetic contexts
- qu-sheng (11, 32): not consistent with diachronic phonetic contexts
  - 11: associated with voiced onsets (synchronous)
  - 32: associated with voiceless onsets (neither synchronous nor diachronic)
Back to our potential results

- The question: p ↔ H & b/m ↔ L?
- For tones that have not undergone flip of tonal register:
  - Yes (44, 42, 24): Phonetic explanation makes sense both diachronically and synchronically
  - No (32): Motivation for sound change is not reflected in synchronic patterns (or anywhere at all?)
The question: $p \leftrightarrow H & b/m \leftrightarrow L$?

For tones that have undergone register flip:
- consistent with diachronic flip of tonal register $(44q, 21q)$:
  - The onset-tone distribution has not changed because of tonal change
  - No phonetic effects on the likelihood of tone-onset combination in syllable types developed/introduced later
- consistent with synchronic tonal register $(11)$:
  - The distribution has changed because of tonal change?
  - after the tonal flip, phonetic effects on the likelihood of tone-onset combination in newly introduced syllable types
Discussion

- The results do not support the strong version of either hypothesis
  - Although, ignoring the low-level and mid-falling tones (qu-sheng tones), the associations follow the original phonetic context
    - The checked tones follow early phonetic contexts rather than current contexts
    - It supports the Blevins/Mielke view of synchronic grammar

- Something special about qusheng tones?
  - a (more) synchronic effect - dispersion?
    - The low-level (11) tone associated with voiced onset
    - The mid-falling tone (32) associated with voiceless onset
  - Something about coda
Next steps

- Type and token frequencies in lexical statistics
- Coda effect
- Psychological reality of such biases
- Lexical strata
- More varieties of Min
Take-home messages

- Lexical statistics and contingency patterns are interesting (albeit limited) tools
- The patterning of onset-voicing and tones may be able to contribute to the wider questions of synchronic and diachronic ‘natural’ explanations
Thank you!

I thank Gillian Gallagher and Scott Seyfarth for their feedback on this project.