The Production and Perception of Foreign Language Sequences
Adelphi University, March 2011
Ellen Broselow
Jiwon Hwang
Stony Brook University (SUNY)

Question
In language contact/foreign language learning, speakers are exposed to structures that do not exist in their native language.

Some of these structures are more readily mastered than other (equally novel) structures.

How can we explain this?

Proposal
1. Differential difficulty in production may stem from differential difficulty in perception.

2. Certain new structures may be perceived less accurately than other (equally new) structures, either because of
   --universal differences in perceptual salience, or
   --native language restrictions.

Outline
1. Perceptual salience independent of native language:
   1.1. Jamaican Creole
   1.2. Japanese loanwords from English

2. Perception determined by native language restrictions:
   Korean L2

1.1. Jamaican Creole

‘that stick’ Basilect: dat tik
          Mesolect: dat stik
          Acrolect: ðat stik

---------------------------------------------------------------
Not found: *ðat tik
---------------------------------------------------------------

Differential difficulty
Meade 1998: any speaker who can produce [ð] can also produce an [st] onset, but not vice versa.

>>[stV]-[tV] contrast more readily produced than [ð]-[d] contrast.
**/d-/ð/ contrast is subtle**

Polka et al. 2001: major acoustic difference is F2

<table>
<thead>
<tr>
<th></th>
<th>[d]</th>
<th>[ð]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean F2 at onset (Hz)</td>
<td>1746</td>
<td>1536</td>
</tr>
<tr>
<td>Mean F2 change (Hz)</td>
<td>604</td>
<td>456</td>
</tr>
<tr>
<td>Mean noise duration (ms)</td>
<td>16.2</td>
<td>18.6</td>
</tr>
<tr>
<td>Mean noise amplitude (dB)</td>
<td>53.7</td>
<td>56.2</td>
</tr>
</tbody>
</table>

**[d]-[ð] contrast is hard for children to hear**

Polka et al. 2001:
- Infants, 6-8 months and 10-12 months
- Francophone and Anglophone
- All 4 groups showed relatively poor [d]-[ð] discrimination relative to other contrasts.

**[d]-[ð] contrast is hard for L2 learners to hear**

Moroson and Jamieson 1989:
After training, adult Canadian Francophones successfully distinguished [θ]-[ð], but still could not distinguish [d]-[ð].

We can explain the pattern *dat tick* for ‘that stick’ by reference to inherent perceptual salience:

[d]-[ð] contrast (*dat-that*) has weak acoustic cues (& low functional load in English).
[s]-[st] contrast has strong cues ([s] is salient).

---

**1. 2. Japanese Loans**

‘city’
- Fully assimilated: [ʃiʃi]
- Partially assimilated: [ʃiʃi]
- Unassimilated: [siti]

Not found: *[ʃiʃi]*

---

**Differential Difficulty**

Crawford 2007: borrowed words with [ti] begin to show up in Japanese in the 1930s (but very few have [si] even now).

>>[ti]-[ʃi] contrast more readily produced than [si]-[ʃi] contrast.
### Native Language Effect

The sequences [ti] and [si] do not occur in native Japanese vocabulary:

- \(kas + u\) ‘lend’ \(kaʃi+ita\) ‘lent’ (si > fi)
- \(kat + u\) ‘win’ \(kaʃi+ita\) ‘won’ (ti > fi)

So the native language equally impedes mastery of BOTH structures.

### Both structures may be assimilated in loanwords

<table>
<thead>
<tr>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘team’</td>
<td>[ʃiimu]</td>
</tr>
<tr>
<td>‘seafood’</td>
<td>[ʃifuudo]</td>
</tr>
<tr>
<td>‘Citizen’</td>
<td>[ʃiizun]</td>
</tr>
</tbody>
</table>

### But

Less assimilated loans may have [ti]:

- ‘teenager’ [tiin]
- ‘party’ [paati]

But not [si]:

- ‘Citibank’ [ʃiibiŋku]
- ‘CD’ [ʃiidi]

(Itô & Mester 1995, 1999; Crawford 2007)

### Difference in perceptual salience?

- [ti]-[ʃi] contrast: different in
  - place (alveolar/palatal)
  - manner (stop/affricate)

- [si]-[ʃi] contrast: different in
  - place (alveolar/palatal)

### English vs. Japanese s/sh

- Main cues for English /ʃ/ | /tʃ/: rise-times and fricative noise duration (Dorman et al. 1980)
- Main cues for English /s/ | /ʃ/: centroid frequency
- Main cues for Japanese /s/ | /ʃ/: onset F2 frequency (Li, Edwards, & Beckman 2007, 2009)

### Question

Does the difference in perceptual salience of these 2 contrasts hold regardless of native language?
Experiment I

Subjects: 14 Japanese NSs (in Japan)
25 English NSs (in U.S.)

Stimuli: nonwords (LHH) produced by balanced bilingual Japanese-English speaker

Task: ABX discrimination
  e.g., hjatire-hjatire
  fasire-fasire

Context contained cues marking words as foreign or native

- [hja] only in native-origin words.
- [fa] only in foreign-origin words.

<table>
<thead>
<tr>
<th>Contrast possible?</th>
<th>Native-origin words</th>
<th>Foreign-origin words</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ti]-[?]</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>[si]-[ʃ]</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

These results suggest that the more easily produced contrast is (inherently) more easily perceived.

Puzzle: The better performance on ti/chi vs. si/shi in English speakers came mainly from the participants who self-identified as African-American—why?

Interim Summary

- Listeners were exposed to 2 structures that did not occur in their native language.
- One structure was produced correctly, creating a new contrast (t-st, ti-chi).
- One structure was assimilated to a native language category (that > dat, seafood > sheafodd).
- We argued that the second structure was inherently more difficult to perceive.

2. Native Language in Perception:
Korean stop-nasal sequences

- In Korean, a stop becomes a nasal before a nasal consonant:
  e.g., /kukmul/ → [kuŋmul] 'soup'

- English stop-nasal sequences as in 'hotmail' or 'segment' pose a problem for Korean speakers.
Production I: Reading

- 20 Korean speakers (Stony Brook) read English pseudowords embedded in sentences.
- Pseudowords were bisyllabic, containing sequence of stop-nasal (e.g., segmal, tibnan, tipnal).
- Productions were transcribed by English native speakers.

Findings

- The Koreans did NOT simply transfer their native language process of nasalizing a stop before a nasal.
- Instead, the most frequent production error involved inserting a vowel between the stop and the nasal.
- However, correct production and vowel insertion were dependent on context.

Voicing Asymmetry

- Korean speakers inserted a vowel much more frequently after a voiced consonant (tegnal) than after a voiceless consonant (teknal).
- This is puzzling, since Korean does not use a voicing contrast.
Production II: Listening+Repetition

- Participants: 20 Korean speakers (Stony Brook)
- Task:

  - Materials: bisyllabic CVC.NV words containing stop and nasal (e.g., segmal, tibnan, tipnal)
- Transcription by English native speakers

Results

Voicing Asymmetry

- Correct
- V insertion

Voicing Asymmetry

- Correct
- V insertion

Voicing effect in both reading and repetition tasks

- Vowel insertion more often applied in voiced context (tegnal → teginal).

- Could this be a perception effect (i.e., the Koreans actually heard a vowel in the English production)?

Illusory vowel in perception

Illusory vowel in perception


Native Language Effect

In Korean,
- Voicing is not contrastive.
- Stops are predictably voiced when they occur between vowels: /kuk + i/ ‘soup’ → [kugi] ‘soup, nom.’
- Therefore, voiced stops occur ONLY before a vowel in Korean.

Hypothesis

- Korean listeners assume that a voiced stop must be followed by a vowel.
- Therefore, Korean listeners will have greater difficulty distinguishing [gn]-[gVn] than [kn]-[kVn].

Perception Experiment

- Categorization and Discrimination for CN-CɪN continua
- 20 Korean and 20 English speakers.
- Stimuli: A modified continuum from /ikna/-/ikɪna/ and from /ɪɡna/-/ɪɡɪnə/.

Categorization Results-voicing

<table>
<thead>
<tr>
<th></th>
<th>kn-kɪN</th>
<th>gN-gɪN</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 ms</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discrimination Results - voicing

<table>
<thead>
<tr>
<th></th>
<th>kn-kɪN</th>
<th>gN-gɪN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

- Voicing effect originates in perception.
  Voiceless unaspirated stops are voiced between sonorants in Korean.

→ The Korean listeners interpreted stop voicing as an indication of a following vowel.

Level of misperception

- Does Korean listeners’ misperception of voiced stop-nasal sequences result from

  (1) the failure to hear the acoustic differences between voiced stop-nasal sequences and voiced stop-V-nasal sequences?

  OR

  (2) the failure to realize that the differences are relevant to linguistic categorization?

Mismatch Negativity

Indicates brain response to change in an auditory stimulus. MMN is elicited even in the absence of attention to stimulus (e.g., while watching a silent movie).

Erp I

Stimuli in ERP

English

Korean

Standard
Deviant

Voiceless unaspirated stops are voiced between sonorants in Korean.

The Korean listeners interpreted stop voicing as an indication of a following vowel.

Mismatch Negativity

Indicates brain response to change in an auditory stimulus. MMN is elicited even in the absence of attention to stimulus (e.g., while watching a silent movie).

Stimuli in ERP

English

Korean

Standard
Deviant
Identification Results-voicing

Categorization-replicated

Summary: Korean

• Voicing Effect originates in misperception.
• Misperception = miscategorization (at least in no V vs. V)
• Miscategorization results from the native language grammar, which allows voiced consonants only before vowels.

Conclusion

• Not all new structures are equally difficult (contrary to the Contrastive Analysis Hypothesis).
• Differential difficulty may result not only from articulatory difficulty (e.g., complexity of English [r]) but also from perceptual difficulty (less salient acoustic cues).
Perception-Production Relationship

- We argued that failure to acquire a new contrast resulted from a failure to perceive that contrast. E.g., a listener who hears English sea and she as the same will not set up separate phonological categories.

Perception must precede production?

Goto 1971, Sheldon & Strange 1982: Japanese speakers could produce acceptable English [r] and [l], even though they could not discriminate (even their own productions).

But this does not necessarily demonstrate that the Japanese speakers had established separate phonological categories for [r] and [l].

Conclusion

Accurate perception of contrasts is a prerequisite to establishment of separate phonological categories.

Foreign contrasts are particularly difficult to perceive when
- their acoustic cues are not particularly salient ([d] vs. [ð], [s] vs. [ʃ]), or
- the cues are subject to misinterpretation in terms of native language structure (voicing signals a following vowel).

Thank you!

This work was supported by NSF grant BCS-07460227 to Ellen Broselow, Marie Huffman, and Nancy Squires.

Thanks to Collaborators: Yu-An Lu, Sophia Kao, Miran Kim, Jennifer Park, Jessica Fareri, Yukiko Asano, Hisako Takahashi, Sara Catlin, Nancy Squires, Marie Huffman.

Selected references


