Dynamic Shift of Word Frequency Effect in the Course of Linguistic Change

Shin-ichiro Sano

Department of Linguistic Sciences
International Christian University

N WAV-Asia Pacific 2,
August 2, 2012 @ NINJAL
Outline

1 Background
   - Neogrammarian Sound Change vs. Lexical Diffusion
   - Word Frequency and Lexical Diffusion
   - Problems and Goals
2 Ongoing Linguistic Change
   - Japanese Morphophonology
   - ra-Deletion
3 Data
   - Corpus
   - Summary of the Data
4 Examination of Word Frequency Effect
   - Method
   - Results
   - Discussion
Outline

1 Background
   - Neogrammarian Sound Change vs. Lexical Diffusion
     - Word Frequency and Lexical Diffusion
     - Problems and Goals

2 Ongoing Linguistic Change
   - Japanese Morphophonology
     - ra-Deletion

3 Data
   - Corpus
   - Summary of the Data

4 Examination of Word Frequency Effect
   - Method
   - Results
   - Discussion
Neogrammrian Regular Sound Change vs. Lexical Diffusion

Neogrammrian Regularity Principle:
- sound change – regular, exceptionless, abrupt
  - **target:** all relevant sounds in every lexical item
  - **manner of change:** all at once, no gap wrt. degree of progress
- e.g. Grimm’s Law, Great Vowel Shift

BUT there are exceptions (Wang 1969, 1977; Wang and Cheng 1977)
- e.g. Latin → Standard French: \( k > f \) palatalization

Lexical Diffusion:
- progress of some changes differs according to lexical items
  ⇒ **irregularity**
Outline

1. Background
   - Neogrammarian Sound Change vs. Lexical Diffusion
   - Word Frequency and Lexical Diffusion
   - Problems and Goals

2. Ongoing Linguistic Change
   - Japanese Morphophonology
   - \textit{ra}-Deletion

3. Data
   - Corpus
   - Summary of the Data

4. Examination of Word Frequency Effect
   - Method
   - Results
   - Discussion
Word Frequency and Lexical Diffusion

Accounts of lexical diffusion:

- regular change – [-social factors]
- word frequency (Hooper 1976; Bybee 2002 et seq.; Phillips 2006)

Word frequency $\Rightarrow$ progress of change
changes diffuse from high-frequency words to low-frequency words

- high-frequency words – susceptible
- low-frequency words – resistant

high-frequency words – more likely to be
locus of change / target of rule application

e.g. English t/d-deletion (Bybee 2002)
Outline

1 Background
   - Neogrammarian Sound Change vs. Lexical Diffusion
   - Word Frequency and Lexical Diffusion
   - Problems and Goals

2 Ongoing Linguistic Change
   - Japanese Morphophonology
   - *ra*-Deletion

3 Data
   - Corpus
   - Summary of the Data

4 Examination of Word Frequency Effect
   - Method
   - Results
   - Discussion
Problems and Goals

Problems
1. But dynamic aspects – underresearched
2. Recent work (phonology) on word frequency effects – synchronic, not diachronic

Goal: demonstrate...
1. dynamic aspects of word frequency effect in ongoing change
2. the mechanism of lexical diffusion
Outline

1. Background
   - Neogrammarian Sound Change vs. Lexical Diffusion
   - Word Frequency and Lexical Diffusion
   - Problems and Goals

2. Ongoing Linguistic Change
   - Japanese Morphophonology
   - ra-Deletion

3. Data
   - Corpus
   - Summary of the Data

4. Examination of Word Frequency Effect
   - Method
   - Results
   - Discussion

Shift of Word Frequency Effect
Verbs in Japanese – 2 types

- **consonant verbs**: stem-final segment – consonant
  - e.g. *hasir-* ‘run,’ *yar-* ‘do’

- **vowel verbs**: stem-final segment – vowel
  - e.g. *mi-* ‘see,’ *tabe-* ‘eat’

Potential forms in Japanese – allomorphy

- **consonant verbs**: potential suffix – e
  - e.g. *hasir*-e– ‘can run,’ *yar*-e– ‘can do’

- **vowel verbs**: potential suffix – rare (traditional variant)
  - e.g. *mi*-rare– ‘can see,’ *tabe*-rare– ‘can eat’

**ra-Deletion** – only in vowel verbs
Outline

1. Background
   - Neogrammarian Sound Change vs. Lexical Diffusion
   - Word Frequency and Lexical Diffusion
   - Problems and Goals

2. Ongoing Linguistic Change
   - Japanese Morphophonology
   - *ra*-Deletion

3. Data
   - Corpus
   - Summary of the Data

4. Examination of Word Frequency Effect
   - Method
   - Results
   - Discussion

Shin-ichiro Sano
Shift of Word Frequency Effect
What is *ra*-Deletion?

1. *ra*-Deletion: change in potential forms in Japanese

2. formation of potential forms (vowel verbs): attaching potential suffix (*rare*) to verb stems

### Configuration of potential forms

<table>
<thead>
<tr>
<th>Traditional potential (traditional variant):</th>
<th>V - rare</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>ra</em>-Deletion (innovative variant):</td>
<td>V - rare</td>
</tr>
</tbody>
</table>

Examples: *tabe-re-ru* ‘can eat,’ *ko-re-ru* ‘can come,’ renewal in single suffix (e.g. Matsuda 1993; Ito and Mester 2004)
ra-Deletion

Background
Ongoing Linguistic Change
Data
Examination of Word Frequency Effect
Conclusion

Japanese Morphophonology
ra-Deletion

ra-Deletion

CSJ

(1) oisii mono-ga **tabe-re-ru**.
delicious stuff-NOM eat-POT-NONPAST
‘(We) can eat delicious foods.’
(cf. traditional: **tabe-rare**)  

(2) onazi keekoo-ga **mi-re-masu**.
same tendency-NOM see-POT-POLITE.NONPAST
‘(We) can observe the same tendency.’
(cf. traditional: **mi-rare**)
Outline

1 Background
   - Neogrammarian Sound Change vs. Lexical Diffusion
   - Word Frequency and Lexical Diffusion
   - Problems and Goals

2 Ongoing Linguistic Change
   - Japanese Morphophonology
   - ra-Deletion

3 Data
   - Corpus
   - Summary of the Data

4 Examination of Word Frequency Effect
   - Method
   - Results
   - Discussion
Corpus of Spontaneous Japanese (CSJ)

- **Size**: 3,302 speech samples (662 hours, 7.5 million words)
- **Organization**: APS (careful) / SPS (relaxed)
- **Rich annotations**: speaker attributes, characteristics of each speech (e.g. style, nervousness, spontaneity) ⇒ detailed analysis of external factors (e.g. style, gender, socioeconomic background)
- **Sampling**: every utterance in CSJ ⇒ all potential forms (vowel verbs)
Outline

1. **Background**
   - Neogrammarian Sound Change vs. Lexical Diffusion
   - Word Frequency and Lexical Diffusion
   - Problems and Goals

2. **Ongoing Linguistic Change**
   - Japanese Morphophonology
   - *ra*-Deletion

3. **Data**
   - Corpus
   - **Summary of the Data**

4. **Examination of Word Frequency Effect**
   - Method
   - Results
   - Discussion
# Overall Distribution

- **# of speakers:** 1,286
- **Birth-year:** 1910s – 1980s

## Distribution of potential forms in CSJ

<table>
<thead>
<tr>
<th>Type</th>
<th>Token</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>traditional potential:</td>
<td>7,615</td>
<td>157</td>
</tr>
<tr>
<td><em>ra</em>-Deletion:</td>
<td>543</td>
<td>53</td>
</tr>
</tbody>
</table>

**probability of *ra*-Deletion:** 6.66%
Figure 1. Chronological transition of potential forms in CSJ

Change of *ra*-Deletion
- Beginning: around 1920s (consistent with previous works)
- in progress, intermediate stage

R² = 0.6012
Outline

1. Background
   - Neogrammarian Sound Change vs. Lexical Diffusion
   - Word Frequency and Lexical Diffusion
   - Problems and Goals

2. Ongoing Linguistic Change
   - Japanese Morphophonology
   - ra-Deletion

3. Data
   - Corpus
   - Summary of the Data

4. Examination of Word Frequency Effect
   - Method
   - Results
   - Discussion
Method

Procedure

1. For each verb stem, calculate the probability of *ra*-Deletion.
2. Assign the frequency of verb stem \(\Leftrightarrow\) NIJL (2005)
   - NIJL (2005) – lists 48,000 lexical items with properties
     e.g. frequency, grammatical information
3. Classify tokens into 4 birth-year periods:
   1910s-20s, 1930s-1940s, 1950s-60s, 1970s-80s
   *Assuming the apparent-time*
4. Calculate the transition of correlation between probability
   of *ra*-Deletion and frequency of verb stem
Method

**Target:** every potential form (vowel verbs) in CSJ
*except for i-ru ‘be’ – outlier:
frequency in NIJL (2005): 8,642 (mean: 365)

**Example** (entire period)

<table>
<thead>
<tr>
<th>verb</th>
<th>prob. of ra-Deletion in CSJ (%)</th>
<th>freq. in NIJL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ku-ru ‘come’</td>
<td>76.09</td>
<td>1,845</td>
</tr>
<tr>
<td>ne-ru ‘sleep’</td>
<td>60.00</td>
<td>53</td>
</tr>
<tr>
<td>de-ru ‘sleep’</td>
<td>45.56</td>
<td>613</td>
</tr>
</tbody>
</table>

**Test for significance of correlation:**
Kendall’s rank correlation tau
Outline

1. Background
   - Neogrammarian Sound Change vs. Lexical Diffusion
   - Word Frequency and Lexical Diffusion
   - Problems and Goals

2. Ongoing Linguistic Change
   - Japanese Morphophonology
   - ra-Deletion

3. Data
   - Corpus
   - Summary of the Data

4. Examination of Word Frequency Effect
   - Method
   - Results
   - Discussion
Transition of Correlation

Figure 2. Correlation between prob. of \( ra \)-Deletion and frequency of verb stem
# Correlation—Summary

## Correlation

Verb stem – **more frequent** ⇒ Prob. of *ra*-Deletion – **higher**

## Degree of Correlation (Slope)

<table>
<thead>
<tr>
<th>Period</th>
<th>Slope Equation</th>
<th>Correlation Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1910s-20s</td>
<td>$y = 0.0044x - 0.854$</td>
<td>(n.s.)</td>
</tr>
<tr>
<td>1930s-40s</td>
<td>$y = 0.0135x - 6.1776$</td>
<td>(p&lt;0.05)</td>
</tr>
<tr>
<td>1950s-60s</td>
<td>$y = 0.0157x - 6.9389$</td>
<td>(p&lt;0.002)</td>
</tr>
<tr>
<td>1970s-80s</td>
<td>$y = 0.0289x - 9.3121$</td>
<td>(p&lt;0.0005)</td>
</tr>
</tbody>
</table>
Outline

1. Background
   - Neogrammarian Sound Change vs. Lexical Diffusion
   - Word Frequency and Lexical Diffusion
   - Problems and Goals

2. Ongoing Linguistic Change
   - Japanese Morphophonology
   - ra-Deletion

3. Data
   - Corpus
   - Summary of the Data

4. Examination of Word Frequency Effect
   - Method
   - Results
   - Discussion
Dynamic Word Frequency Effect and Lexical Diffusion

Consistent with Previous Works

1. Progress of change and word frequency – positively correlate
2. Words with high frequency – preferred context for change

**ra-Deletion:**

- **Outset of change:** restricted to [-frequent] verbs
  
  ↓

- **Change progresses:** more likely to occur in [+frequent] verbs

As the change progresses, *ra*-Deletion . . . diffuses into [+frequent] verbs, but not into [-frequent] verbs

Shin-ichiro Sano
Shift of Word Frequency Effect
Dynamic Word Frequency Effect and Lexical Diffusion

Generalization

Higher frequency – not always preferred context
not preferred context (resistant) $\Rightarrow$ preferred context (susceptible)

Lower frequency – stable
always not preferred context (resistant)

Lexical diffusion

<table>
<thead>
<tr>
<th></th>
<th>lower freq.</th>
<th>higher freq.</th>
<th>progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outset</td>
<td>resistant</td>
<td>resistant</td>
<td>$\Rightarrow$ same</td>
</tr>
<tr>
<td>Later stages</td>
<td>resistant</td>
<td>susceptible</td>
<td>$\Rightarrow$ different</td>
</tr>
</tbody>
</table>

Word Frequency Effect

1. not stable, but dynamic
2. strengthened: (1) early stage – n.s. $\Rightarrow$ lower p value
   (2) slope – progressively steeper
Effects of Factors: Stable or Dynamic?

**Constant Rate Effect:** (Kroch 1989)
- **Internal factors:** rate of progress – **stable/same** in all contexts
  \( \uparrow \) driven by a single underlying rule

**Dynamic Effect:** (Present Research)
- **External factors:** rate of progress – **different** in each context
  \( \uparrow \) driven by a variable/dynamic factor (e.g. word frequency)

<table>
<thead>
<tr>
<th>Parametrization of factor effects</th>
<th>effect</th>
<th>progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal factors</td>
<td>stable</td>
<td>( \Rightarrow ) same</td>
</tr>
<tr>
<td>External factors</td>
<td>dynamic</td>
<td>( \Rightarrow ) different</td>
</tr>
</tbody>
</table>
Conclusion

- Dynamic/variable effects of Word frequency
- Ongoing change in potential form – *ra*-Deletion

Correlation between:
1) probability of *ra*-Deletion and 2) frequency of verb stem
⇒ chronological transition

Results
1. positive correlation ⇒ higher frequency – preferred context
2. word frequency effect – variable/dynamic, strengthened
Conclusion

- Words with lower frequency – consistently resistant
  Words with higher frequency – resistant $\Rightarrow$ susceptible

$\downarrow$

- As the change progresses, innovative forms are more likely to occur in words with higher frequency.
  $\Rightarrow$ gap between lower and higher freq. *lexical diffusion*

- Word frequency – one of the causes for lexical diffusion
  Internal factors – regular change
  External factors – lexical diffusion (irregular change)
Thank you!
References


References II


References III


