EVALUATION OF MORA-BASED AND SYLLABLE-BASED TEXT-SETTING BY NATIVE SPEAKERS AND LEARNERS OF JAPANESE

REBECCA L. STARR

NATIONAL UNIVERSITY OF SINGAPORE, starr@nus.edu.sg

STEPHANIE S. SHIH

UNIVERSITY OF SOUTHERN CALIFORNIA, shihs@usc.edu

MAY 27, 2018

KEIO UNIVERSITY
THE MORA IN JAPANESE

- Japanese is a prototypical mora-based language: the mora, rather than the syllable, is the primary prosodic unit.

- Rhythmic timing depends on mora units (Homma 1981; Port et al. 1987; cf., Beckman 1982).

- Phonological/morphological implications:
  - e.g., accent placement, compensatory lengthening

  \[\text{kai.zen} = \text{[ka]}_{\mu} \cdot \text{[i]}_{\mu} \cdot \text{[ze]}_{\mu} \cdot \text{[n]}_{\mu}\]
HOW DO MORAS AND SYLLABLES COMBINE?

- Moras are units nested within syllables.
- A single syllable in Japanese may consist of a single ‘regular’ CV or V mora (e.g., ka).
- A multi-moraic syllable must consist of a regular mora plus one of a small set of ‘special’ mora:
  - Ex. 1: **coda nasal**: kan, [ka]_μ [N]_μ
  - Ex. 2: **Vi**: kai, [ka]_μ [i]_μ
  - and others
The role of the syllable in Japanese phonology has been the subject of controversy. Some argue that the syllable is unnecessary and does not exist in Japanese phonology (Labrune 2012). Phonological and psycholinguistic evidence, however, suggests that the syllable does play a role (Kubozono 1999; et seq.; Kawahara 2016; a.o.).
LEXICAL STRATA IN JAPANESE

- Major lexical strata in Japanese (Itô and Mester 1999):
  - Yamato (native Japanese)
    - Ex: 好き suki (‘to like’)
  - Sino-Japanese (Chinese origin)
    - Ex: 人間 ningen (‘human’)
  - Foreign (~85% English origin)
    - Ex: ベンチ benchi (‘bench’)
  - Mimetic
    - Ex: フワフワ fuwafuwa (‘fluffy’)
Strata characterized by different phonotactics, and phonological rules (Itô and Mester 1999).
- e.g., long a never occurs in Sino-Japanese words.

Because Chinese and English are syllable-based, is it possible that the syllable is a more salient unit in the Sino and Foreign strata?

Alternatively, does widespread knowledge of English contribute to increased salience of the syllable in just the Foreign stratum?
Text-setting: the pairing of language and music in song.

Typologically, text-setting makes use of salient prosodic units particular to a language.

English: syllables, lexical and phrasal stresses (Halle & Lerdhal 1993; Shih 2008; Hayes 2009; a.o.).

Text-setting in Japanese has been described as mora-based (Kubozono 1999; Hayes and Swiger 2008; cf. Manabe 2009).

- Each mora must receive (at least) one note.

5 sylls, 7 moras: do.ra.go-n bo-o.ru  
(Dragonball Z theme, 1989)
7 notes: x x x x x x x

- In these settings, the syllable receives at least one note.

Many cases of moraic and syllabic settings of the same word, even within a single song:

**moraic:** sa-n.ta no o.ji.sa-n ga

**syllabic:** de.mo so.no sa{n}.ta wa

*(I Saw Mommy Kissing Santa Claus, 1952, trans. 1962)*
MAIN RESEARCH QUESTIONS

- What constraints govern moraic vs. syllabic text-setting variation in Japanese?
- Do Japanese listeners perceive the syllable as an acceptable segmentation unit in text-setting?
  - Is it as acceptable as the mora?
- What about L2 learners of Japanese?
TWO APPROACHES

- Corpus study (Starr & Shih 2017)
- Experiment
Three corpora of Japanese songs compared:

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Type</th>
<th>Time Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anime theme songs</td>
<td>native</td>
<td>Late 1980s – 90s</td>
</tr>
<tr>
<td>Disney songs</td>
<td>translated</td>
<td>Late 1980s – 90s</td>
</tr>
<tr>
<td>Christmas songs</td>
<td>translated</td>
<td>Late 19^{th} c. – 21^{st} c.</td>
</tr>
</tbody>
</table>
WHY LOOK AT TRANSLATED SONGS?

- Japanese has a lower information density than English (and many other languages) (Pellegrino et al. 2011).
  - It takes longer to say the same thing in Japanese.

Ex:

*English:*
Well Ali Baba had them forty thieves Scheherazad-ie had a thousand tales

*Japanese:*  
Sou ai-ri-ba-ba ni wa yon-juu-nin mo-no to-u-zo-ku ga i-ta

(Friend Like Me, Aladdin, 1992)
Hypothesis: translation will put pressure on the text-setting system due to information density mismatch, thus making syllabic settings more likely.

This is interesting in itself, and also useful for us, as it gives us more opportunities to examine what linguistic environments favor syllabic settings.
CORPUS STUDY: VARIABLES

- Coda-Nasal
- Vi
- Long Vowels
- Inter-voiceless-consonant $i$ and $u$
Example | Moraic | Syllabic
---|---|---
*ningen* (‘human’) | *ni-n-ge-n* | *ni-n-gen*
*sekai* (‘world’) | *se-ka-i* | *se-ka*I
*hosshi* (‘want’) | *ho-shi-i* | *ho-shii*
*su*ki (‘like’) | *su-ki, suₕ-ki* | *(suₕ)*ki
CODING METHODOLOGY

- **Excluded:**
  - Code-switching involving two or more consecutive words in English with different parts of speech.
    - Ex: “It’s my day”
  - Content that was spoken rather than sung.
  - Discourse particles like ああ *aa*.
  - Nonsense words for which a stratum could not be determined.
  - Mimetic stratum words (not enough for analysis).
CORPUS STUDY: DISNEY EXAMPLE
FINDINGS: DISNEY VS. ANIME

Coda-N
Long-V
Vi
Inter-IU

% total settings
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%

syllabic
moraic

dis. anime
dis. anime
dis. anime
dis. anime

STARR & SHIH, MAY 27, 2018
FINDINGS: SETTING BY STRATUM

<table>
<thead>
<tr>
<th></th>
<th>Coda Nasal</th>
<th>Long V</th>
<th>VI</th>
<th>Inter-IU</th>
</tr>
</thead>
<tbody>
<tr>
<td>foreign</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sino</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yamato</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

% total setting

- Coda Nasal
- Long V
- VI

- syllabic
- moraic
FINDINGS: DISNEY VS. CHRISTMAS VS. ANIME

![Bar chart showing the percentage of syllabic and moraic settings for Disney, Christmas, and Anime.]
## FINDINGS: STATISTICAL MODELS

### Coda Nasal:

| Factor             | Estimate | Std. Error | z value | Pr (>|z|) |
|--------------------|----------|------------|---------|----------|
| (Intercept)        | -0.204   | 0.343      | -0.59   | 0.55     |
| Corpus = Anime     | -1.944   | 0.141      | -4.69   | < 0.0001 *** |
| Corpus = Disney    | -0.232   | 0.355      | -0.65   | 0.51     |
| Strata = Foreign   | 1.784    | 0.323      | 5.51    | < 0.0001 *** |
| Strata = Sino      | 0.210    | 0.299      | 0.70    | 0.48     |
| Word final = Yes   | -0.135   | 0.253      | -0.53   | 0.59     |

N = 368; syllabic = 158; moraic = 210
Long V:

| Factor            | Estimate | Std. Error | z value | Pr (>|z|) |
|-------------------|----------|------------|---------|----------|
| (Intercept)       | -0.579   | 0.509      | -1.136  | 0.26     |
| Corpus = Anime    | -1.271   | 0.356      | -3.56   | 0.0003 *** |
| Corpus = Disney   | -0.0939  | 0.306      | -0.307  | 0.76     |
| Strata = Foreign  | 1.108    | 0.288      | 3.84    | 0.0002 *** |
| Strata = Sino     | -0.181   | 0.253      | -0.714  | 0.48     |
| Word final = Yes  | 0.260    | 0.241      | 1.08    | 0.28     |
| Sonority          | 0.095    | 0.177      | 0.54    | 0.588    |

N = 414; syllabic = 179; moraic = 235
## FINDINGS: STATISTICAL MODELS

**Vi:**

| Factor                   | Estimate | Std. Error | z value | Pr (>|z|)    |
|--------------------------|----------|------------|---------|-------------|
| (Intercept)              | -4.385   | 0.791      | -5.55   | < 0.0001 ***|
| Corpus = Anime           | 1.227    | 0.74       | 1.66    | 0.0971 .    |
| Corpus = Disney          | 2.103    | 0.713      | 2.95    | 0.0032 **   |
| Strata = Foreign         | 1.152    | 0.515      | 2.24    | 0.0253 *    |
| Strata = Sino            | 0.662    | 0.300      | 2.21    | 0.0272 *    |
| Word final = Yes         | 0.201    | 0.312      | 0.65    | 0.5182      |
| Is ai? = Yes             | 1.002    | 0.382      | 2.63    | 0.0087 **   |

N = 381; syllabic = 72; moraic = 309
### FINDINGS: STATISTICAL MODELS

**Inter-i/u**

| Factor                        | Estimate | Std. Error | z value | Pr (>|z|)          |
|-------------------------------|----------|------------|---------|-------------------|
| (Intercept)                   | -6.5043  | 0.9753     | -6.67   | < 0.0001 ***      |
| Corpus = Anime                | 0.0291   | 0.7444     | 0.04    | 0.969             |
| Corpus = Disney               | 1.5882   | 0.6677     | 2.38    | 0.017 *           |
| Strata = Foreign              | 2.9482   | 0.5501     | 5.36    | < 0.0001 ***      |
| Strata = Sino                 | 0.5330   | 0.4868     | 1.09    | 0.274             |
| Good cluster? = Yes           | 3.7177   | 0.7254     | 5.12    | < 0.0001 ***      |
| Devoiced vowel = u            | -0.1214  | 0.4214     | -0.29   | 0.773             |

N = 381; syllabic = 72; moraic = 309
FINDINGS: CHANGE OVER TIME IN CHRISTMAS CORPUS
While all variables show instances of syllabic setting, some are more likely to receive syllabic settings.

- Coda N and Long V particularly likely to receive syllabic settings.

Syllabic settings most likely in Foreign stratum, for all four variables.

- For 3 of 4 variables, Foreign contrasts with Sino-Japanese and Yamato. For Vi, Foreign & Sino contrasts with Yamato.
- Most likely explanation is knowledge of English.
CORPUS STUDY: KEY FINDINGS

- Syllabic settings more likely in translated songs than in native songs.
  - As predicted, information density mismatch puts pressure on the system.

- When time period is accounted for, Disney & Christmas corpus (translated) pattern together, contrasting with anime corpus (native).
  - Consistent with previous work indicating that syllabic setting is becoming more frequent (Tanaka 2000).
BUILDING ON THESE FINDINGS

- A corpus study can tell us what trained composers and lyricists prefer to do.
- But, what about ordinary Japanese speakers?
- And what about L2 Japanese learners, who have not grown up speaking a mora-based language?
  - Can they acquire a new prosodic system?
EXPERIMENT RESEARCH QUESTIONS

- How do native Japanese speakers rate moraic vs. syllabic settings?
- Do native speakers show a preference for syllabic settings for Foreign stratum words relative to other words?
  - Does this vary depending on their exposure to English?
- Do native speakers show a preference for syllabic settings for certain segments (e.g., coda nasal)?
EXPERIMENT RESEARCH QUESTIONS

- Do L2 learners of Japanese match the text-setting intuitions of native speakers?
  - Does their level of experience with the language make a difference?
  - Does their own native language make a difference?
METHODOLOGY: VOCALOID

- Creating experimental stimuli involving multiple sung arrangements presents a challenge.
- Yamaha’s Vocaloid software allows for the production of synthesized sung Japanese.
METHODOLOGY: EXPERIMENT DESIGN

- **Linguistic variables:**
  - Coda Nasal (e.g., *minto*)
  - Vi (limited to *ai*) (e.g., *gaido*).

- **Strata: Foreign vs. Sino-Japanese**
  - Not possible to make minimal pairs with Yamato
  - Selected 10 near-minimal pairs.
    - Ex: ベンチ *benchi* / 便宜 *bengi*

- **Five-note melodies created using common chord progressions.**
METHODOLOGY: EXPERIMENT DESIGN

Settings:

- Mora: mi-n-to da-yo
- Syllable: min-to da-yo-ne
- Split Syllable (melisma): mi-in-to da-yo
- Bad 1 (too small): m-in-to da-yo
- Bad 2 (too large): mi-i-i-i-intodayo
PREDICTIONS

Native speakers:

- Will rate Moraic highest
- Will prefer Syllabic over Split-Syllabic
  - Syllabic: (min-to da-yo-ne)
  - 6 moras for only 5 notes. Since there is no moraic solution, listeners will be okay with syllabic solution.
  - Split-Syllabic: (mi-in-to da-yo)
  - 5 moras for 5 notes, but arrangement is non-moraic anyway. Turns short vowel into long vowel.

- Both syllabic settings will be preferred over the “bad” settings, which do not correspond to any proposed prosodic unit.
PREDICTIONS

Native speakers:

- Participants who use English frequently will give higher ratings to syllabic settings, particularly for Foreign stratum words.
Learners:

- Beginners will rate Moraic, Syllabic and Split-Syllabic settings equally.
  - While learners will quickly absorb the novel phenomenon of moraic settings in Japanese songs, it will take much longer to realize that syllabic settings are dispreferred.

- Will give lower ratings to Coda Nasal than Vi in the Moraic setting (e.g., mi-n-to vs. ga-i-do).
  - Coda nasals do not occur as independent units in the native languages of our learners (primarily English and/or Mandarin).
Variables and lexical strata:

- If judgments follow corpus pattern, Vi rating will not differ between Foreign & Sino strata, while Coda-N will receive more positive ratings for Foreign stratum.
METHODOLOGY: TASK DESIGN

Motsu-kun is learning to arrange lyrics to a melody. Help him improve by rating his work!

ミントだよね

(minto dayone)

• Asked to rate on four-point Likert scale.
PARTICIPANTS

- Survey conducted online using Qualtrics
- 97 native Japanese speakers
  - Asked for frequency of English use
- 30 L2 Japanese learners
  - Asked for native language(s), length of Japanese study
- Data analyzed using linear mixed-effects modeling via lmerTest in R, with participant and word as random effects.
FINDINGS

Rating by setting and speaker status

- Moraic
- Syllabic
- Split-Syllabic
- Bad

Rating (higher more positive)

- Interm. Learner (1-3 yrs)
- Adv. Learner (>3 yrs)
- Native
FINDINGS: RATINGS CONSISTENT OVERALL WITH EXPECTATIONS

Rating by setting and speaker status

Rating (higher more positive)

Moraic  
Syllabic  
Split-Syllabic  
Bad  

- Interm. Learner (1-3 yrs)  
- Adv. Learner (>3 yrs)  
- Native
FINDINGS: NATIVE SPEAKERS GENERALLY MORE CAUTIOUS ON TASK

Rating by setting and speaker status

- Moraic
- Syllabic
- Split-Syllabic
- Bad

Rating (higher more positive)

- Interm. Learner (1-3 yrs)
- Adv. Learner (>3 yrs)
- Native
# FINDINGS: INTERMEDIATE LEARNERS

## Significant fixed effects:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Est.</th>
<th>Std. Error</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syll. vs. Split-syll.</td>
<td>-1.10510</td>
<td>0.52584</td>
<td>54.24</td>
<td>-2.102</td>
<td>0.040250 *</td>
</tr>
<tr>
<td>Syll. vs. Bad</td>
<td>-1.03678</td>
<td>0.45756</td>
<td>55.28</td>
<td>-2.266</td>
<td>0.027394 *</td>
</tr>
<tr>
<td>Split syll.: native English speaker</td>
<td>1.65425</td>
<td>0.61072</td>
<td>52.95</td>
<td>2.709</td>
<td>0.009079 **</td>
</tr>
</tbody>
</table>
FINDINGS: INTERM. LEARNERS: ENGLISH SPEAKERS LOVE MELISMA (SPLIT-SYLL.)

Intermediate learner ratings by English background

Rating (higher more positive)

- Moraic
- Syll.
- Split-syll.
- Bad

Native English speaker
Not Native English speaker

STARR & SHIH, MAY 27, 2018
As predicted, while intermediate learners have figured out that moraic settings are acceptable in Japanese, they have not yet learned that syllabic settings are dispreferred.

- This effect may be limited to native English speakers: more diverse learner pool needed.
- Ratings very high among native English speakers for melisma (extending vowel over more than one notes), which is common in English-language pop music.

Surprisingly, ratings unaffected by nasal vs. Vi or Foreign vs. Sino-Japanese stratum.
Significant fixed effects:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Est.</th>
<th>Std. Error</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syll. vs. Moraic</td>
<td>0.2826</td>
<td>0.1619</td>
<td>225</td>
<td>-10.855</td>
<td>0.08221***</td>
</tr>
<tr>
<td>Syll. vs. Split-syll.</td>
<td>-0.2826</td>
<td>0.1619</td>
<td>225</td>
<td>-1.746</td>
<td>0.08221***</td>
</tr>
<tr>
<td>Syll. vs. Bad</td>
<td>-1.5217</td>
<td>0.1402</td>
<td>225</td>
<td>-10.855</td>
<td>&lt;.0001***</td>
</tr>
<tr>
<td>Native Chinese speaker</td>
<td>0.3808</td>
<td>0.1075</td>
<td>225</td>
<td>3.543</td>
<td>0.00048 ***</td>
</tr>
</tbody>
</table>
CHINESE SPKRS MORE POSITIVE AT EVERY STEP: MAYBE THEY LOVE VOCALOID?

Advanced learner ratings by Chinese background

- **Moraic**
  - Native Chinese speaker: [Rating (higher more positive)]
  - Not Native Chinese speaker: [Rating (higher more positive)]

- **Syll.**
  - Native Chinese speaker: [Rating (higher more positive)]
  - Not Native Chinese speaker: [Rating (higher more positive)]

- **Split-syll.**
  - Native Chinese speaker: [Rating (higher more positive)]
  - Not Native Chinese speaker: [Rating (higher more positive)]

- **Bad**
  - Native Chinese speaker: [Rating (higher more positive)]
  - Not Native Chinese speaker: [Rating (higher more positive)]
PRELIMINARY CONCLUSIONS: ADVANCED LEARNERS

- Advanced learners (>3 years of Japanese study) are approaching native speaker judgment levels.

- Native Chinese speakers (generally from Singapore) show more positivity on the task overall, unrelated to lexical stratum, setting type, etc.

- Unlike for intermediate learners, no effect here of being a native English speaker.

- Again, no effect of lexical stratum or nasal vs. Vi.
  - However: if we consider all learners in a single model, Vi items do receive higher ratings than nasal items in the moraic setting, as predicted.
FINDINGS: NATIVE SPEAKERS

Significant fixed effects:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Est.</th>
<th>Std. Error</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syll. vs. Moraic</td>
<td>0.30772</td>
<td>0.08469</td>
<td>856.2</td>
<td>3.633</td>
<td>0.000296 ***</td>
</tr>
<tr>
<td>Syll. vs. Split-Syll.</td>
<td>-0.27685</td>
<td>0.08471</td>
<td>855.8</td>
<td>-3.268</td>
<td>0.001125 **</td>
</tr>
<tr>
<td>Syll. vs. Bad</td>
<td>-0.78452</td>
<td>0.07337</td>
<td>856.9</td>
<td>-10.693</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td>Phon Vi</td>
<td>-0.13709</td>
<td>0.07152</td>
<td>17.30000</td>
<td>-1.917</td>
<td>0.071876 .</td>
</tr>
</tbody>
</table>

Note: Split-syll. rating also significantly better than Bad (p < .0001) when model is releveled.
FINDINGS: NATIVE SPKR RATINGS FOR NASAL VS. VI CONTRARY TO PREDICTIONS

Native speaker ratings for Coda Nasal vs. Vi

Rating (higher more positive)

- Coda Nasal
- Vi

moraic
syll.
split-syll.
bad
PRELIMINARY CONCLUSIONS: NATIVE SPEAKERS

- As predicted, a clear ranking of settings among native speakers:
  
  moraic < syllabic < split-syllabic < bad

- Contrary to findings of corpus study, however, there is no effect of lexical stratum, and coda nasal is rated as more acceptable in moraic, not syllabic settings.
  
  - This is also contrary to prediction of sonority.

- No effect found for frequency of speaking English.
EXPERIMENT: KEY FINDINGS

- While moraic settings are slightly preferred, native speakers rate syllabic settings as acceptable.
- Native speakers show a preference for a one-to-one correspondence between note and prosodic unit.
- No evidence for stratum effect, English exposure effect, or sonority effect.
- Learners with 1-3 years of study acquire knowledge of moraic setting, but lag in dispreference for syllabic settings.
  - Even at the advanced level, learners don’t look quite like native speakers in their judgments.
CONCLUSIONS

- Both corpus and experimental evidence support the view that the syllable is an active prosodic unit in Japanese.
  - Moreover, the salience of the syllable appears unrelated to frequency of English use.

- L2 learners of Japanese demonstrate quick acquisition of novel prosodic categories, but slow re-ranking of prosodic segmentation preferences.
FUTURE DIRECTIONS

- Further work needed to account for discrepancies between corpus study and experimental findings.
  - Is the preference for syllabic settings for Foreign stratum words limited to expert text-setters?
  - Why doesn’t more frequent exposure to syllabic settings for coda nasal in Japanese songs result in higher ratings in this experiment?

- Additional data collection needed to identify when beginner Japanese learners first acquire knowledge of moraic text-setting.
ACKNOWLEDGMENTS

Thanks to our participants, and to the many individuals who helped us distribute the survey: Mie Hiramoto, Kazuko Matsumoto, Kevin Heffernan, Elin McCready, Wilkinson Gonzales, Jun Jie Lim, Zhenhao Guo, Abigail Starr, the NUS Japanese Studies Society, and others.

Thanks again to Shigeto Kawahara for organizing this conference, and for his helpful comments on this project over the years.
REFERENCES


REFERENCES (CONT’D)


