Target vowel asymmetry in Brazilian Veneto metaphony

Guilherme D. Garcia and Natália B. Guzzo
BALL STATE UNIVERSITY • MCGILL UNIVERSITY

{guilhermegarcia, nataliaguzzo}.github.io

51st LSRL • University of Illinois at Urbana-Champaign
Apr 29th – May 1st 2021
Introduction

- Metaphony is observed in many Romance languages

- **Goal:** to explore the structural conditioning of metaphony in Brazilian Veneto (BV or Talian), an understudied dialect of Veneto

- **Why:** Although metaphony in Talian is assumed to be variable, little is known about how such variation is structured

- **How:** Through a corpus study (based on written data)
Brazilian Veneto

- Italian immigrants settled in Brazil starting ≈1875
- These immigrants settled in several areas of Brazil, especially in the southeastern and southern states
- Most immigrants came from Northern Italy (Veneto)
Brazilian Veneto

- Few communities spoke a single language
- Contact between varieties + scarce contact with Portuguese

This contributed to the development of a Veneto-based dialect: Brazilian Veneto (locally known as Talian)

<table>
<thead>
<tr>
<th>Region of origin</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Veneto</td>
<td>54.0</td>
</tr>
<tr>
<td>Lombardia</td>
<td>33.0</td>
</tr>
<tr>
<td>Trentino-Alto Adige</td>
<td>7.0</td>
</tr>
<tr>
<td>Friuli Venezia-Giulia</td>
<td>4.5</td>
</tr>
<tr>
<td>Others</td>
<td>1.5</td>
</tr>
</tbody>
</table>
Brazilian Veneto

- In the southern state of Rio Grande do Sul, Italian immigrants were assigned land in areas where there were virtually no other population groups.
Brazilian Veneto

- BV is closely related to Central Veneto
  - a dialect of Veneto spoken in Italy
- Both exhibit a trisyllabic window for stress assignment
- Both share a large portion of their lexica

(Frosi and Mioranza 1983; Belloni 2009)
Metaphony in (Brazilian, Central) Veneto

- In Veneto, stressed /e, o/ variably raise to [i, u] when followed by unstressed /i/ (Zamboni 1974; Walker 2005, 2010; Belloni 2009)
- The trigger is usually a separate morpheme
  - (plural marker or 2Ps inflection)
- Metaphony targets all stress positions (examples from Talian)

1. **Metaphony with penultimate stress:**
   - 'ov-i ~ 'uv-i
   - 'pes-i ~ 'pis-i
   - 'bev-i ~ 'biv-i
   - 'kor-i ~ 'kur-i

   - ‘egg.PL’
   - ‘fish.PL’
   - ‘drink.2Ps’
   - ‘run.2Ps’
2. **Metaphony with antepenultimate stress:**
   - The unstressed vowel in penultimate position also raises
     
     \[ 'z\text{oven}-i \sim 'z\text{uv}i\text{n}-i \]
     
     ‘young.PL’

     - It may also raise without a target in stressed position
     
     \[ '\text{omen}-i \sim '\text{omin}-i \]
     
     ‘man.PL’

3. **Metaphony with final stress:**

   \[ \text{fa}'z\text{o}-i \sim \text{fa}'z\text{u}-i \]  
   
   ‘bean.PL’ (sing. /fazol/)

   \[ \text{ni}'s\text{o}-i \sim \text{ni}'s\text{u}-i \]  
   
   ‘bedsheet.PL’ (sing. /nisol/)

---

Metaphony in (Brazilian, Central) Veneto

Garcia & Guzzo

Target vowel asymmetry in Brazilian Veneto metaphony

7 of 26
We examine metaphony in data from the Talian Corpus↑

Why written data?
1. Talian has no standardized orthography and grammar, so orthographic variation may reflect at least in part variation in the authors’ spoken language
2. Despite having no official orthography, authors are consistent in their writing, and grapheme-phoneme mapping is constant (e.g., letter $u = [\text{u}]$, letter $o = [\text{o, o}]$)
Methods

Talian corpus

• Little to no digitized text in Talian
• OCR\(^1\) using Tesseract with trained Italian data

**Materials:**

- Book excerpts and newspaper articles
- Newspaper: *Correio Riograndense* (founded 1909)

Another newspaper (*O Florense*, founded 1986) is also part of the corpus, but its excerpts can be accessed online in digitized form

---

\(^1\)Optical Character Recognition.
Methods

Talian corpus

1. Data preparation
   - >1 article per page
   - figures
   - multiple columns
   - line breaks
   - faded text (book excerpts)

2. OCR
   - proofreading
   - general fixes

3. Corpus compilation
   - R  
     (R Core Team 2020)

Newspaper article example
Methods
Talian corpus – available at the Open Science Framework
Methods
Talian corpus – available at the Open Science Framework

- Format: RData file with tidy data ready for analysis
- Size (as of April 2021): **237,774 words**

Available at [http://guilhermegarcia.nataliaguzzo.github.io/talian](http://guilhermegarcia.nataliaguzzo.github.io/talian)
### Methods

**Talian corpus**

- Currently coded for 25 variables:

<table>
<thead>
<tr>
<th>line</th>
<th>logFreq</th>
<th>nSyl</th>
<th>v_3</th>
<th>onset_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>sentence</td>
<td>author</td>
<td>syl_3</td>
<td>coda_3</td>
<td>v_1</td>
</tr>
<tr>
<td>wd</td>
<td>title</td>
<td>syl_2</td>
<td>onset_2</td>
<td>coda_1</td>
</tr>
<tr>
<td>sLength</td>
<td>year</td>
<td>syl_1</td>
<td>v_2</td>
<td>stressed_V</td>
</tr>
<tr>
<td>freq</td>
<td>IPA</td>
<td>onset_3</td>
<td>coda_2</td>
<td>stress</td>
</tr>
</tbody>
</table>
Methods

Analysis of metaphony

- Extract all words that (could) undergo metaphony (target items):
  - polysyllabic words ending in unstressed /i/
  - with an underlying /e, o/ in stressed position
  - manually check for mid-vowel quality in all words

- Words extracted using R script; checked for stress assignment and vowel quality

- Examples (orthographic form):
  - senti ‘feel.2ps’
  - amori ‘love.PL’
  - dóveni ‘young.PL’
  - curri ‘run.2PS’
  - cagniti ‘dog.DIM.PL’
  - fasui ‘bean.PL’

  (no metaphony)

  (metaphony)
Methods
Analysis of metaphony

• Given the distribution of tokens in our sample, we focus on:
  1. words with **penultimate** stress
  2. words with **2 and 3** syllables

• Total number of items: 2095 ($n = 490$ unique)

• Items coded for application of metaphony (response variable)
• Predictors examined:
  ○ **target vowel, number of syllables, morphology**, onsets, codas
Results & analysis

- **Asymmetry**: more metaphony with /o/ than /e/ (apparent interaction)
- Predictors related to phonotactic profile (e.g., onset, coda) had no clear effect
Results & analysis

- Hierarchical logistic regression

\[
\text{metaphony} \sim \text{targetV} \times \text{nSyl} + (1 \mid \text{author})
\]

Model estimates ($\hat{\beta}$), given in log-odds
Discussion

- What could be driving the asymmetry in question?
- One potential factor: **lexical statistics**

![Graph showing % of metaphony and number of syllables for target vowels e and o.]

![Graph showing % in lex and number of syllables for target vowels e and o.]

Entire corpus
Discussion

- Potential morphological effect: again asymmetrical pattern emerges
Discussion

- Hierarchical logistic regression

\[
\text{metaphony} \sim \text{targetV} \times \text{targetInRoot} + (1 \mid \text{author})
\]

Model estimates ($\hat{\beta}$), given in log-odds
## Discussion

<table>
<thead>
<tr>
<th></th>
<th>/e/</th>
<th>/o/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole corpus</td>
<td>more /e/s in 3-σ</td>
<td>more /o/s in 2-σ</td>
</tr>
<tr>
<td>Target items</td>
<td>more metaphony with /e/ in 3-σ</td>
<td>more metaphony with /o/ in 2-σ</td>
</tr>
<tr>
<td>$\sqrt{V}$</td>
<td>less metaphony</td>
<td>more metaphony</td>
</tr>
</tbody>
</table>

To formalize these results, we need:

- a **probabilistic** framework like MaxEnt (e.g., Goldwater and Johnson 2003; Wilson 2006; Hayes and Wilson 2008)
- some form of "lexical regulation" (e.g., Coetzee and Kawahara 2013)
- a motivation for [+hi] spreading (Walker 2010)
Formalization

- MaxEnt (weighted constraints, probabilistic)
- Constraints:
  - IDENT[hi]: don’t change [hi]
  - Lic([+hi] _posttonic_, σ): posttonic [+hi] is associated to σ²
  - OCP[hi]: don’t have sequences of identical [hi]

---

See Walker (2010) and references therein.
Formalization

Illustrative example for /e/ inputs

- Scaling factor regulates IDENT weight based on lexical stats
- Here, \( e-i_{\sigma\sigma} = 1 \) and \( e-i_{\sigma\sigma\sigma} = 0.4 \)

<table>
<thead>
<tr>
<th>‘fish.PL’</th>
<th>( w = 1^+ )</th>
<th>( w = 1 )</th>
<th>( w = 0.8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>/pes-i/</td>
<td>IDENT</td>
<td>OCP</td>
<td>Lic [+hi]</td>
</tr>
<tr>
<td>[pesi]</td>
<td></td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>[pisi]</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>‘poor.DIM.PL’</th>
<th>( w = 1^+ )</th>
<th>( w = 1 )</th>
<th>( w = 0.8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>/poaret-i/</td>
<td>IDENT</td>
<td>OCP</td>
<td>Lic [+hi]</td>
</tr>
<tr>
<td>[poareti]</td>
<td></td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>[poariti]</td>
<td>1</td>
<td>1</td>
<td>2.4</td>
</tr>
</tbody>
</table>
Final remarks

- Asymmetry between /e/ and /o/ in 2- and 3-σ words
- Patterns mirror what we see in the corpus as a whole (lexicon)
- Asymmetry also found when we look into morphology

- Variable application → probabilistic approach
- Lexical influence → some scaling factor

- So far: written data as a proxy for the grammar of Talian
- Next step: gather empirical data to evaluate that assumption
References


Acknowledgments:

• our RAs: Émilie Dubé, Alexandra Lancaster, Ray Marks, Fabian McCarthy, Lamia Oudni, Carolyn Rathgeber, and Jovia Wong (from McGill); and Jenna Gramlich (from Ball State)

• Dr. Valdemir Guzzo