# Target vowel asymmetry in Brazilian Veneto metaphony

Guilherme D. Garcia and Natália B. Guzzo

**BALL STATE UNIVERSITY • MCGILL UNIVERSITY**

$\left\{ \begin{array}{c} \texttt{guilhermegarcia} \\ \texttt{nataliaguzzo} \end{array} \right\}$ `.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 (Frosi and Mioranza 2009)
- Contact between varieties + scarce contact with Portuguese
- ☞ This contributed to the development of a Veneto-based dialect: Brazilian Veneto (locally known as **Talian**)

| Region of origin | % |
|---|---:|
| Veneto | 54.0 |
| Lombardia | 33.0 |
| Trentino-Alto Adige | 7.0 |
| Friuli Venezia-Giulia | 4.5 |
| Others | 1.5 |

# 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
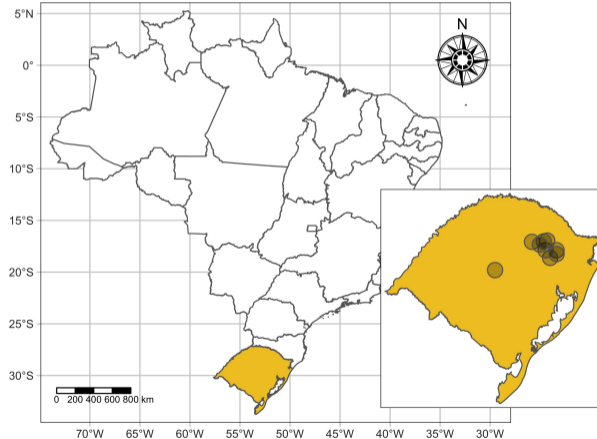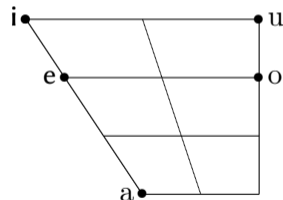
# Brazilian Veneto

- BV is closely related to Central Veneto <span>(Frosi and Mioranza 1983; Belloni 2009)</span>
  - a dialect of Veneto spoken in Italy
- Both exhibit a trisyllabic window for stress assignment
- Both share a large portion of their lexica
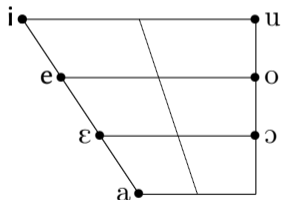


Stressed position



Unstressed (pre-tonic) position

# Metaphony in (Brazilian, Central) Veneto

- In Veneto, stressed /e, o/ variably raise to [i, u] when followed by unstressed /i/ <small>(Zamboni 1974; Walker 2005, 2010; Belloni 2009)</small>
- 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                                                                                    'egg.PL'

    'pes-i ∼ 'pis-i                                                                                  'fish.PL'

    'bev-i ∼ 'biv-i                                                                               'drink.2PS'

    'kor-i ∼ 'kur-i                                                                                 'run.2PS'

# Metaphony in (Brazilian, Central) Veneto

2. **Metaphony with antepenultimate stress:**

- The <u>unstressed vowel</u> in penultimate position also raises

  ˈzoven̲-i ∼ ˈzuvin̲-i                                                    'young.PL'

- It may also raise without a target in stressed position

  ˈɔmen̲-i ∼ ˈɔmin̲-i                                                    'man.PL'

3. **Metaphony with final stress:**

  faˈzo-i ∼ faˈzu-i                                        'bean.PL' (sing. /fazol/)

  niˈso-i ∼ niˈsu-i                                    'bedsheet.PL' (sing. /nisol/)

# This presentation

- We examine metaphony in data from `the Talian Corpus`[↑]   (Garcia and Guzzo 2021)

**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* = [u], letter *o* = [o, ɔ])

# Methods
**Talian corpus**

- Little to no digitized text in Talian
- OCR[1] using Tesseract with trained Italian data <span style="float:right">(Smith 2007)</span>
- **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
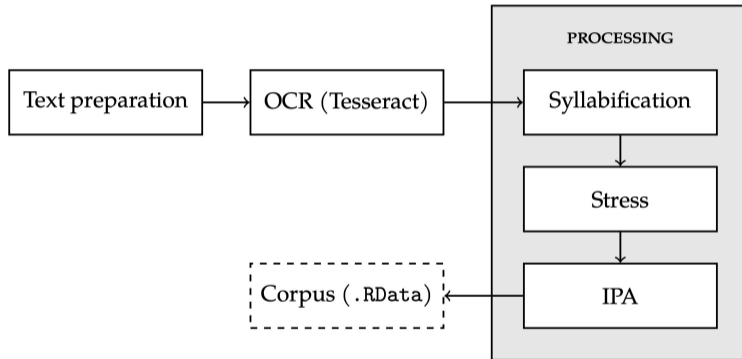   - R                    (R Core Team 2020)



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   (Wickham et al. 2014)
- Size (as of April 2021): **237,774 words**

☞ Available at `http://`$\left\{\begin{array}{c}\texttt{guilhermegarcia}\\\texttt{nataliaguzzo}\end{array}\right\}$`.github.io/talian`

## Methods

**Talian corpus**

- Currently coded for 25 variables:

| line | logFreq | nSyl | v_3 | onset_1 |
|------|---------|---------|--------|-----------|
| sentence | author | syl_3 | coda_3 | v_1 |
| wd | title | syl_2 | onset_2 | coda_1 |
| sLength | year | syl_1 | v_2 | stressed_V |
| freq | IPA | onset_3 | coda_2 | stress |

## 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 ($n = 3088$)
  - 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.2ᴘs' (no metaphony)
  - amori 'love.ᴘʟ'
  - dóveni 'young.ᴘʟ'
  - curri 'run.2ᴘs' (metaphony)
  - cagniti 'dog.ᴅɪᴍ.ᴘʟ'
  - fasui 'bean.ᴘʟ'

## 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
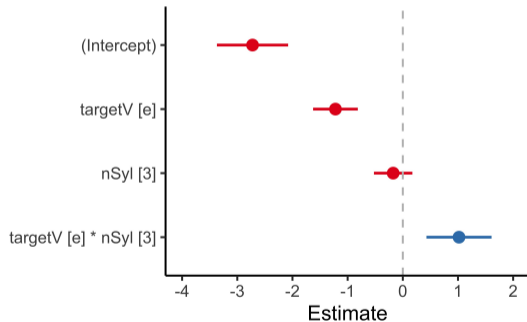
# Results & analysis
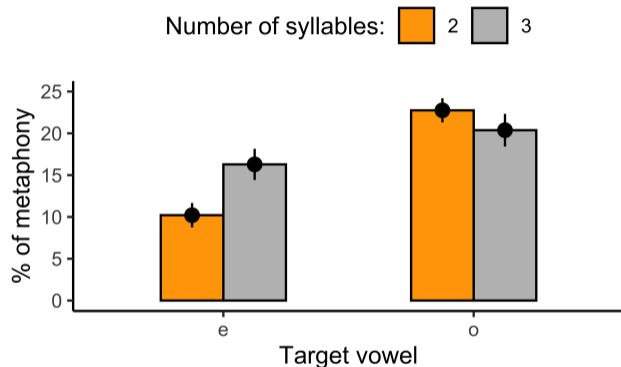


- **Asymmetry:** more metaphony with /o/ than /e/ (apparent interaction)
- Predictors related to phonotactic profile (e.g., onset, coda) had no clear effect
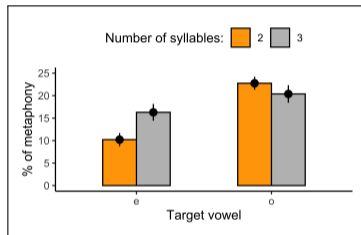
# Results & analysis

- Hierarchical logistic regression
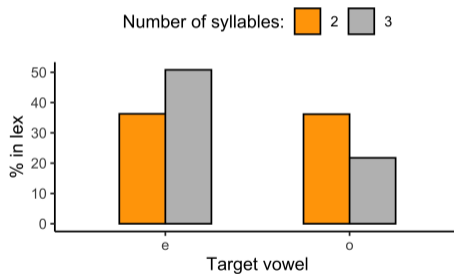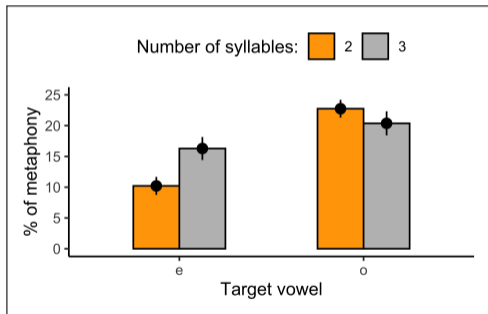
  `metaphony ~ targetV * nSyl + (1 | author)`



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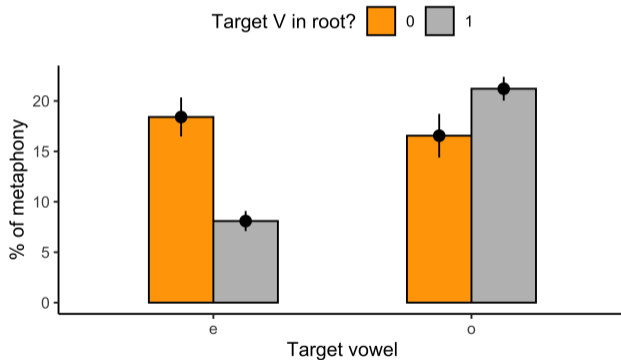
# Discussion

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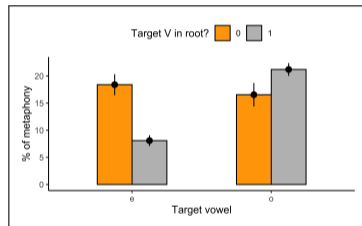- One potential factor: **lexical statistics**





Entire corpus
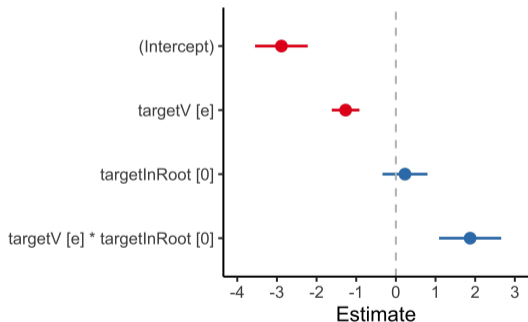
# Discussion

- Potential morphological effect: again asymmetrical pattern emerges

# Discussion

- Hierarchical logistic regression

  metaphony $\sim$ targetV $*$ targetInRoot + (1 | author)



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

# Discussion

|  | /e/ | /o/ |
|---|---|---|
| **Whole corpus** | more /e/s in 3-$\sigma$ | more /o/s in 2-$\sigma$ |
| **Target items** | more metaphony with /e/ in 3-$\sigma$ | more metaphony with /o/ in 2-$\sigma$ |
| $\sqrt{\textbf{V}}$ | less metaphony | more metaphony |

☞ To formalize these results, we need:

• a **probabilistic** framework like **MaxEnt** <span>(e.g., Goldwater and Johnson 2003; Wilson 2006; Hayes and Wilson 2008)</span>

• some form of **"lexical regulation"** <span>(e.g., Coetzee and Kawahara 2013)</span>

• a motivation for [+hi] spreading <span>(Walker 2010)</span>

# Formalization

- MaxEnt (weighted constraints, probabilistic)
- Constraints:
  - IDENT[hi]: don't change [hi]
  - LIC([+hi]$_{posttonic}$, $\acute{\sigma}$): posttonic [+hi] is associated to $\acute{\sigma}$[2]
  - OCP-V: don't have sequences of identical vowels



---
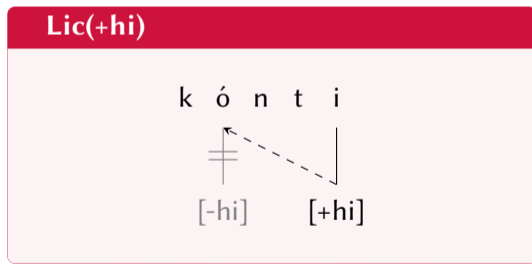
[2]See Walker (2010) and references therein.

# 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

'fish.PL'  $w = 1^+$   $w = 1$   $w = 0.8$

| /pes-i/ | IDENT | OCP | Lic[+hi] | h(x) | P*(x) | P(x) | R |
|---------|-------|-----|----------|------|-------|------|---|
| [p**e**si] | | | 1 | 0.8 | 0.449 | 0.90 | 0.898 |
| [p**i**si] | 1 | 1 | | 3 | 0.05 | 0.10 | 0.102 |

'poor.DIM.PL'  $w = 1^+$   $w = 1$   $w = 0.8$

| /poaret-i/ | IDENT | OCP | Lic[+hi] | h(x) | P*(x) | P(x) | R |
|------------|-------|-----|----------|------|-------|------|---|
| [poar**e**ti] | | | 1 | 0.8 | 0.449 | 0.83 | 0.837 |
| [poar**i**ti] | 1 | 1 | | 2.4 | 0.091 | 0.17 | 0.163 |

# Final remarks

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

- Variable application $\rightarrow$ probabilistic approach
- Lexical influence $\rightarrow$ 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

Belloni, S. (2009). *Grammatica veneta*. Esedra, Padova.

Coetzee, A. W. and Kawahara, S. (2013). Frequency biases in phonological variation. *Natural Language & Linguistic Theory*, 31(1):47–89.

Frosi, V. M. and Mioranza, C. (1983). *Dialetos italianos: um perfil linguístico dos ítalo-brasileiros do Nordeste do Rio Grande do Sul*. EDUCS, Caxias do Sul.

Frosi, V. M. and Mioranza, C. (2009). *Imigração italiana no nordeste do Rio Grande do Sul*. EDUCS, Caxias do Sul, 2nd edition.

Garcia, G. D. and Guzzo, N. B. (2021). Talian corpus: a written corpus of Brazilian Veneto.

Goldwater, S. and Johnson, M. (2003). Learning OT constraint rankings using a Maximum Entropy model. In *Proceedings of the Stockholm workshop on variation within Optimality Theory*, pages 111–120.

Hayes, B. and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39(3):379–440.

R Core Team (2020). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Smith, R. (2007). An overview of the tesseract OCR engine. In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, volume 2, pages 629–633.

Walker, R. (2005). Weak triggers in vowel harmony. *Natural Language and Linguistic Theory*, 23:917–989.

Walker, R. (2010). Nonmyopic harmony and the nature of derivations. *Linguistic Inquiry*, 41:169–179.

Wickham, H. et al. (2014). Tidy data. *Journal of Statistical Software*, 59(10):1–23.

Wilson, C. (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science*, 30(5):945–982.

Zamboni, A. (1974). *I dialetti del Veneto*. Pacini, Pisa.

# Thank you • Obrigad$_a^o$ • Gràssie

**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