

# The transmission of dialect features between first- and second-generation Bahian migrants in the state of São Paulo

Leonardo Ferraz (Unicamp)  
Advisor: Livia Oushiro (USP/Unicamp)

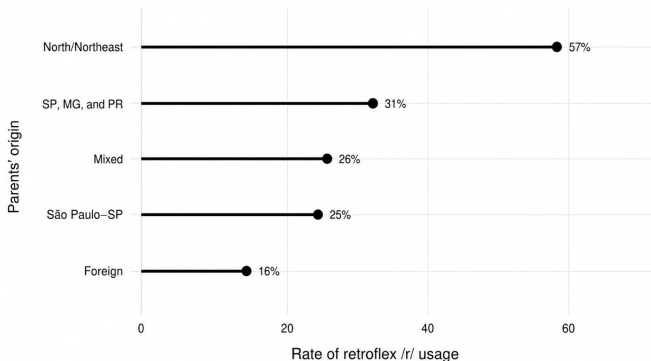
FAPESP Grant 2025/12308-7  
CAPES Grant 88887.949895/2024-00



# Migrants' children in speech communities

Oushiro (2015)

- One of the linguistic variables analyzed: alternation between tap [ɾ] and retroflex [ɻ] coda /r/ in São Paulo Portuguese (as in *porta* 'door')
- Parents' place of origin taken into account
- **Main finding:** The retroflex variant was favored by the children of migrants from Northern and Northeastern Brazil



# Dialect contact and migrants' children

Payne (1976, 1980)

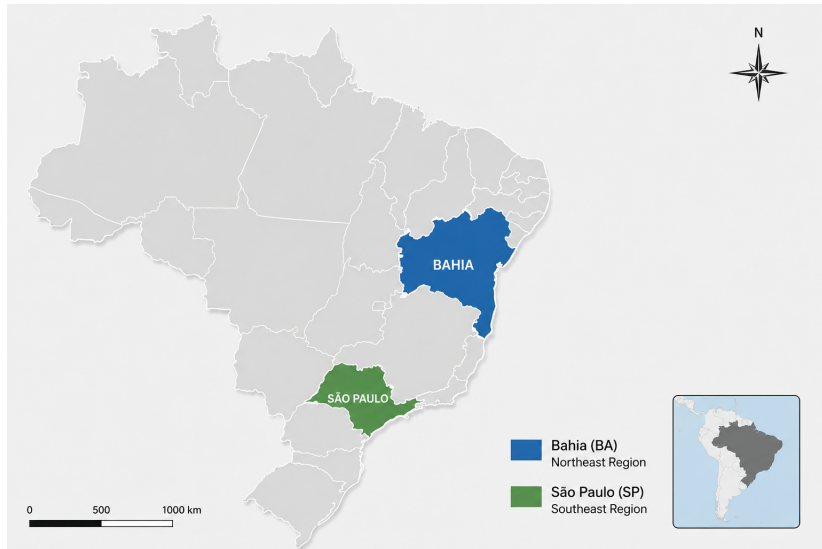
- Examining the relative influence of parents vs. peers
- Linguistic variables analyzed:
  - *Phonetic variables*: /ay/ (as in time) and /aw/ (as in house)
  - *Phonological variable*: the /er–ər/ merger (as in ferry vs. furry)
- Phonetic variables are acquired more readily than phonological variables
- **Main finding**: Migrants' children play an important role in the processes of language variation and change observed in Philadelphia

# Variable rules in a dialect contact situation

Guedes (2019) and Oushiro and Guedes (2019)

- Dialect contact between Paraiban migrants and native speakers from São Paulo
- Linguistic variable analyzed:
  - Presence vs. absence of the definite article before possessives: *a minha mãe* vs.  $\emptyset$  *minha mãe* (lit.: '(the) my mother')
- Separate analyses of Northeastern migrants, non-migrant Northeasterners, and native speakers from São Paulo
- **Main finding:** Migrants adapted to the host community not only in their rates of article use, but also in the variable rules governing article use

# Bahian migrants in São Paulo



# Bahian families in Palmares Paulista (SP)



- Located in the the Northwestern region of São Paulo State
- Area: 80 km<sup>2</sup> (IBGE, 2022)
- Population: 9,650 (IBGE, 2022)
- Main economic activity: sugarcane industry

# Research question

*Given the presence of Bahian migrant families in Palmares Paulista, how are surface level features and abstract patterns of variation transmitted from one generation to the next?*

# Corpus

## Control sample

- Iboruna corpus (Gonçalves, 2007): 10 non-mobile speakers from São José do Rio Preto-SP
- Semiarido corpus (Almeida and Carneiro, 2008): 10 non-mobile speakers from Feira de Santana-BA

## Target sample

53 Bahian migrants living in Palmares Paulista (Ferraz, 2024; 2025):

- First generation: 26 speakers
- Second generation: 27 speakers (children of the first-generation migrants)

# The variables

Coda /r/ and pretonic mid-vowels /e, o/

## Linguistic variables analyzed:

- Coda /r/ (e.g., *sorvete* 'ice cream' and *cantar* 'to sing')
  - Rhotic deletion
- Pretonic mid-vowels /e, o/ (e.g., *menino* 'boy' and *colega* 'colleague')
  - Vowel raising and lowering

## Levels of analysis:

- Usage rates
  - Frequency and proportion of variant use
- Phonetic detail
  - Acoustic parameters (e.g., syllable duration, F1 and F2 values, F2–F3 convergence, etc.)
- Variable rules (Labov, 1969; Guy, 1991; Oushiro, 2015; Guedes, 2019; Oushiro and Guedes, 2019)

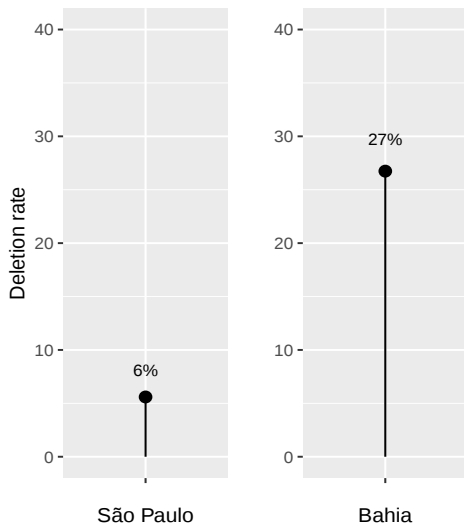
# Methods

- Transcription in ELAN (Sloetjes and Wittenburg, 2008)
- Auditory coding in ELAN (op. cit.) and forced alignment with the Montreal Forced Aligner (McAuliffe et al., 2019)
- Automated data extraction and coding of predictor variables via scripts
- Categorical and continuous analyses of sociolinguistic variables
- Bayesian mixed-effect logistic and linear regression models (R Core Team, 2026)
- Separate analyses for each (sub)sample and comparisons among them

# Some preliminary results

Variable rules for the /r/-deletion in the speech of native Bahians and Paulistas

- **Response variable:**  
*Deletion* (colhe/ø/ 'spoon') vs.  
*Realization* (colhe[ʎ,h])
- **Total number of tokens:**  
 1,867 in the Iboruna corpus and  
 2,025 in the Semiarido corpus



# Some preliminary results

Variable rules for the /r/-deletion in the speech of native Bahians and Paulistas

**Table:** Effects of linguistic predictors on coda /r/-deletion in São Paulo and Bahia

| Predictor variable                     | São Paulo | Bahia |
|----------------------------------------|-----------|-------|
| Preceding. phon. cont. (Height)        | ✓         | x     |
| Preceding. phon. cont. (Backness)      | ✓         | x     |
| Following phon. cont. (Manner of art.) | x         | x     |
| Following phon. cont. (Voicing)        | x         | x     |
| Word class                             | ✓         | x     |
| Syllable stress                        | x         | x     |
| Coda position                          | ✓         | ✓     |

# Discussion

## From control samples to migrant speech

- Native Bahia and São Paulo speakers exhibit distinct patterns of variation
- Differences involve both:
  - Overall deletion rates
  - Variable rules for the coda /r/-deletion
- These “control grammars” provide the baseline for analyzing intergenerational transmission

## Next steps

- Complete the auditory coding of the corpus
- Review the phonetic alignments produced by the MFA (McAuliffe et al., 2019)
- Analyze coda /r/ and the pretonic mid-vowels (acoustically and categorically) in the migrant samples and the two control samples
- Perform statistical analyses for each sample and compare the linguistic patterns across all four samples

# References

-  Almeida, N. L. F. and Z. O. N. Carneiro (2008). *Coleção Amostras da Língua Falada no Semiárido Baiano*. Feira de Santana: UEFS.
-  Ferraz, L. (2024). “Palmamirim: um estudo sobre a plasticidade dialetal de migrantes baianos residentes no município de Palmares Paulista-SP”. MA thesis. Campinas: IEL, UNICAMP.
-  — (2025). *A transmissão intergeracional de características dialetais do português baiano em situação de contato com o português paulista*. Projeto de Tese (FAPESP 2025/12308-7). Campinas: IEL, UNICAMP.
-  Gonçalves, S. C. L. (2007). *Banco de dados Iboruna: amostras eletrônicas do português falado no interior paulista*. Processo FAPESP nº 2003/08058-6. URL: <http://www.alip.ibilce.unesp.br>.
-  Guedes, S. (2019). “Emprego do artigo definido em situação de contato dialetal”. In: *Domínios de Lingu@gem* 13.4, pp. 1401–1432. DOI: 10.14393/dl40-v13n4a2019-4.
-  Guy, G. R. (1991). “Explanation in variable phonology: An exponential model of morphological constraints”. In: *Language Variation and Change* 3, pp. 1–22.

# References

-  IBGE (2022). *Censo Demográfico de 2022*. Instituto Brasileiro de Geografia e Estatística.
-  Labov, W. (1969). “Contraction, Deletion, and Inherent Variability of the English Copula”. In: *Language* 45.4, pp. 715–762.
-  McAuliffe, M. et al. (2019). *Montreal Forced Aligner*.  
<https://montreal-forced-aligner.readthedocs.io/>. Accessed: 2023-10-19.
-  Oushiro, L. (2015). “Identidade na pluralidade: avaliação, produção e percepção linguística na cidade de São Paulo”. PhD thesis. São Paulo: FFLCH, USP. 394 pp. DOI: 10.11606/t.8.2015.tde-15062015-104952.
-  Oushiro, L. and S. Guedes (2019). *Usage rates and variable rules: what changes in migrants’ speech*. Paper presented at NWAV48, Eugene-OR.
-  Payne, A. (1976). “The acquisition of the phonological system of a second dialect”. PhD thesis. Philadelphia: University of Pennsylvania. 287 pp.
-  Payne, A. C. (1980). “Factors controlling the acquisition of the Philadelphia dialect by out-of-state children”. In: *Locating Language in Time and Space*. Ed. by W. Labov. New York, NY: Academic Press, pp. 143–178.

# References

-  R Core Team (2026). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
-  Sloetjes, H. and P. Wittenburg (2008). *ELAN. Linguistic annotator*. Computer software. Version 6.3. Nijmegen, The Netherlands. URL: <https://archive.mpi.nl/tla/elan/>.

**Vielen Dank!**

✉ [lferraz@dac.unicamp.br](mailto:lferraz@dac.unicamp.br)

