

Pronominal system variation: What can we learn from Twitter data?



Mariana Centanin Bertho
Adriana Picoral
Isabella Calafate de Barros
Marina Carcamo-Garcia
Maria Carolina Zuppardi



Introduction

Gap:

- Pronominal system variation has been extensively studied, but not on Twitter data
- The study of linguistic variation in Twitter data has become more common only recently, and has focused on English (Abitbol et al., 2018; Grieve et al., 2019)

Our Study:

- Brazilian Twitter pronominal system variation
- Focus on first and second person pronouns in Portuguese
- Language and variation approach

Why Tweets?

Unmonitored written data (McCulloch, 2019)

Large amount of linguistic data with geographic metadata (Eisenstein, 2018)

Studies on English internet language and variation:

- patterns of lexical innovation in American Twitter (Grieve, Nini & Guo, 2018)
- lexical dialect variation in British English (Grieve et al., 2019)

Studies on Spanish internet language and variation:

- lexical variation across dialects (Gonçalves & Sánchez, 2020)
- innovations in Spanish Internet orthography (Myslín & Gries, 2010)

Portuguese is the 5th language used in the web (Internet World Stats, 2020)

Why Tweets?

Variation is present in tweets, as shown in this super token (Tagliamonte, 2019) example from our data:

Ninguém tá passando pano aqui, **nós** queremos respostas e que eles falem sobre isso, só que **nos** vivemos ataques das gringas por mais de um mês e isso faz com que **a gente** fique com medo disso ser fake e ser mais uma armação delas.

*No one is ignoring what is happening, **we** want answers and that they talk about this, but **we** are under foreign attack for over a month and because of that **we** fear that this is fake and that this is another scheme of theirs.*

What we know about pronominal variation: nós vs a gente

1st person plural forms:

<i>old standard</i>	"nós comemos" (we eat)
<i>non-standard</i>	"nós come" (we eats)
<i>emerging standard</i>	"a gente come" (we eat)

Scherre, Naro & Yacovenko (2018)

The noun *gente* ("people") lost its [+plural] marking, but kept its collective interpretation with the singular agreement

→ *a gente* has the discourse features of first-person plural but the grammatical features of third-person singular

What we know about pronominal variation: nós vs a gente

Zilles (2005): *a gente* is favored by different linguistic and social factors

- linguistic factors: proximity of the subject, reference, and subject in previous clause → *A gente* is favored when the subject is distant from the verb, the reference is generic and *a gente* is used in the previous clause.
- social factors: age, and the interaction between gender and education → *A gente* is favored by young speakers and is “avoided” by males with a lower level of education.

Travis & Silveira (2009): the high type frequency of *a gente* favors the spread of this form over *nós*; *nós* is still frequent, but only in a small number of frequent contexts and formulaic forms

What we know about pronominal variation: *você* vs *tu*

General 2nd person singular forms: ***tu*** and ***você***

*Additional forms: **cê**, **ocê*** - growing tendency on using variations of ***você*** (Othero, 2013)

Sociopragmatic (relationship among the speakers) and **geographic** variation

(Gonçalves, 2008; Ponzo Peres, 2006; Loregian-Penkall & Menon, 2012; Almeida Ferrari, 2013; Guimaraes, De Araújo & Pereira, 2018; Scherre, Andrade & Catão, 2020)

Other Romance languages:

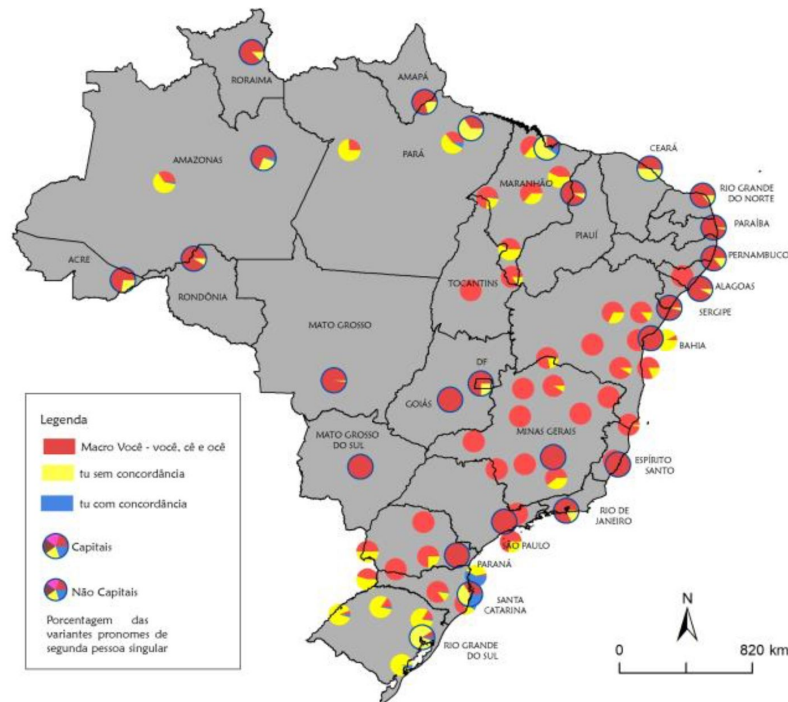
- Spanish (**both types of variation**, cf. Blas Arroyo, 2008; Moyna, Kluge & Simon, 2019)
- French (**more sociopragmatic**, cf. Brown & Gilman, 1960; Gardner-Chloros, 2007)

What we know about pronominal variation: você vs tu

Geographic variation

Scherre, Andrade & Catão (2020):

- **Macro você - você, ocê, cê:** generally used across Brazil
- **Tu without agreement - tu vai (3rd p. sg.):** mostly used in Rio Grande do Sul state, also present in some specific areas (e.g., Rio de Janeiro, Fortaleza)
- **Tu with agreement - tu vais (2nd p. sg.):** more common in Santa Catarina and Paraná area



Methods

Tweets containing target pronouns (all orthographic variations) were downloaded from June 12 2020 to July 20 2020 using the rtweet R package (Kearny, 2019)

Envelope of variation

- We kept only **subject pronouns** in our analysis:

pq **a gente** sempre quer, quem n quer **a gente**???
*why do **we** always love those that don't love **us**???*

5,002 tweets have been coded by hand to date

- Approx. 40% of tweets retained for analysis (i.e., 60% of tweets were eliminated)

Methods

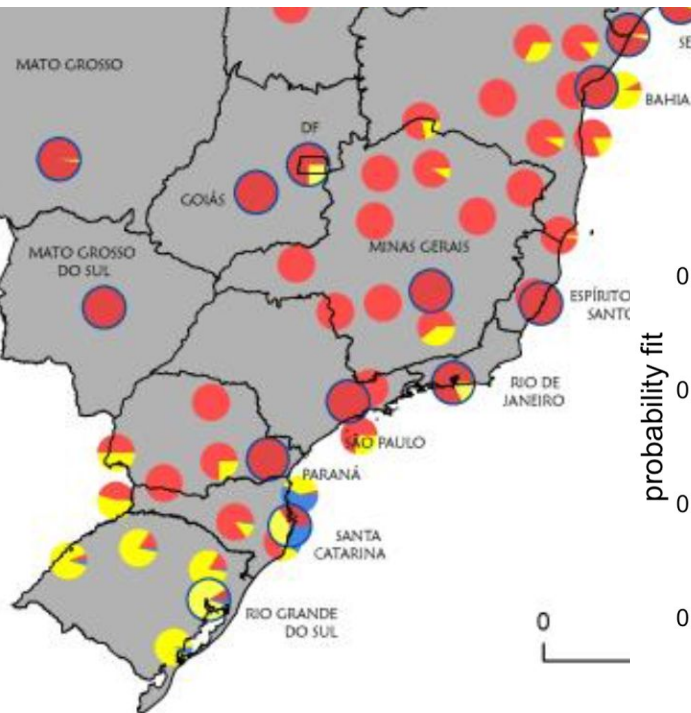
Region	First Person Token Count	Second Person Token Count
Florianópolis	359	26
João Pessoa	445	15
Porto Alegre	255	642
Recife	415	0
Rio de Janeiro	607	704
Salvador	125	365
São Paulo	417	627

For more details on our methods, see Lívio et al. (2021)

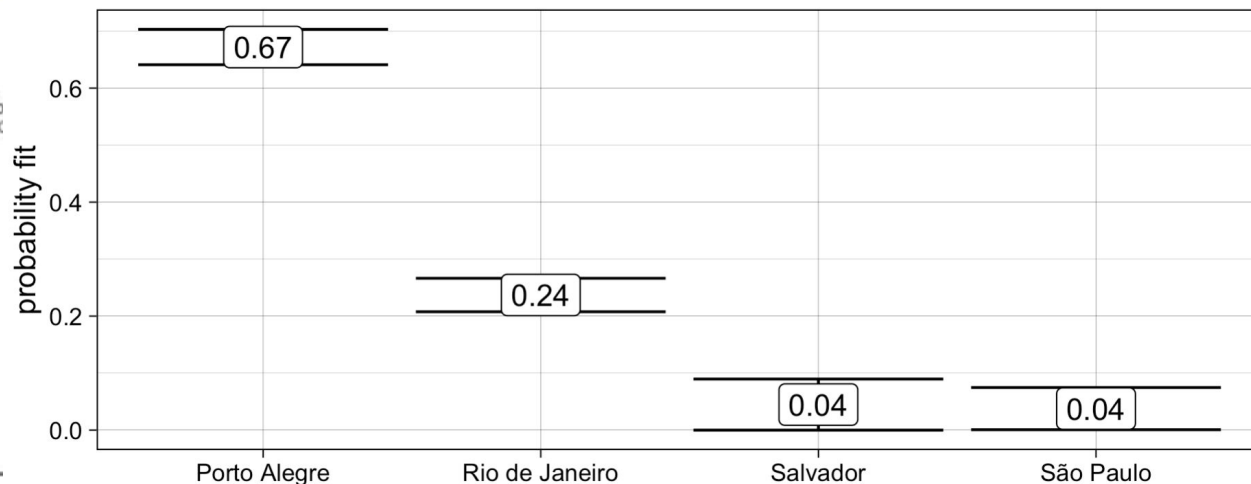
Results - Second Person (você vs. tu)

Location ($p < .0001$)

	n	proportion	logodds	factor	weight
Porto Alegre	621	66.02	2.23	0.90	
Rio de Janeiro	673	22.29	0.31	0.58	
Salvador	351	7.98	-1.03	0.26	
São Paulo	410	4.88	-1.51	0.18	72

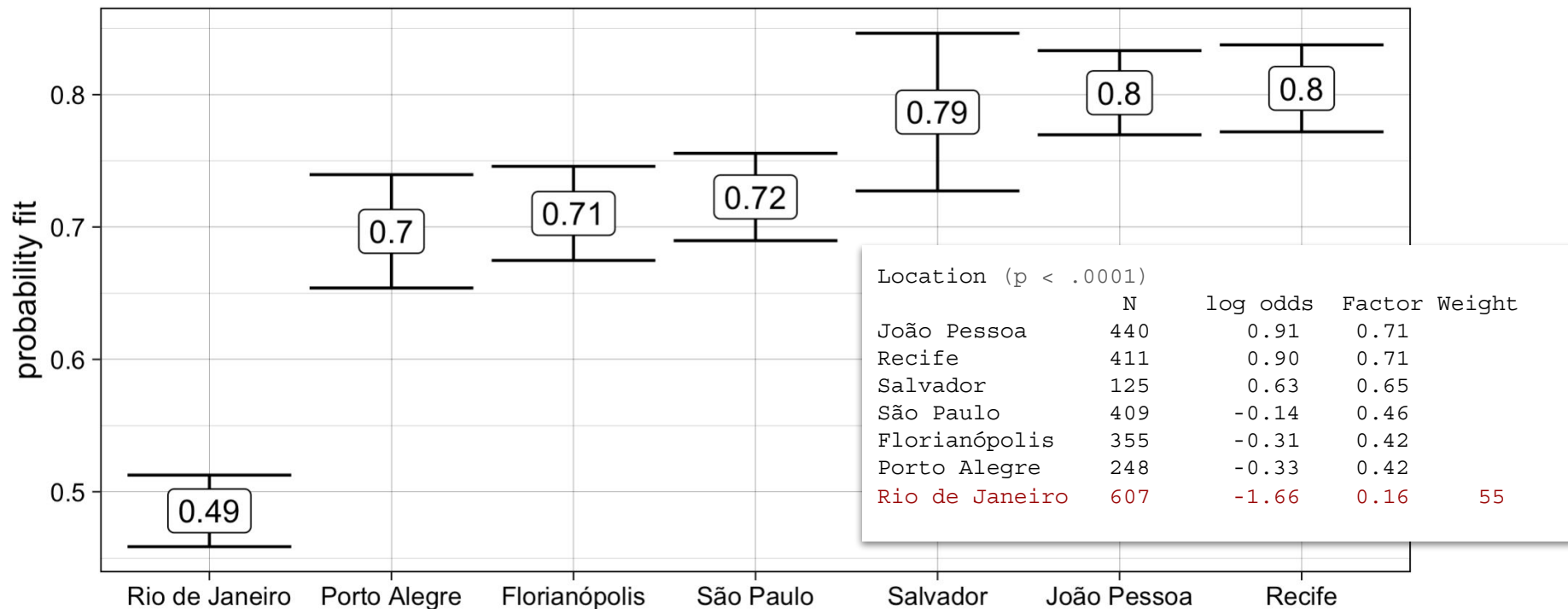


Probability fit of use of `tu` (vs. você)



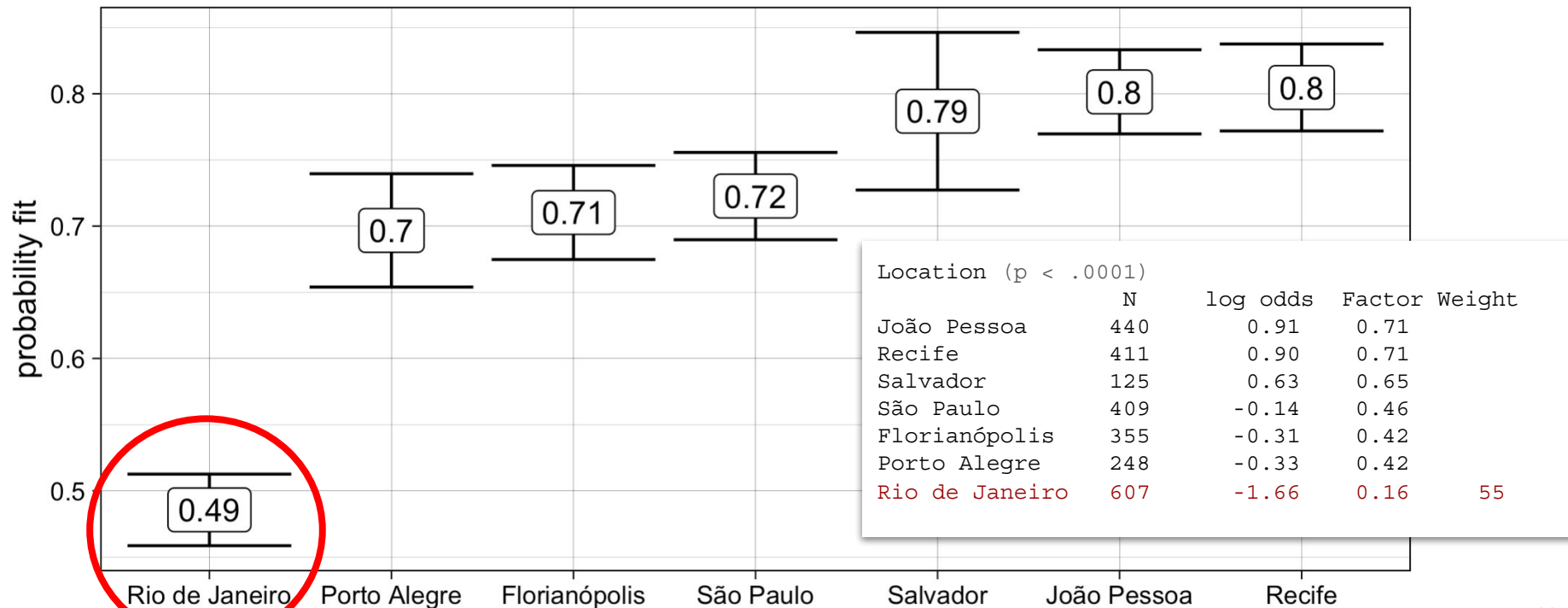
Results - First Person (nós vs. a gente)

Probability fit of use of `a gente` (vs. nós)



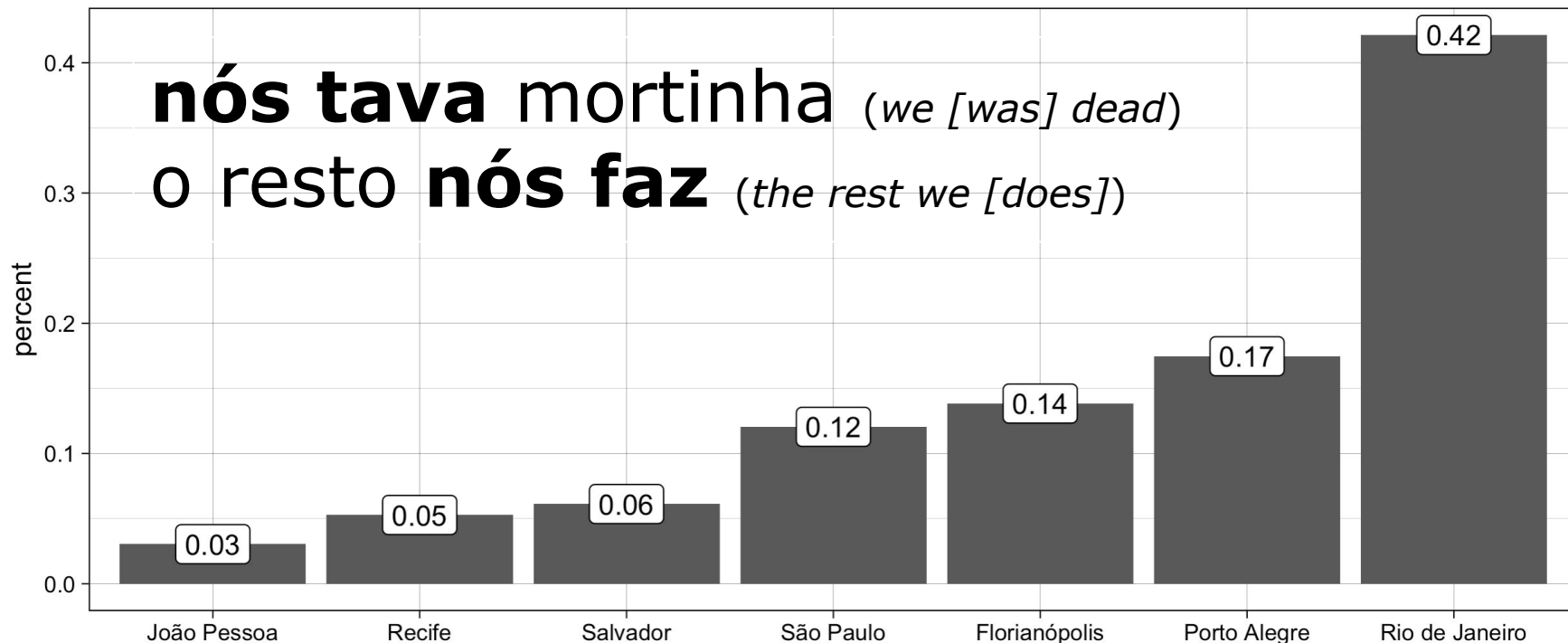
Results - First Person (nós vs. a gente)

Probability fit of use of `a gente` (vs. nós)



Results - First Person (nós vs. a gente)

Distribution of `third singular` agreement for `nós` (we) across locations



Discussion

Overall, location is a significant factor in the variation of first and second person pronouns.

- *tu* vs *você* → Results of the twitter data analysis reflect what has been shown in the literature: Porto Alegre highly favors *tu*, Rio de Janeiro also favors *tu* although to a lesser extent, and Salvador and São Paulo favor the use of *você*.
- *nós* vs *a gente* → Our findings indicate that João Pessoa, Recife and Salvador favor *a gente*. Rio de Janeiro is the location that most favors the use of *nós*. In addition, looking at verb agreement, Rio de Janeiro highly retains *nós* with third singular agreement (considered to be *non-standard*).

Discussion

Previous studies have shown that

- variation between tu/você is highly conditioned by geographic location (in addition to sociopragmatic factors).
- On the other hand, variationist studies on nós/a gente, although investigating data from different regions, have highlighted the effect of linguistic factors and social factors that do not depend on location.

Twitter data show us that geographic location is in fact very relevant for both variables. In other words, in the case of first person pronoun, twitter data allow us to observe that location seems to be more relevant than expected and not only in terms of pronoun selection but also in terms of agreement. Compared to other locations analyzed, Rio de Janeiro presents a high use of *nós* with third singular agreement (considered to be the *non-standard* use).

Next steps

Other factors have been included in the analysis but not yet discussed:

- use of internet language (e.g., contractions, emoji, laughter)
- generic/specific reference
- SV/VS order

Code more data

- Now that we have a coding scheme in place, an interface is being developed for more consistent coding

Code for pragmatic function

- We still need to decide on a coding scheme for pragmatic function

Our research group

girlanguages.com



Adriana Picoral
University of Arizona



Bruna Sommer-Farias
Michigan State University



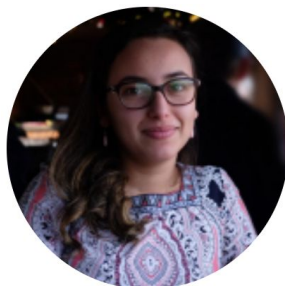
Camila Lívio
University of Georgia



Elisa Stumpf
Pelotas Federal University



Isabella Calafate de Barros
University of Arizona



Larissa Goulart
Northern Arizona
University



Laura Fontana Soares
University of Arizona



Mariana Centanin Bertho
University of Arizona



Marina Carcamo Garcia
University of Arizona



Marine Laísa Matte
Rio Grande do Sul Federal
University

Pronominal system variation: What can we learn from Twitter data?



Mariana Centanin Bertho

Adriana Picoral

Isabella Calafate de Barros

Marina Carcamo-Garcia

Maria Carolina Zuppardi

marianabertho@email.arizona.edu

adrianaps@arizona.edu

isabellacb@email.arizona.edu

mcarcamogarcia@email.arizona.edu

carolzupp@gmail.com

girlanguages.com

References

- Jacob Levy Abitbol, Márton Karsai, Jean-Philippe Magué, Jean-Pierre Chevrot, and Eric Fleury. (2018) Socioeconomic Dependencies of Linguistic Patterns in Twitter: a Multivariate Analysis. In *Proceedings of the 2018 World Wide Web Conference (WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1125–1134. DOI:<https://doi.org/10.1145/3178876.3186011>
- Brown, R. & Gilman, A. (1960). The Pronouns of Power and Solidarity. In Sebeok, T. A. *Style in Language*, pp. 253-276. Cambridge, Massachusetts: MIT Press.
- Blas Arroyo, J. L. (2008). *Sociolingüística del español: desarrollos y perspectivas en el estudio de la lengua española en contexto social*. Madrid: Cátedra.
- Cook, M. (2013) Portuguese Pronouns and Other Forms of Address, from the Past into the Future— Structural, Semantic and Pragmatic Reflections. In *Ellipsis* 11: pp. 267-290. American Portuguese Studies Association.
- Eisenstein, J. E. (2018). Identifying Regional Dialects in On-Line Social Media. In C. Boberg, J. Nerbonne, & D. Watt (Eds.), *The Handbook of Dialectology* (pp. 368–383). Wiley.
- Gardner-Chloros, P. (2007). T/V choices: An act of identity? In W. Ayres-Bennett & M.C. Jones (Eds.), *The French Language and Questions of Identity* (pp. 106–115). Oxford: Legenda.
- Grieve, J., Montgomery, C., Nini, A., Murakami, A., & Guo, D. (2019). Mapping lexical dialect variation in British English using Twitter. *Frontiers in Artificial Intelligence*, 2, 11.
- Grieve, J., Nini, A., & Guo, D. (2018). Mapping Lexical Innovation on American Social Media. *Journal of English Linguistics*, 46(4), 293–319. <https://doi.org/10.1177/0075424218793191>
- Gonçalves, C. R. (2008). *Uma abordagem sociolinguística dos usos das formas 'você', 'ocê' e 'cê' no português*. Doctoral Dissertation, University of São Paulo.
- Gonçalves, B., & Sánchez, D. (2016). Learning about Spanish dialects through Twitter. *Revista Internacional De Lingüística Iberoamericana*, 14(2 (28)), 65-75. Retrieved February 27, 2021, from <http://www.jstor.org/stable/26379776>
- Guimaraes, T. A. A. S., De Araújo, A. A. & Pereira, M. L. D. (2018). Os pronomes você(s) e cê(s) no falar de Fortaleza-CE. *Letras Escreve* 8 (2). Macapá, Brazil.
- Internet World Stats (2009). Internet world users by language: top 10 languages. URL <<http://www.internetworldstats.com/stats7.htm>>, accessed 12 July 2009.
- Lívio, C., Sommer-Farias, B., Marchioro Stumpf, E., Goulart, L., & Picoral, A. (2021). Investigating pronominal variation using Twitter data: advantages, challenges, and lessons learned. *American Association for Applied Linguistics*.
- Loregian-Penkal, L. & Menon, O. P. D. (2012). Você, Ocê (?) e Cê em Curitiba, Paraná. *SIGNUM: Estud. Ling.*, 15/1, p. 223-243.
- McCulloch, G. (2020). *Because internet: Understanding the new rules of language*. Riverhead Books.
- Moyna, M. I., Kluge, B., & Simon, H. J. (2019). Kluge, B., & Moyna, M. I. (Eds.). *It's not all about you : New perspectives on address research*. Amsterdam ; Philadelphia : John Benjamins Publishing Company
- Myslín, M., & Gries, S. T. (2010). k dixez? A corpus study of Spanish Internet orthography. *Literary and linguistic computing*, 25(1), 85-104.
- Scherre, M. M. P., Andrade, C. Q. & Catão, R. C. (2020). Redesenhando o mapa dos pronomes tu/você/cê/ocê no português brasileiro falado In: *Resumos expandidos das Comunicações Individuais do V Congresso Nacional de Estudos Linguísticos - CONEL*, 2019, Vitória- ES.
- Scherre, M. M. P., Naro, A. J., & Yacovenco, L. C. (2018). Nós e a gente em quatro amostras do português brasileiro: revisitando a escala da saliência fônica. *Diadorim* (Rio de Janeiro), v. 20, n. 1, p. 420-450.
- Tagliamonte, S. (2016). *Making waves : the story of variationist sociolinguistics*. Wiley Blackwell.
- Travis, C. & Silveira, A. S. (2009). The role of frequency in first-person plural variation in Brazilian Portuguese: Nós vs. a gente. *Studies in Hispanic and Lusophone Linguistics* 2, 347- 376.
- Zilles, A. M. S. (2005). The development of a new pronoun: the linguistic and social embedding of a gente in Brazilian Portuguese. *Language Variation and Change* 17, 19-53.