# Null subjects in Portuguese
# and the typology of conditioning on variation

Gregory R. Guy

New York University

Penn Linguistics Conference 47

March 18, 2023

# Why null subjects?

- A prototypical syntactic variable, occurring in many languages

    1. Eu vou.   'I'm going'
    2. [Ø] vou.   'I'm going'

- Addressed in the literature from a variety of perspectives
  - parameter theory [NSLs vs Non-NSLs]
  - functionalist models
  - variationist studies
- It appears to be widely subject to several potentially universal constraints, as well as language-specific constraints and some social factors

# This project

- Variation in subject pronoun expression [**SPE**] between null and expressed pronominal subjects in dialects of Portuguese.
- Data: sociolinguistic interviews from the following corpora
  - São Paulo – **SP2010 corpus** (Mendes, USP)
  - Lisbon – **ComparaPort corpus** (Vieira and Brandão, UFRJ)
  - Rio – **ComparaPort corpus**
  - Funchal **– ComparaPort corpus**
- Thanks to these colleagues for their generosity and assistance: Ronald Mendes, Silvia Vieira, Silvia Brandão, Karen Beaman, Danny Erker
- Special thanks to my research assistants:
  - Emilia Vieira Branco and Julia Vieira Branco (Portugal)
  - Thaís Melo (Brasil)
- Part of a bigger project comparing SPE across five languages, with D. Erker, R. Orozco, A. Adli, R. Bayley and K.Beaman

# The corpus*

| Speech community | Number of Speakers | Social dimensions | Tokens | % overt subject pronouns |
|---|---|---|---|---|
| São Paulo | 44 | 2 ed., 3 age, 2 gender | 15895 | 66.2 |
| Rio de Janeiro | 36 | 3 ed., 3 age, 2 gender | 9776 | 64.6 |
| Lisbon | 36 | 3 ed., 3 age, 2 gender | 9746 | 32.0 |
| Funchal | 18 | 3 ed., 3 age, 2 gender | 4602 | 32.3 |
| Total | 101 | | 40,019 | |

- Coding underway for Maputo, Moçambique

# Constraints on linguistic variation

- Tamminga, MacKenzie & Embick 2016
  - s-conditioning (social)
  - i-conditioning (internal linguistic)
  - p-conditioning (psychophysiological)
- Social distribution: who uses what variant more and when?
  - dialects, speakers, age, class, gender, ethnicity…
  - change: leaders and laggers
  - style: indexicality and social meaning
- Linguistic conditioning: what contexts favor what variants?
- Universals: what constraints derive from universal properties of the language faculty, articulation, cognition, neurology…

# s-conditioning: Social distribution

- Four dialects
  - São Paulo, Brazil
  - Rio de Janeiro, Brazil
  - Lisbon, Portugal
  - Funchal, Madeira, Portugal

- Speaker Gender (binary)

- Not significant:
  - Speaker Age
  - Speaker Education (proxy for social class)

# Dialects:
## The big picture

- SPE rate in BP is double the EP rate
- Within each country, the dialects are very similar
- In each city, the speakers are clustered around common values
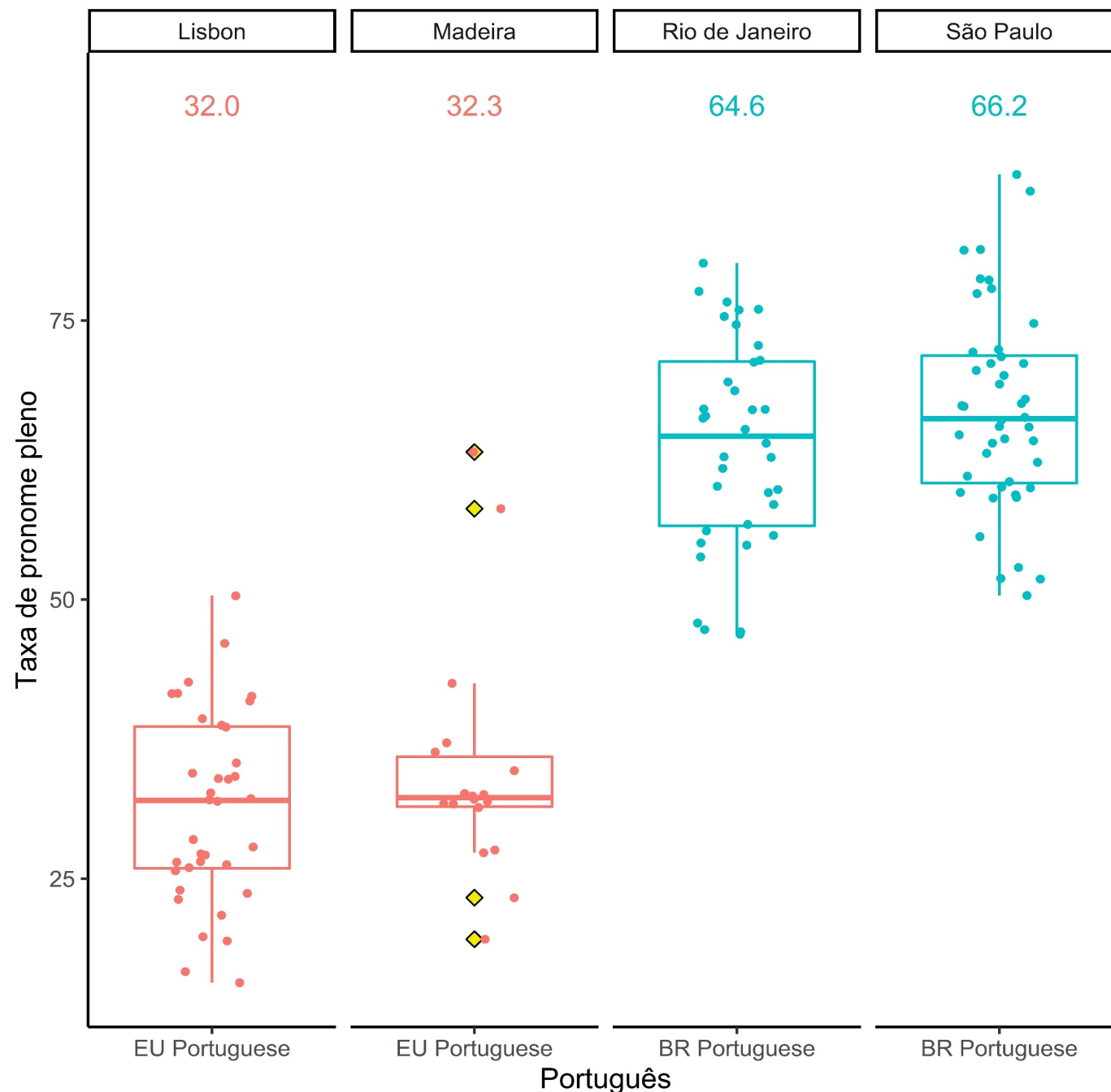
# i-conditioning: Internal linguistic constraints

- Person/number of subject
    - 1st and 3rd person singular and plural
    - 2nd singular *tu* (familiar - Portugal only), *você* (all cities)
    - 2nd plural *vocês (*archaic *vós* does not occur)
    - *a gente:* a newly grammaticized pronoun for 'we' taking 3rd sg inflection

- Syntax
    - main clause
    - subordinate clause

- Other constraints not considered here:
    - Verbal semantics
    - Tense/Mood/Aspect: weak effect, not considered here
    - Animacy. Inanimates strongly disfavor, excluded from this corpus

# p-conditioning (Cognitive, potential universals?)

- Priming or persistence
  - Was a prior pronominal subject expressed?
    - overt pro
    - null
- Continuity of Reference/ Discourse accessibility
  - Was the subject referent of the previous tensed clause the same as or different from the current tensed clause?
    - same
    - different

# Priming: null favors null, overt favors overt

| | Priming Context | | | | | | |
|---|---|---|---|---|---|---|---|
| | null | | | overt | | | |
| | *n* | % overt | weight | *n* | % overt | weight | est. |
| SP | 3172 | 50.0 | 0.35 | 6342 | 71.6 | 0.58 | 0.840*** |
| RIO | 1993 | 43.5 | 0.33 | 3284 | 69.9 | 0.60 | 1.047*** |
| LIS | 3844 | 21.8 | 0.44 | 1662 | 37.8 | 0.62 | 0.679*** |
| MAD | 1543 | 21.9 | 0.41 | 768 | 46.4 | 0.69 | 1.007*** |

Table 1
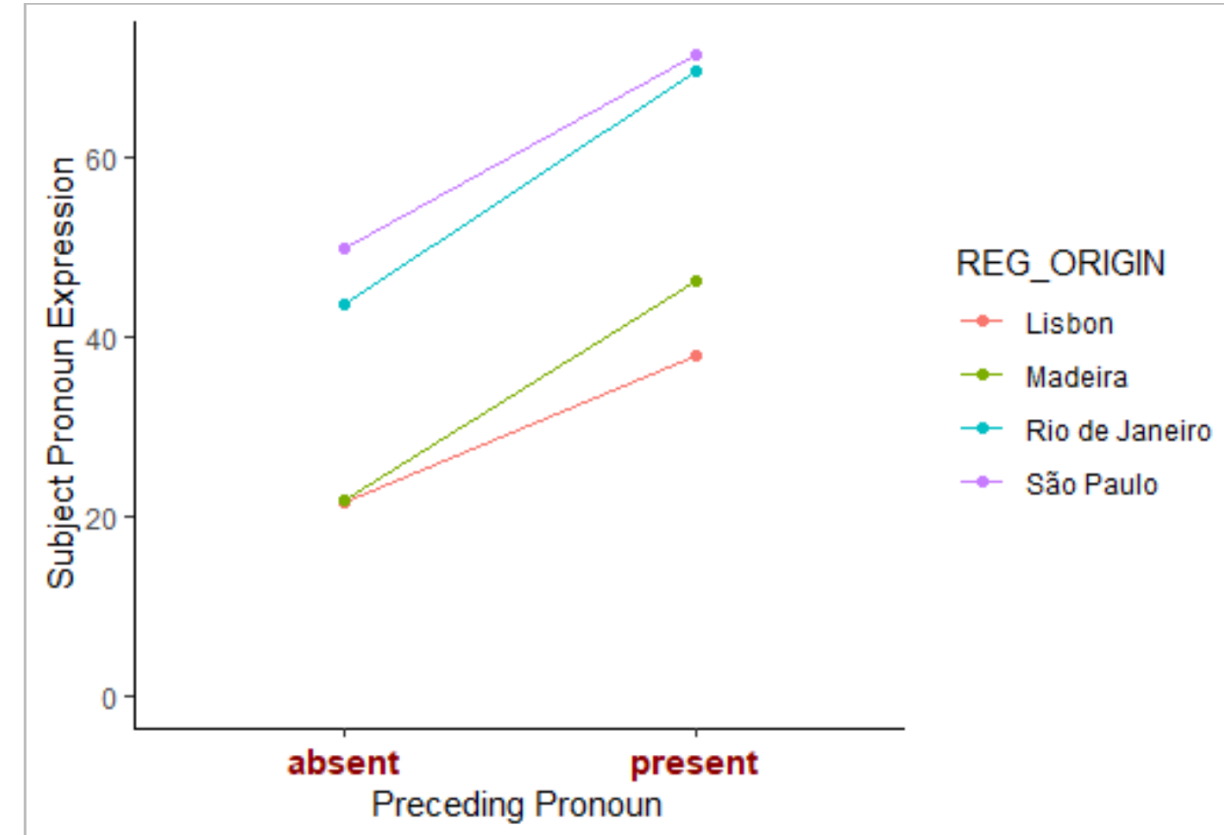


Figure 1

# Reference continuity: new referents favor overt pro

| | Reference continuity | | | | | | |
|---|---|---|---|---|---|---|---|
| | same | | | different | | | |
| | N | % overt | weight | N | % overt | weight | est |
| SP | 8564 | 60.8 | 0.442 | 4653 | 75.1 | 0.605 | 0.701*** |
| RIO | 4590 | 53.5 | 0.409 | 3081 | 73.6 | 0.633 | 0.906*** |
| LIS | 4716 | 18.8 | 0.376 | 2936 | 46.1 | 0.694 | 1.305*** |
| MAD | 2210 | 22.9 | 0.384 | 1185 | 49.0 | 0.707 | 1.353*** |

Table 2
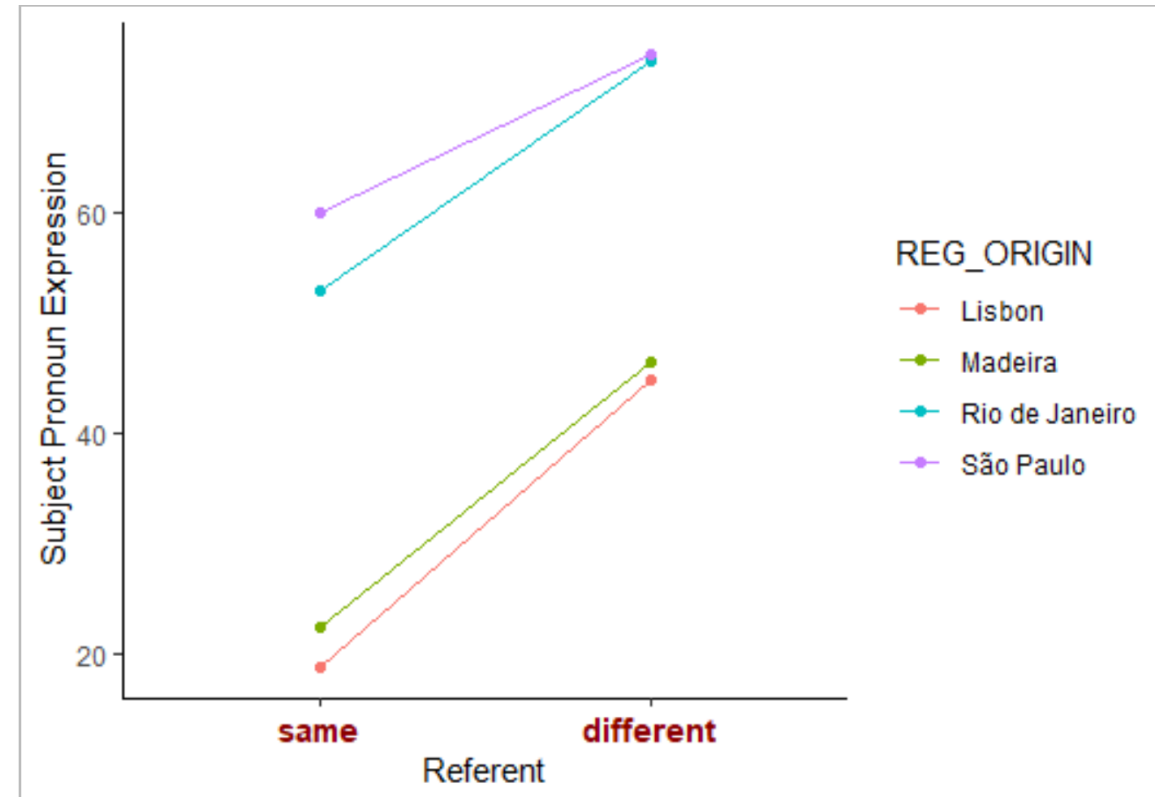


Figure 2

# Syntax: subordinate clauses favor overt pro

| | Clause Type | | | | | | |
|---|---|---|---|---|---|---|---|
| | main | | | subordinate | | | |
| | N | % overt | weight | N | % overt | weight | est. |
| SP | 12407 | 63.5 | 0.457 | 3091 | 80.6 | 0.666 | 0.929*** |
| RIO | 7778 | 59.9 | 0.448 | 1892 | 81.9 | 0.703 | 1.062*** |
| LIS | 7434 | 28.3 | 0.459 | 2165 | 44.9 | 0.639 | 0.745*** |
| MAD | 3566 | 30.5 | 0.449 | 939 | 51.9 | 0.686 | 1.089*** |

Table 3



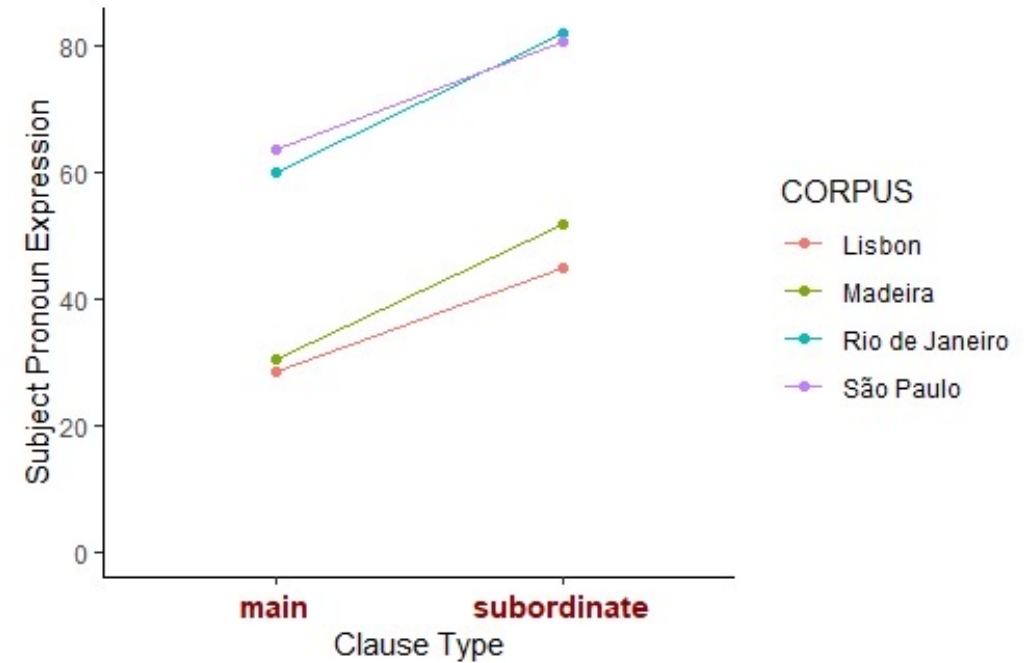Figure 3

# Subordinate clauses and the null subject parameter

- Subordinate clause SPE

  1. Jânia disse que Ø vem amanhã. 'Jania said that Ø is coming tomorrow.'
  2. Jânia disse que ela vem amanhã. 'Jania said that she is coming tomorrow.'

  Claim:  In NSLs (1) must mean that Jania is coming tomorrow, and (2) must mean that someone else is coming tomorrow.

  We find no such categorical association in any dialect.  In our data, this is merely a reflection of the switch reference effect: same subject as higher clause disfavors overt pro, but does not disallow it.

# i-conditioning: person-number

Classic

|      | Singular | Plural |
|------|----------|--------|
| 1st  | eu falo  | nós falamos |
| 2nd  | tu falas | vós falais |
| 3rd  | ele/ela fala | eles/eles falam |

Contemporary colloquial

|      | Singular | Plural |
|------|----------|--------|
| 1st  | eu falo  | nós falamos/a gente fala |
| 2nd  | (tu falas) você fala | vocês falam |
| 3rd  | ele/ela fala | eles/eles falam |

# i-conditioning: person/number. 1st singular

| | 1st singular (weights) | 2nd singular *tu* | 2nd singular *você* | 3rd singular | 1st plural *nós* | 1st plural *a gente* | 2nd plural *vocês* | 3rd plural |
|---|---|---|---|---|---|---|---|---|
| SP | 0.483 | --- | 1.009*** | -0.195*** | -0.967*** | 0.920*** | 12.522 | -0.533*** |
| RIO | 0.521 | --- | 0.269*** | -0.538*** | -1.112*** | 0.927*** | 0.454 | -0.895*** |
| LIS | 0.547 | -0.241 | -1.235*** | -0.260** | -0.563*** | 2.236*** | 1.069 | -0.724*** |
| MAD | 0.525 | -1.209*** | -1.372** | 0.414*** | -0.781*** | 1.920*** | -0.467 | -0.798*** |

Table 4
NB: 1st singular: above the grand mean in Rio, Lisbon and Madeira, below it in São Paulo

# i-conditioning: person/number – 3rd singular

- lower SPE rate than 1st sg (reference value) for all communities except Madeira

|  | 1st singular (weights) | 2nd singular *tu* | 2nd singular *você* | 3rd singular | 1st plural *nós* | 1st plural *a gente* | 2nd plural *vocês* | 3rd plural |
|---|---|---|---|---|---|---|---|---|
| SP | 0.483 | --- | 1.009*** | -0.195*** | -0.967*** | 0.920*** | 12.522 | -0.533*** |
| RIO | 0.521 | --- | 0.269*** | -0.538*** | -1.112*** | 0.927*** | 0.454 | -0.895*** |
| LIS | 0.547 | -0.241 | -1.235*** | -0.260** | -0.563*** | 2.236*** | 1.069 | -0.724*** |
| MAD | 0.525 | -1.209*** | -1.372** | 0.414*** | -0.781*** | 1.920*** | -0.467 | -0.798*** |

# i-conditioning: person/number
# 1st and 3rd plural strongly disfavor overt pro

| | 1st singular (weights) | 2nd singular *tu* | 2nd singular *você* | 3rd singular | 1st plural *nós* | 1st plural *a gente* | 2nd plural *vocês* | 3rd plural |
|---|---|---|---|---|---|---|---|---|
| SP | 0.483 | --- | 1.009*** | -0.195*** | -0.967*** | 0.920*** | 12.522 | -0.533*** |
| RIO | 0.521 | --- | 0.269*** | -0.538*** | -1.112*** | 0.927*** | 0.454 | -0.895*** |
| LIS | 0.547 | -0.241 | -1.235*** | -0.260** | -0.563*** | 2.236*** | 1.069 | -0.724*** |
| MAD | 0.525 | -1.209*** | -1.372** | 0.414*** | -0.781*** | 1.920*** | -0.467 | -0.798*** |

NB: These each have distinctive inflections in all TMA: -mos and –am/-ão/-ram

# i-conditioning: person/number – 2nd singular

- Portugal has T/V contrast, both forms disfavor overt pro, especially *você*
- *você* is the only 2nd sg pronoun in SP and Rio, favors overt pro in both cities

| | 1st singular (weights) | 2nd singular *tu* | 2nd singular *você* | 3rd singular | 1st plural *nós* | 1st plural *a gente* | 2nd plural *vocês* | 3rd plural |
|---|---|---|---|---|---|---|---|---|
| SP | 0.483 | --- | 1.009*** | -0.195*** | -0.967*** | 0.920*** | 12.522 | -0.533*** |
| RIO | 0.521 | --- | 0.269*** | -0.538*** | -1.112*** | 0.927*** | 0.454 | -0.895*** |
| LIS | 0.547 | -0.241 | -1.235*** | -0.260** | -0.563*** | 2.236*** | 1.069 | -0.724*** |
| MAD | 0.525 | -1.209*** | -1.372** | 0.414*** | -0.781*** | 1.920*** | -0.467 | -0.798*** |

# i-conditioning: person/number – new pronoun *a gente* 'we'
- strongly favors overt pro in all communities, especially Portugal

|  | 1st singular (weights) | 2nd singular *tu* | 2nd singular *você* | 3rd singular | 1st plural *nós* | 1st plural *a gente* | 2nd plural *vocês* | 3rd plural |
|---|---|---|---|---|---|---|---|---|
| SP | 0.483 | --- | 1.009*** | -0.195*** | -0.967*** | 0.920*** | 12.522 | -0.533*** |
| RIO | 0.521 | --- | 0.269*** | -0.538*** | -1.112*** | 0.927*** | 0.454 | -0.895*** |
| LIS | 0.547 | -0.241 | -1.235*** | -0.260** | -0.563*** | 2.236*** | 1.069 | -0.724*** |
| MAD | 0.525 | -1.209*** | -1.372** | 0.414*** | -0.781*** | 1.920*** | -0.467 | -0.798*** |

NB: recently grammaticalized from NP 'the people'; takes 3rd sg agreement.

# Person/number puzzles

- Why should any one p/n form favor SPE more than others?
  - Diachronic variability and change
- Why such disparity between dialects in *você*?
  - Social attitudes about the T/V distinction
- Why is *a gente* so often overt?
  - Novelty?

# Person across time in Brazilian Portuguese: no consistent diachronic preference



Figure 2. OVERT SUBJECTS ACCORDING TO GRAMMATICAL PERSON (%)
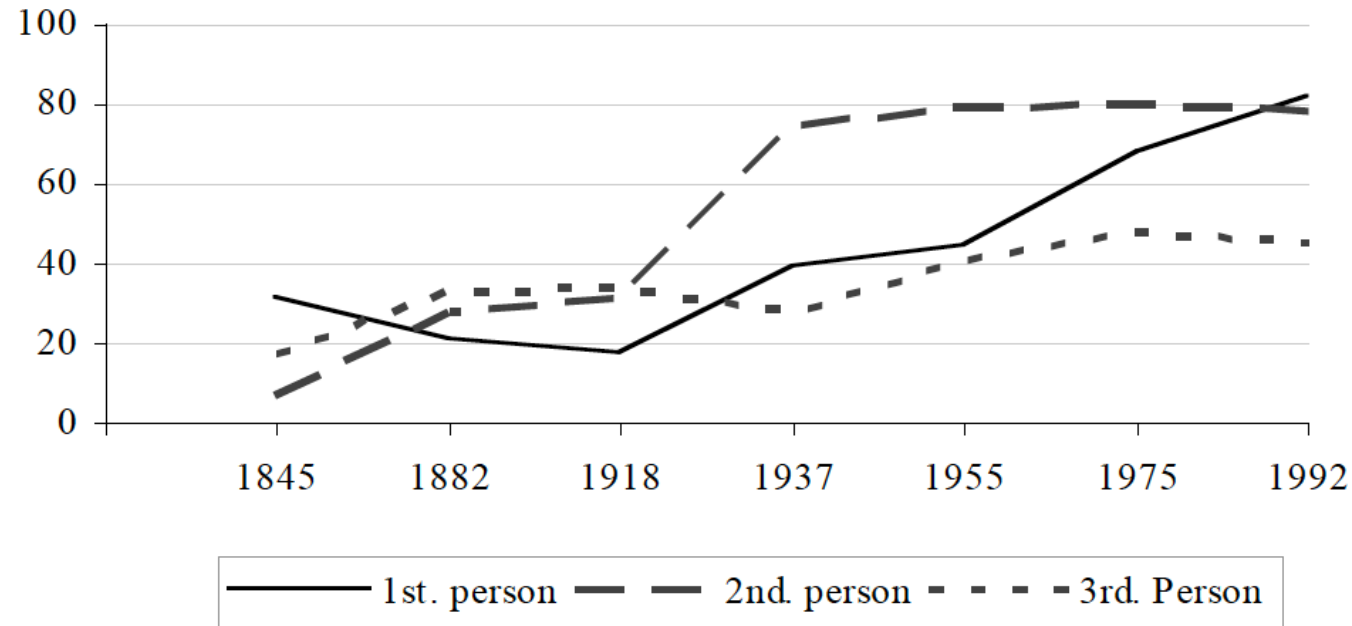
Barbosa, Duarte & Kato 2005

# Person/number puzzles: *você*

- Why does ***você*** favor SPE in Brazil and disfavor in Lisbon?

- Social intuitions of my Portuguese research assistants: a prevailing egalitarian ideology favors addressing everyone as ***tu***; the overt pronoun ***você*** is avoided, but a null pro with the accompanying 3sg verbal inflection sneaks under the radar. <span style="color:red">(s-condition?)</span>
  - We had only 16 tokens of overt *você* in the entire Lisbon corpus

- Note that the plural *vocês* does not compete with defunct *vos*, so is socially unmarked and favors overt pro.

# s-conditioning

- The social distribution of a variable should reflect or depend on social practice, norms, and identity expression.

- These are likely to be community-specific, culturally-specific.

- Significant social dimensions here:
  - Dialect: Portugal (two cities), vs. Brazil (two cities with substantial dialect differences)
  - Gender

  - Age and education not significant.

# Results:  s-conditioning

- Dialect: **huge difference between Portugal and Brazil**.  SPE rate is twice as high in Brazil.

- Gender: women have higher SPE rate than men; this is significant in SP and Lisbon

- Interaction between gender and age in Lisbon: gender difference is evident only in younger generation – under 45; i.e., born after the revolution of 1974 and establishment of democracy.
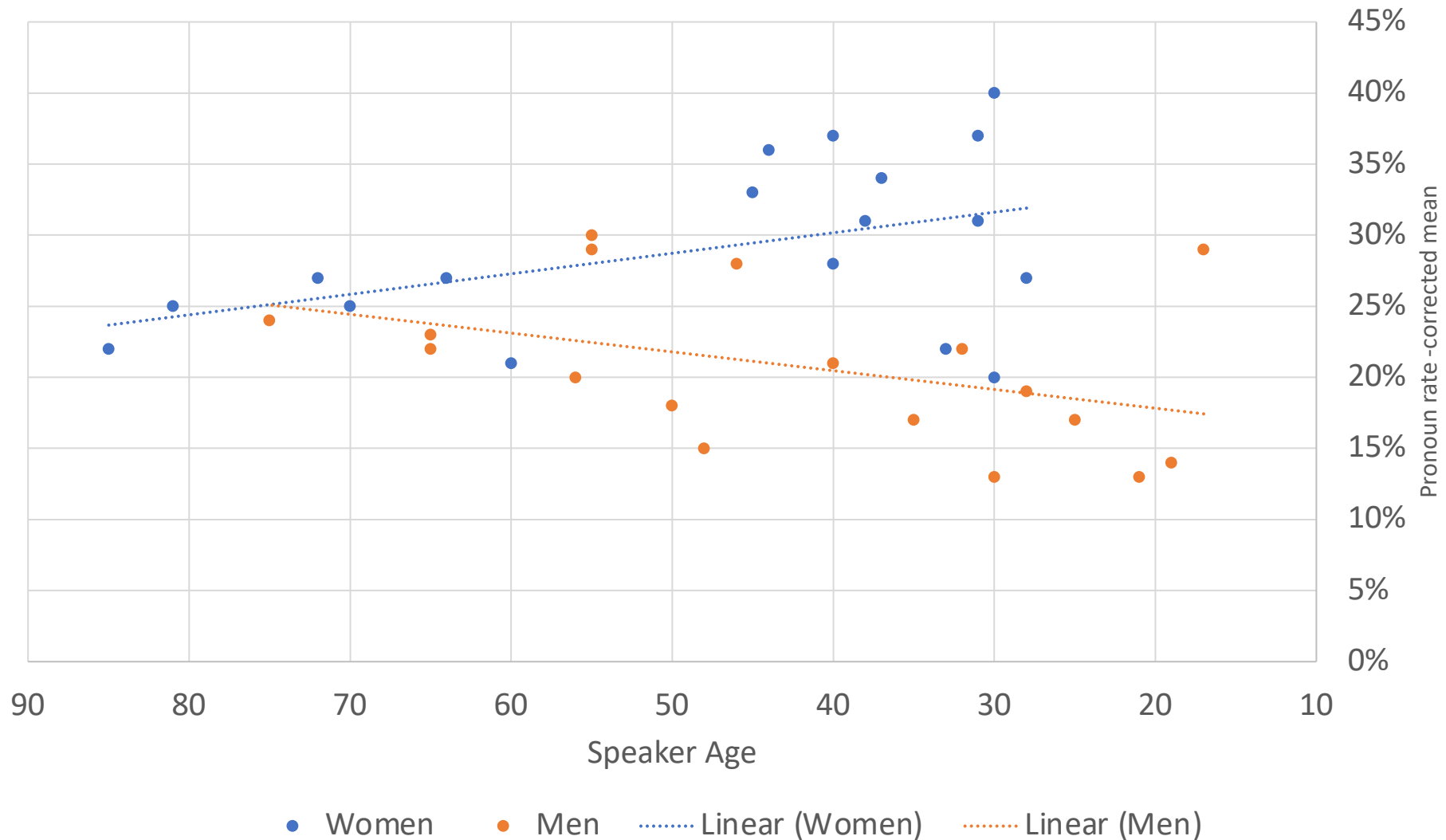
# s-conditioning: speaker gender women favor overt pro

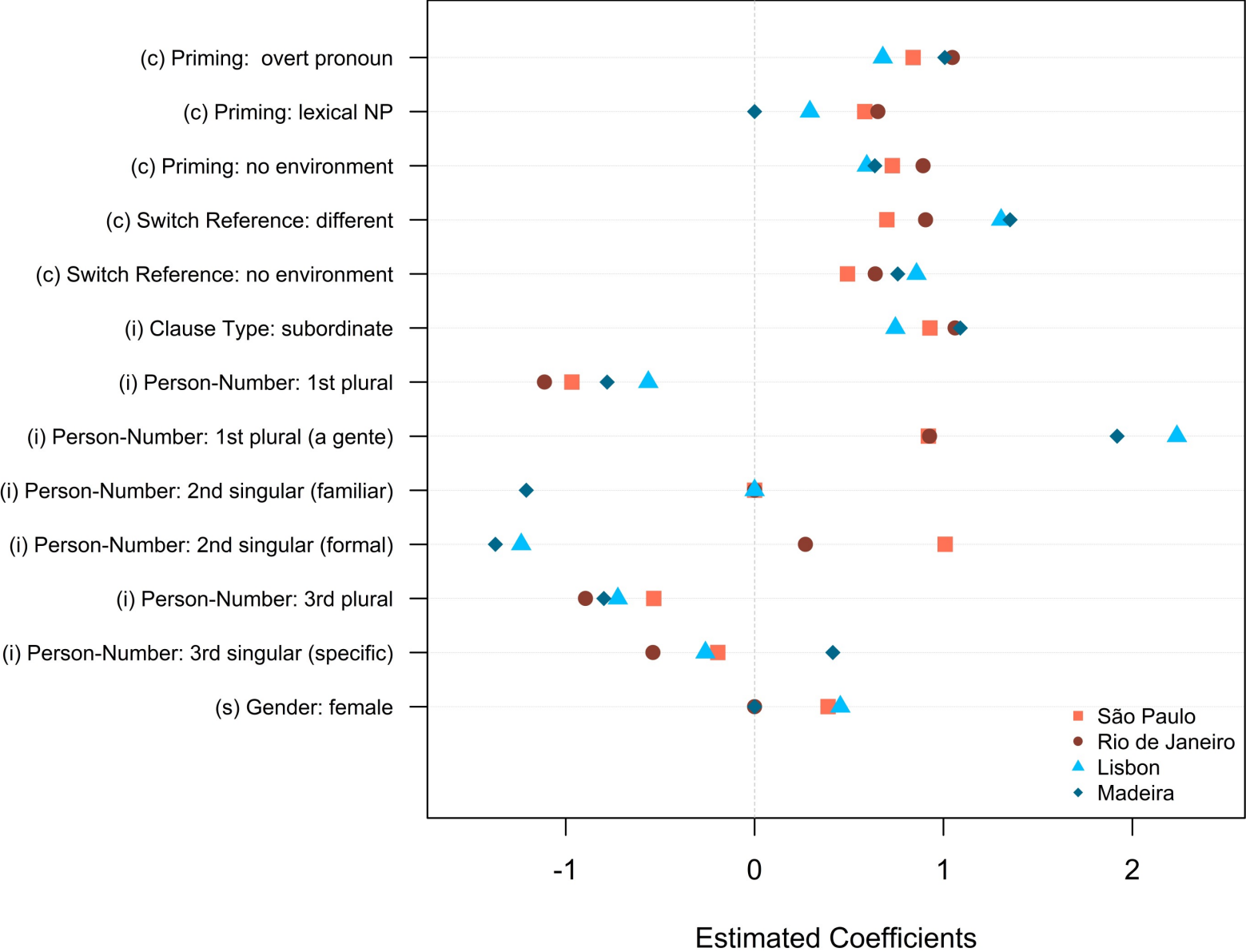| | Gender | | | | | | |
|---|---|---|---|---|---|---|---|
| | male | | | female | | | |
| | N | % overt | weight | N | % overt | weight | est. |
| SP | 7000 | 63.0 | 0.454 | 8498 | 70.1 | 0.538 | 0.389** |
| RIO | 4993 | 62.8 | 0.485 | 4677 | 65.7 | 0.516 | 0.108 |
| LIS | 4492 | 26.7 | 0.441 | 5107 | 36.8 | 0.552 | 0.454*** |
| MAD | 2103 | 33.6 | 0.495 | 2402 | 36.2 | 0.505 | 0.094 |
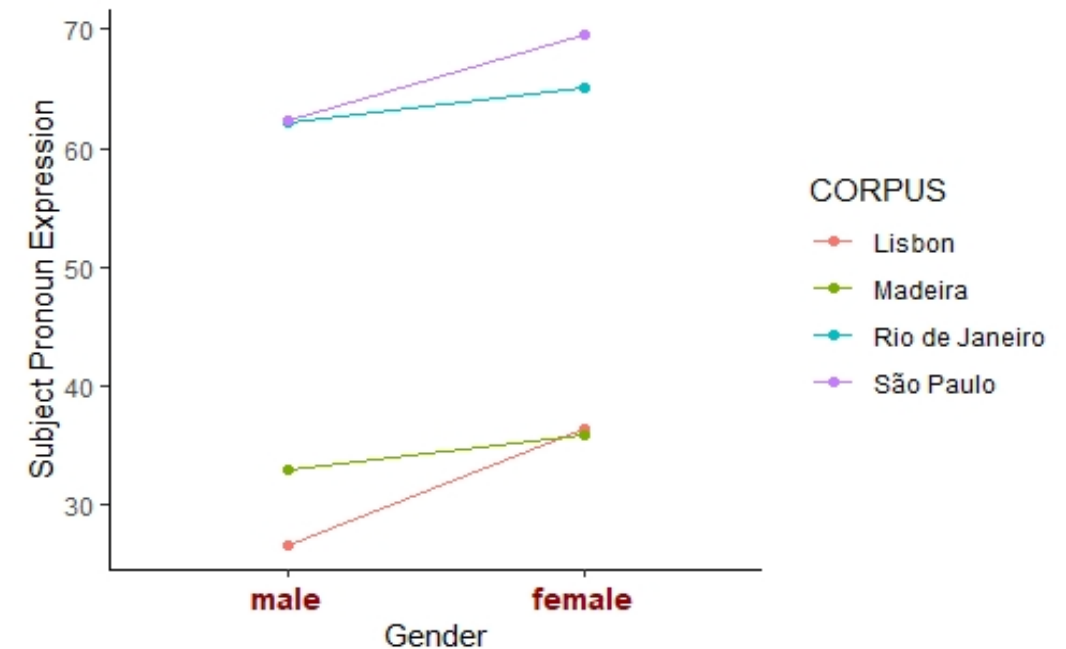
Lisbon: younger men and women diverge

# Summary

- p-conditions
  - effects are very similar across communities, consistent with Tamminga et al.
- i-conditions
  - clause type effect is very similar across communities. Maybe it's really a p-condition?
  - person/number constraint effects vary considerably across communities. This is consistent with Tamminga et al., but much of the disparity is related to social factors: attitudes toward T/V usage, novelty.
- s-conditions
  - gender difference is systematic, but only significant in two communities.
  - dialect difference between Portugal and Brazil is dramatic.
  - these disparities are consistent with Tamminga et. al.

**SPE in Portuguese - Relative Weight of Coefficients**

(c) Priming:  overt pronoun
(c) Priming: lexical NP
(c) Priming: no environment
(c) Switch Reference: different
(c) Switch Reference: no environment
(i) Clause Type: subordinate
(i) Person-Number: 1st plural
(i) Person-Number: 1st plural (a gente)
(i) Person-Number: 2nd singular (familiar)
(i) Person-Number: 2nd singular (formal)
(i) Person-Number: 3rd plural
(i) Person-Number: 3rd singular (specific)
(s) Gender: female

São Paulo
Rio de Janeiro
Lisbon
Madeira

Estimated Coefficients

# Potential confounds: i-, p- or s-conditioning?

- Portuguese has pronouns with social implications
  - second person: *você,* vs. *tu.*  These pronouns encode differences in social relationships, so their expression or non-expression may be socially motivated.
- These pronouns also confound with p-conditioning
  - *tu* takes second singular verbal inflection, *você* takes third singular and *vocês* takes third plural.  Hence null *você/vocês* creates greater ambiguity than null *tu*.
- *A gente* is new and has an extremely high rate of expression.  Why?
  - **Is novelty an i-, p- or s- condition?**

# Conclusions

- The constraint typology proposed by Tamminga, MacKenzie and Embick (the Penn model?) is mostly confirmed by these results…

- But the three categories may not be entirely discrete

- There is some ambiguity about the typological status of some constraints

- Some of the constraints interact across categories

# FIM

**Obrigado**

**("the end, thank you")**