Experimental evidence for 2SG direct object pronoun preferences in Brazilian Portuguese

Scott A. Schwenter[1]

Mark Hoff[2]

Kendra V. Dickinson[3]

Justin Bland[4]

Luana Lamberti[5]

ABSTRACT

We examine the competition in Brazilian Portuguese 2SG direct object pronoun expression between clitic *te* and tonic *você* (e.g. *Eu te vi ~ Eu vi você*). We offer data from an online forced-choice survey, analyzed using mixed-effects logistic regression, to show that dialectal subject pronoun preference (*tu/você*) and contrast both play a significant role in conditioning this choice. Furthermore, we find that contrast, despite its traditional treatment as binary, shows gradient effects on pronoun choice- while *te* is the preferred DO pronoun overall, *você* is the variant preferred in contrastive contexts, especially in cases of double contrast.

1   The Ohio State University. E-mail: schwenter.1@osu.edu.
2   The Ohio State University. E-mail: hoff.96@osu.edu.
3   The Ohio State University. E-mail: dickinson.130@osu.edu.
4   The Ohio State University. E-mail: bland.97@osu.edu.
5   The Ohio State University. E-mail: lambertinunes.1@osu.edu.

RESUMO

Examinamos a competição no português brasileiro entre o pronome clítico *te* e o tônico *você* para a expressão de objetos diretos de segunda pessoa singular (*Eu te vi ~ Eu vi você*). Apresentamos dados recolhidos de um questionário, e analisados utilizando a regressão logística de efeitos mistos, para mostrar que tanto a preferência do pronome de sujeito (*você/tu*) quanto o contraste têm um papel importante na escolha do pronome de objeto direto. Também, o contraste, apesar do tratamento tradicional como binário, mostra efeitos gradáveis na escolha de pronome- enquanto o *te* **é preferido** em termos globais, *você* é preferido em contextos contrastivos, especialmente nos casos de duplo contraste.

Palavras-chave: objetos diretos, contraste, português brasileiro, variação

1. Introduction

Personal pronouns in Brazilian Portuguese (BP) (and also European Portuguese; see Preto-Rodas 1972; Lara Bermejo 2017) are well-known for displaying a so-called *mistura de tratamento* ('address mixture') whereby distinct parts of the pronominal paradigm combine to create a system that "mixes" forms with different person/number combinations.[6] Perhaps the most famous of the many cases of mixture in BP is the use of the 2nd person singular (2SG) pronominal forms, which is characterized as "o uso simultâneo, por parte do falante, dos pronomes *tu* e *você* (e suas formas oblíquas e possessivas correspondentes, além do imperativo verbal) quando ele se dirige a uma segunda pessoa" (Bagno 2009: 237).[7] This case of mixture thus originated from the development and adoption of the grammatically 3rd person singular (3SG) form *você* (< *vossa mercê* 'your mercy'), and has been partially combined with the original paradigm corresponding to the 2SG pronoun *tu*. Despite considerable disdain from normative purists who deem this and other kinds of *mistura* unacceptable distortions of the language (see Bagno 2012: chapter 15 for ample discussion and a descriptivist

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista Linguística **260**

defense), the paradigmatic combination of pronominal forms from different sources is today not felt to be strange in the least to most Brazilian speakers; they simply consider *te* as a marker of direct or indirect objects (1, 2) that refers to the same referent that would be indicated by *você* when occurring as a subject or the object of a preposition (3, 4). We assume in (3, 4) that these are speakers who would use *você*, not *tu*, as their preferred subject pronoun, such as most speakers in the states of São Paulo or Minas Gerais.

(1)   Eu **te** vi ontem.   (Direct object)

'I saw you yesterday.'

(2)   Eu **te** dei esse livro como presente.   (Indirect object)

'I gave you that book as a present.'

(3)   **Você** come muito feijão.   (Subject)

'You eat a lot of beans.'

(4)   Essa casa é para **você**.   (Object of preposition)

'That house is for you.'

A particular site of pronominal mixture that has garnered nearly no attention from linguists can be found in direct object (DO) position for 2SG referents. While an example like that in (1) with the clitic pronoun form *te* is abundantly frequent in BP, it alternates with the use of the tonic, nominative form *você* as in (5):

(5)   Eu vi **você** ontem.

'I saw you yesterday'

In this paper, we employ experimental survey methods to investigate the alternation between the clitic form *te* and the tonic form *você* when they occur as direct objects, as in (1) and (5), respectively (see section 3 below for explanation of why we include only these two DO forms and no others). We

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista Linguística **261**

hypothesize that these two pronouns correspond to a heretofore unidentified instance of the clitic/weak vs. strong pronoun distinction in BP (cf. Cardinaletti & Starke 1999 for this distinction in Romance), and in particular that the strong pronoun *você* is actually the preferred choice in contrastive contexts, where there is contrastive focus on a direct object referent that competes with another referent. Thus, we find that the choice of 2nd person singular DO pronoun has a discourse-pragmatic basis that is strongly tied to functional motivations. And, unlike the typical understanding of the clitic/weak vs. strong pronoun distinction, which conceives of it as categorical, we find that the distribution of the two pronouns as DOs in both contrastive and non-contrastive contexts is probabilistic and constrained in predictable fashion by contextual factors.

The rest of the paper is structured as follows. In section 2 we review what grammars and the scarce prior research on the topic have said about DO *te/você* variation. Section 3 presents the methodology of our study. The results of our multivariate statistical analysis in R are offered in Section 4. Section 5 presents additional discussion of our findings, while Section 6 offers up some conclusions and directions for the future.

2. Background

In the literature on BP pronouns, when the possibility of using both *te* and *você* for direct object function is noted (sometimes the option of *você* as a DO is not even mentioned), they are presented simply as distinct variants without any functional differentiation between them. Indeed, for Perini (2010: 116), there is no difference whatsoever: "*Você* tem a forma oblíqua *te*, mas esta é usada em concorrência com a forma reta [*você*], de maneira que se pode dizer *eu te amo* ou *eu amo você*, indiferentemente."[8] In his English-language grammar of BP, Perini (2002: 100) espouses the same view: "the only case form intensively used nowadays in the spoken language is *te*, objective form of *você*; but even this competes with the regular form, so that one hears both *eu te amo* and *eu amo você*." Perini (2002: 385) goes on to say that these "two pronouns are used interchangeably, with no difference in meaning or degree of formality," and Azevedo (2005: 231) says in addition that the use of *você* for direct object function "is universal." Bagno (2012: 754) presents both options (along with *lhe* and the enclitic forms *-lo* and *-la*) as possible expressions of 2nd person singular DOs but does not offer any explanation as to how they might differ. Thomas (1974: 27) states explicitly that, "All the words used as subjects of verbs with the meaning *you* [...] may also be used as direct objects of

---

8   Translation: '*Você* has the oblique form *te*, but the latter is used in conjunction with the full form [*você*], so that one can say *eu te amo* or *eu amo você*, interchangeably.'

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*
Lingüística Revista **262**

a verb" and provides the example *Eu não compreendo você* 'I don't understand you' as illustration.[9] Finally, Whitlam (2011: 57) opines in a similar fashion, but appears to reverse the preferences: "In the spoken language [of Brazil], the 2SG object pronoun *te* may be used as an alternative to *você*." Overall, then, the received wisdom in the previous literature, while admittedly both scarce and terse, uniformly holds that there is no functional (or, for that matter, social) differentiation between *te* and *você* when used as DO pronouns.

The two hearts pictured below, found via a quick Google search, would seem to support these views, as would the selection of items in many souvenir shops in Rio de Janeiro, where both ways of expressing love on dolls and trinkets seem indeed to be found indiscriminately.

Figure 1. 2SG DO variation in the wild

While a few prior studies have compiled the frequencies of distinct forms for 2SG DOs in different varieties of BP (cf. Duarte & Ramos 2015: 186 for a brief summary), the only prior variationist study of the *te/você* alternation of which we are aware is Zampaulo (2014), who uses 20th century spoken data from the online Corpus do Português (Davies & Ferreira 2006; corpusdoportugues.org) in order to carry out his analysis. In terms of overall frequency (N = 559), Zampaulo found that 38% (n = 213) of 2nd person singular DOs correspond to *você*, while 62% (n = 346) are realized as *te*.[10] As we will show below, this rate is similar to the overall distribution of each form in our study. Zampaulo used Varbrul analysis to identify the significant constraints and their relative contribution to the variation between the two forms, and ultimately found that only one factor significantly affected the variation, namely, the number of verbs in the verb phrase. The clitic pronoun *te* was favored by simple VPs consisting of only one verb (e.g. *Eu te convido para a festa* 'I invite you to the party'), while *você*

---

9   This example with DO *você* is especially interesting, since it contains negation, which is typically considered in Portuguese-language grammars (e.g. Cunha and Cintra 2001) to be a proclitic trigger. Following the rules of proclisis, the example ought to have *te* instead of *você*.

10   Note however that, in order to extract this number of direct object tokens, Zampaulo (2014: 179) first needed to examine 14,409 occurrences of *você* and 3951 of *te*, i.e. over 18,000 occurrences of the two pronouns combined. The relative scarcity of these pronouns in direct object function in large corpora was one of our motivations for adopting an experimental approach to the variation.

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista Linguística **263**

was favored in more complex VPs made up of two (e.g. *Eu quero convidar você para a festa* 'I want to invite you to the party') or three verbs (e.g. *Eu deveria ter convidado você para a festa* 'I should have invited you to the party'). None of the other factors that Zampaulo included in his multivariate analysis (grammatical person, number, polarity, tense) turned out to be significant predictors of the choice between *te* and *você*.

As we will show below in the statistical analysis of our results, we were unable to replicate Zampaulo's (2014) finding of the importance of VP complexity for the choice of 2nd person DO pronoun. We believe that this lack of effect is due to several factors: our experimental data differ greatly from the corpus data analyzed by in Zampaulo's article; our hypotheses led us to test several factors that Zampaulo did not include in his study; and our analysis used mixed-effects logistic regression, while Zampaulo's only employed fixed effects. However, a comment made in passing at the end of Zampaulo's paper is one that has proven crucial for our research. Referring to the example in (6), he says (our translation from the original text in Spanish):

(6) a. Você convidou aquele seu amigo para a festa?

'Did you invite that friend of yours to the party?'

b. Não, eu não o convidei--eu convidei ***você***!

'No, I didn't invite him--I invited **you**!'

"While the use of *te* in place of *você* in this case would be considered ungrammatical, it is evident that the contrastive value of speaker B's sentence favors the tonic pronoun *você* instead of the atonic *te*" (Zampaulo 2014: 191).

We consider this example and the short accompanying explanation to be crucially important for studying the choice of 2SG DO pronoun in BP. As would be expected given its tonicity, the bisyllabic pronoun *você* lends itself to use in contrastive contexts more than the atonic, monosyllabic clitic form *te*, since contrastive pitch accent is possible in the first case but not in the second. Unlike Zampaulo, however, we do not consider the use of *te* in (6) to be ungrammatical, and the native speaker results we report below corroborate our view that *te* is still a possible option in this context. Note as well that although in European Portuguese it is possible (as in Spanish) to use clitic doubling in such an

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

264

example--see (7) as an alternate, yet truth-conditionally equivalent, response by B in the context of (6)--this option does not exist in BP, which does not permit clitic doubling in any context in the spoken language (we thank Patrícia Amaral for the EP example; note as well that the use of *você* as in [6] would be odd in EP, mainly because the use of this pronoun is are in Portugal):

(7)   Não, eu não o convidei--eu convidei-te **a ti**!

'No, I didn't invite him--I invited you!'

The clitic doubled sentence in (7) provides in the prepositional phrase *a ti* a grammatical structure that can host contrastive pitch accent; in fact, according to the native speakers consulted, the example would be odd in EP **without** clitic doubling and the contrastive pitch accent on *a ti*.
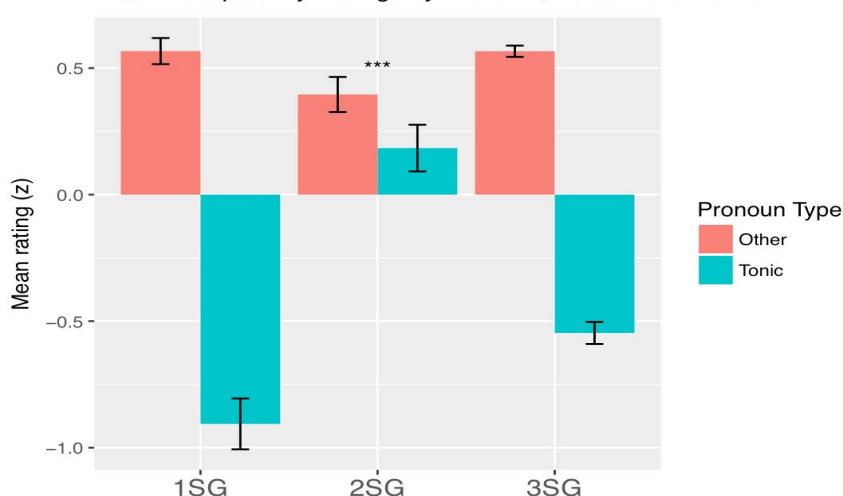
Looking outside of Portuguese to other languages with a clitic/weak vs. strong pronoun distinction, we find in the literature a hearty debate about the functions of these pronouns. In the original paper that proposed this distinction (Cardinaletti & Starke 1999), it is stated that the difference in function between these forms is linked to accessibility: clitics and weak pronouns refer to referents that are more accessible in the discourse model than those encoded by strong pronouns, which are used in turn to encode less accessible referents (e.g. to re-introduce referents that have been dormant in the discourse for a relatively longer period of time than those encoded by clitics or weak pronouns). However, Kaiser (2010) challenges this view in a study of what is termed the 3rd person "long pronoun" (*tema*) in Estonian in comparison with its "short pronoun" counterpart (*ta*). Using corpus data from Estonian and also parallel corpora in Finnish, Kaiser finds that the accessibility (or "salience") theory cannot account for the pronominal patterns. Rather, she finds that a more abstract notion of discourse contrast is at work, whereby the strong pronoun *tema* is employed preferentially in specifically contrastive contexts, and the weak pronoun *ta* in other contexts (i.e. it is the discourse-pragmatically unmarked form).

This finding is important for our research since, as 2nd person forms, *te* and *você* should not encode distinct degrees of discourse accessibility or salience of their referents. As part of the conversational dyad along with 1st persons, 2nd person referents (or their denotata) are always salient in the discourse model; they are not like 3rd person referents whose degree of accessibility can increase and decrease as the discourse unfolds. This leads us to hypothesize, following Kaiser's (2010) analysis of Estonian and also Zampaulo's (2014) brief mention for BP, that the clitic/strong pronoun

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista
Linguística 265

distinction found in *te/você* must be associated with the notion of contrast, and specifically that the strong form *você* is the one that will show greater association with contrastive contexts. In that sense, *você* is the marked form in discourse-pragmatic terms, while *te* is the form that is found elsewhere in unmarked discourse-pragmatic contexts. Note also that *você* displays other characteristics of strong pronouns (Cardinaletti and Starke 1999), for example, it can be coordinated (*Você/*Te e eu fomos para o cinema* 'You and I went to the movies') and it (but not *te*) can occur as the focused element of focus particles like *mesmo, até* (meaning 'even') or *só/somente* ('only'). As we will show in the results and discussion below, the weak/strong distinction is strongly upheld in our data, but at the same time it remains one that is probabilistic in nature: both forms can be selected in both contrastive and non-contrastive contexts, and either pronoun may be chosen by speakers in either context type.

Previous discussions of the perceived acceptability and social evaluation of object pronoun choice in BP have found that 2SG forms are evaluated differently than their 1st person singular (1SG) and 3SG counterparts. In their experimental survey of these forms, Lamberti et al. (2017) found that the canonical pronominal forms for 1SG (*me*) and 3SG (∅) DOs were rated as much more favorable than their tonic forms (*eu* and *ele/ela*, respectively). However, while *te* received a significantly higher acceptability rating overall, *você* was the only tonic form whose rating was above the mean of the normalized acceptability judgements. These results are illustrated in Figure 2 below, where "Other" refers to the canonical DO pronoun forms for each person-number combination: the clitics *me* (1SG; n=100) and *te* (2SG; n=100), and a null pronoun (3SG; n=500). "Tonic" refers to the corresponding non-canonical forms, i.e. the nominative forms used for DO function (*eu*, n=100; *você*, n=100; and *ele/ela*, n=500, respectively).

Figure 2. Normalized acceptability ratings for 1SG, 2SG, and 3SG pronominal forms (Lamberti et al. 2017)

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista Linguística **266**

As Figure 2 shows, while both 1SG and 3SG tonic pronouns are negatively evaluated by BP speakers, in the case of 2SG the tonic form receives positive evaluation just like its atonic counterpart. These results are critical to our own investigation, since they illustrate that in the choice between *te* vs. *você,* the less common tonic form *você* is not subject to the same degree of negative social evaluation as the non-canonical variants of other person/number combinations. Thus, the negative evaluation cannot be based strictly on the choice of tonic vs. clitic/null forms, since in the case of 2nd persons respondents did not rate *você* overall in negative fashion, while they did in the case of both *eu* as a 1SG DO pronoun and *ele/ela* as 3SG DO pronouns.

In the following section, we describe the experimental survey methodology we employed to test our hypotheses about the choice of *te* vs. *você* as DOs.

3. Methodology

3.1 Survey Methods

To examine the linguistic and social conditioning of *te/você* choice in BP, we created an online questionnaire hosted on Qualtrics. The link to the questionnaire was published on social media and distributed using the friend-of-a-friend method. Upon beginning the questionnaire, participants were presented with a brief description of the study's purpose, which was presented as an investigation of *a fala cotidiana dos brasileiros* 'Brazilians' everyday speech,' and given details about their rights as participants, pursuant to the study's IRB protocol.

Next, participants were asked a series of demographic questions in order to gather social data for later analysis and to ensure that only native speakers of BP over 18 years of age and who had lived most of their lives in Brazil were included in the sample. A total of 146 native speakers of BP, ranging in age from 18 to 74 years (mean age 31), completed the questionnaire. Table 1 reports the distribution of these participants by gender, level of education, socioeconomic status (SES), state of residence, and reported preferred 2SG subject pronoun. To arrive at this last factor, participants were asked, before beginning the target section of the survey, what pronoun they use among friends and were given the option of selecting *tu* or *você*.

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista
Lingüística 267

Table 1. Participant distribution by social factors

Gender	Education	SES	State of Residence	Subj. Pronoun
Female: 103 Male: 42 Other: 1	High school: 20 University: 52 Post-university: 74	Lower: 2 Lower-middle: 40 Middle: 80 Upper-middle: 23 Upper: 1	São Paulo: 82 Minas Gerais: 23 Rio de Janeiro: 12 Other: 29[11]	*tu*: 18 *você*: 128

Following the collection of these demographic data, participants were presented with a series of forced-choice questions with the instruction to choose the form they would employ in their everyday speech. A total of 15 target items and 15 filler items were presented to each participant, in random order. The order of the two choices for each question was also randomized for each participant to avoid bias.

An example of a target item is given in (8), and a sample filler item is presented in (9).

(8)   *Escolha a opção que você falaria:* 'Choose the option you would say:'

  ◯ *Eu estava te escutando.* 'I was listening to you (*te*).'

  ◯ *Eu estava escutando você.* 'I was listening to you (*você*).'

(9)   *Escolha a opção que você falaria:* 'Choose the option you would say:'

  A Maria comeu demais ontem e 'Maria ate too much yesterday and'

  ◯ *passou mal.* 'felt sick'

  ◯ *ela passou mal.* 'she felt sick.'

After responding to all 30 survey stimuli, participants were presented with a final, open-ended question where they were invited to provide additional comments about the questionnaire. The qualitative data obtained from this question, while interesting, will not be discussed in detail here due to space concerns. However, it is worth noting that several participants commented on the use of *você* to signal contrast, an important finding we discuss in greater detail in the sections to follow. Also interesting was some participants' perception, much like that of certain authors mentioned above, that *te* and *você* may be used interchangeably--we will have much to say on this misconception as well in the remainder of this paper.

---

11   Since the remaining states were sparsely represented, they were grouped together for statistical analysis, with Other including the following states in order of the number of participants living in each. Paraná: 5; Santa Catarina: 5; Bahia: 4; Ceará: 3; Rio Grande do Sul: 3; Fortaleza: 1; Goiás: 1; Pará: 1; Paraíba: 1; Pernambuco: 1; Piauí: 1. Three participants raised in but currently living outside of Brazil were also included in Other.

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Linguística **268**

We note that we did not include other 2SG variants besides *te* and *você*, such *lhe*, *cê*, or *ocê* as options in the survey. These forms are not necessarily found in the speech of the respondents while both *te* and *você* are and are recognized as competing variants (at least for those who chose *você* as their usual subject pronoun). In addition, it is not clear in many cases whether BP speakers are choosing *lhe* for 2SG DOs as a kind of "intermediate" option in terms of formality, or even as the clitic corresponding to formal *o senhor/a senhora* (Monteiro 1994: 86), and our interest here is strictly in the informal 2SG forms. As regards the reduced forms of *você* (*cê, ocê*), which have been the focus of some research in BP (Petersen 2008; Ramos 1996; Vitral 1996), these are mostly limited to subject function and they have also been mainly associated with certain regions of Brazil (e.g. Minas Gerais). Moreover, it was unclear to us how familiar respondents would be with the orthographic rendering of these forms on the survey. Nevertheless, we recognize that including these forms in future research may provide additional nuance to our analysis.

3.2 Statistical Methods

Statistical analysis was performed in R (R Core Team 2017) using the lme4 package (Bates et al. 2015). We constructed five mixed-effects logistic regression models on relevant subsets of the data. In each model, we only included factors and factor levels that were fully crossed, i.e. where every level of one factor co-occurred with every level of each other factor. For example, all of the items in our dataset with three verbs had positive polarity, so we excluded three-verb constructions from models that examined positive and negative polarity.

The first model investigated the effect of the number of verbs (1, 2), polarity (negative, positive), and contrast (contrastive, non-contrastive) on object pronoun choice (*te*, *você*). The second model looked more closely at the effect of number of verbs by including items with 1, 2, or 3 verbs. The third model examined the effect of different negative polarity items (*não* 'no', *ninguém* 'nobody') and number of verbs (1, 2) on object pronoun choice. The fourth model investigated whether object pronoun selection is sensitive to the prescriptive prohibition against sentence-initial clitics by comparing a context with sentence-initial *te* (e.g. *Te vi ontem* 'I saw you yesterday') and a context with non-initial *te* (e.g. *Eu te vi* 'I saw you'). The fifth and final model investigated the effect of contrast types. We compared an instance of direct object contrast, i.e. single contrast (e.g. *Eu vi a Luana, mas queria te ver/mas queria ver você* 'I saw Luana, but I wanted to see you'), where the main verb in each clause was the same, with an instance of direct object and verbal contrast, i.e. double contrast (e.g. *Eu vi a Luana, mas te procurava/mas procurava você* 'I saw Luana, but I was looking for you'), where the

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista Linguística **269**

main verb was different by clause.

For all of the models above, we also considered the following external predictors: gender, level of education, SES, state of residence, and reported subject pronoun (*tu, você*). All of these predictors were self-reported by our respondents. While none of these external factors were selected as significant in any of the models, this may be due in part to uneven participant distribution in the sample. Thus, this finding should therefore be interpreted with some caution. In the remainder of the paper our focus will be on the linguistic, rather than social conditioning of *te/você* choice. We asked respondents to provide both their state of residence and reported subject pronoun in order to assess whether the previously attested relationship between subject pronoun choice and region (Monteiros 1994) was borne out in our data. Indeed, we found these two factors to be non-orthogonal (i.e. there was strong collinearity between subject pronoun choice and respondent region). Therefore, we included only reported subject pronoun in our final models since the output of the random forest function in R (Hothorn et al. 2006) revealed it to be the stronger of the two predictors. In all models, random intercepts were included for respondent in order to account for repeated measures. For each of our five models, we built multiple models using a step-up method and compared them with ANOVAs to determine the best-fit model.

4. Results

Overall, our data (N=2119) show that *te* (67.7%, n=1434) is selected as an object pronoun just over twice as often as *você* (32.3%, n=685). This is one particular site in the pronominal grammar of BP, therefore, where the clitic form remains more frequent than the tonic form, as opposed to, for instance, 3rd person DOs, where the clitics are nearly extinct in conversation (Schwenter & Silva 2003). Still, a fine-grained analysis is needed to further investigate the constraints on this variation between the 2SG forms. Descriptions of the variation and the results of our mixed-effects logistic regression analyses follow.

Model 1: Number of verbs, polarity, and contrast

Model 1 examined the effects of number of verbs, polarity, and contrast on object pronoun choice. We also considered reported subject pronoun and the social factors described above as predictors.

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Lingüística **270**

It is important to note that all instances of contrastive focus with positive polarity also included the contrastive conjunction *mas* 'but' (10) in our stimuli, while none of those with negative polarity contained *mas* (11). For this reason, it is not possible to tease apart the effect of polarity from the effect of null vs. overt conjunction in contrastive contexts in our stimuli. In the following model, we acknowledge the possibility that either polarity or conjunction type may be responsible for the effect labeled "Polarity" shown below.

(10)   Eu vi a Luana, **mas** queria te ver/mas queria ver você.

         'I saw Luana, but I wanted to see you.'

(11)   Eu vi a Luana, **Ø** eu não te vi/eu não vi você.

         'I saw Luana, I didn't see you.'

First, we note that *te* is preferred overall both among survey respondents who report using subject pronoun *tu* and those who use *você.* However, this preference is much stronger for *tu* users (93.3% *te,* n=251) than for those who use *você* as their main subject pronoun (64% *te,* n=1183). It is somewhat surprising to note that some *tu* users nevertheless chose *você* in DO contexts (6.7%).

Figure 3. Survey results for object pronoun choice by reported subject pronoun
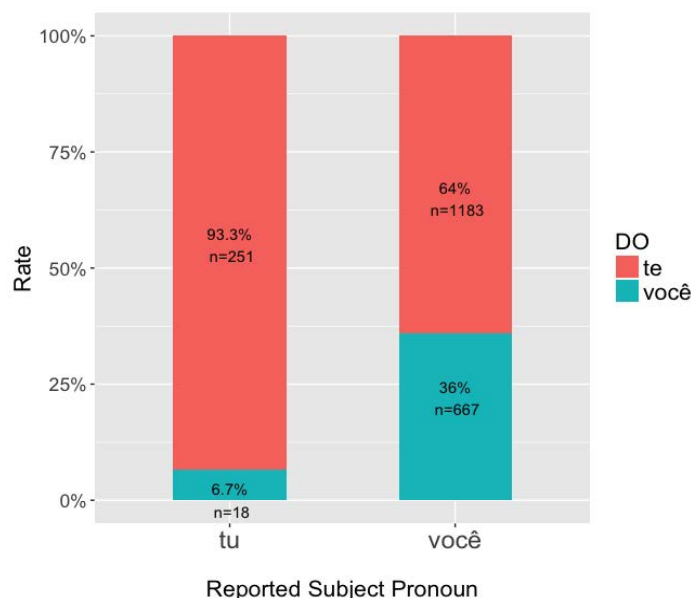
Second, we note a significant difference in object pronoun choice between contrastive and non-contrastive contexts (*p*<2e-16). In non-contrastive contexts, *te* is the default selection, but in

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*
Revista Linguística 271

contrastive contexts, *você* is selected at a rate of 52.1%, i.e. more often than *te* (47.9%). As we show below, however, the type of contrastive context is crucial to this choice of *você* over *te*.

Figure 4. Survey results for object pronoun choice by context type
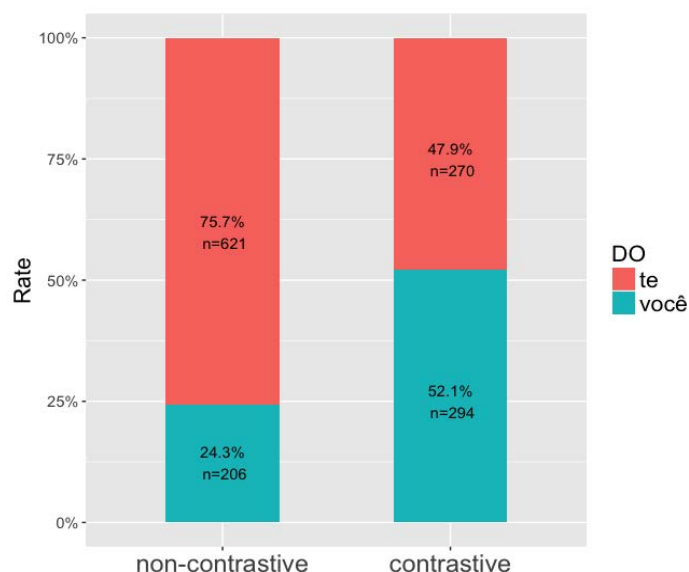
To build a model for these data, we used a step-up function followed by an ANOVA comparison to determine which factors to include in the statistical model. These results suggested that the best-fit model would include the predictors of contrast, reported subject pronoun, and polarity, and that the remaining predictors of age, SES, polarity, gender, level of education, and number of verbs did not contribute enough to variation in the dependent variable to be included in the best-fit model.
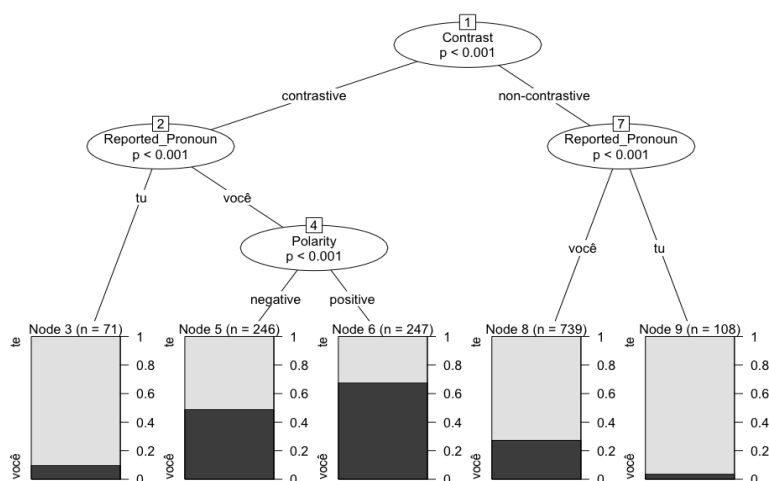
We ran a mixed-effects logistic regression model using these three predictors and random intercepts for respondent. The results of the logistic regression for model 1 are presented in Table 2. All three predictors of contrast, reported subject pronoun, and polarity were significant. In this subset of the data, participants chose *te* (65.7%, n=372) more often than *você* (34.3%, n=194) as a direct object; however, speakers who reported using subject *você* chose object *você* significantly more often than those who reported using subject *tu* (p=3.46e-08). Additionally, participants chose *você* significantly more often in sentences with contrastive focus (p<2e-16) and in sentences with positive polarity (p=0.00141).

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista
Lingüíſtica **272**

Table 2. Logistic regression results for model 1

		Estimate	Std. Error	Z value	Pr(>|z|)
	Intercept	-4.6541	0.6171	-7.524	4.63e-14***
Contrast (Reference level: Non-contrastive)	Contrastive (n=564; 40%)	1.9646	0.1628	12.068	<2e-16***
Reported subject pronoun (Reference level: Tu)	Você (n=1232; 87.3%)	3.4332	0.6224	5.516	3.46e-08***
Polarity (Reference level: Positive)	Negative (n=566; 40.1%)	-0.4837	0.1516	-3.193	0.00141**

In order to explore these effects more thoroughly, we built a conditional inference tree for model 1 using the party package in R (Hothorn et al. 2006). The output of the conditional inference tree, in concert with the output of the logistic regression model, shows that contrast is the primary predictor of pronoun choice in these data. Overall, respondents preferred *te* over *você* as an object pronoun and also preferred *te* in non-contrastive contexts. However, those who reported use of *você* as a subject pronoun were less likely to select *te* as an object pronoun in non-contrastive contexts than *tu* users. In contrastive contexts, *tu* subject pronoun users continued to prefer *te* as an object pronoun overall, but selected *você* more often than they did in non-contrastive contexts. Finally, in contrastive contexts with negative polarity, *você* users selected *você* and *te* as an object pronoun at nearly the same rate. However, in contrastive contexts with positive polarity, *você* users demonstrated a true preference for *você* as the object pronoun.

Figure 5. Conditional inference tree for model 1

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista Linguística **273**

Model 2: Number of verbs

The second model explored further the effect of number of verbs on object pronoun choice, in order to test our results against those of Zampaulo (2014). Model 2 differs from model 1 in that we included sentences with three verbs as well as sentences with one or two verbs, and we only considered sentences with positive polarity and without contrastive focus. As above, reported subject pronoun and social factors were included as possible predictors.

The step-up function and ANOVA comparison for model 2 suggested that the best-fit model would only include reported subject pronoun as a predictor. That is, the remaining predictors, including number of verbs, are likely not significant. It was not entirely unexpected that number of verbs would be excluded from the best-fit model for model 2, given that it was not included in model 1. Since the purpose of model 2 was to investigate the effect of number of verbs on object pronoun choice, and that factor was shown to not likely to be significant, it became unnecessary to run the final model.

Model 3: Negative polarity

Model 3 further explored the effect of negative polarity on object pronoun choice by testing whether the specific negative indefinites *não* and *ninguém* have different effects on the selection of *você* or *te*. For this model, we also considered number of verbs (1, 2), reported subject pronoun, and social factors as possible predictors. Only non-contrastive sentences were considered in this model.

The results from the step-up function and ANOVA comparison were the same as for model 2: the only factor included in the best-fit model was reported subject pronoun. Neither negative item (*não, ninguém*) nor number of verbs (1, 2) were included in the best-fit model. Therefore, the use of a negative item does not appear to influence object pronoun choice in non-contrastive sentences. It is worth noting that, in the conditional inference tree for model 1, polarity only influenced object pronoun choice in contrastive sentences for users of the subject pronoun *você,* and model 3 only considered non-contrastive sentences. Therefore, further investigation of the influence of different polarity items in contexts with contrastive focus would be a fruitful direction for future research.

Model 4: Clitic placement

Model 4 investigated the effect of clitic placement on object pronoun choice. Prescriptive rules

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*
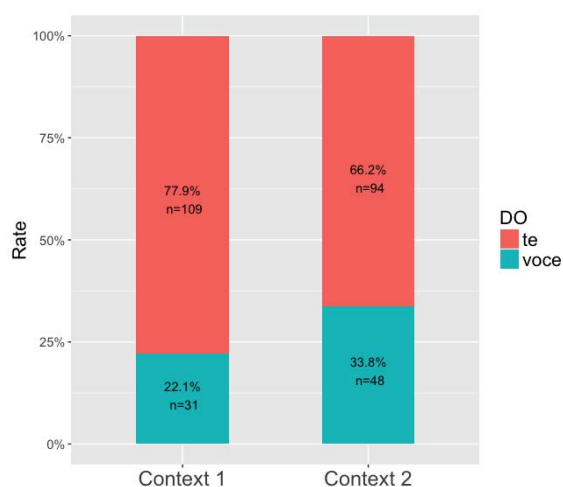
Revista
Lingüíſtica **274**

in Portuguese prohibit the use of clitic pronouns in sentence-initial position. Therefore, this model aimed to assess whether speakers showed an increased preference for the tonic pronoun *você* when the alternative is *te* as the first element in the sentence. For this model, we considered two specific survey items:

(12)   Eu te vi. / Eu vi você.

         'I saw you.'

(13)   Te vi ontem. / Vi você ontem.

         'I saw you yesterday.'

In Context 1, as in (12), the presence of an overt subject pronoun means that even if *te* is selected as the 2SG DO, it will not occur in sentence-initial position. However, in Context 2, as in (13), selecting *te* results in a sentence-initial *te*. The question addressed in model 4 is whether *você* is selected more often in Context 2 than in Context 1 because of prescriptive pressures to avoid sentence-initial clitics. DO pronoun selection results for the items considered in model 4 are displayed in Figure 6.

First, we note that object pronoun selection for these two survey items is similar to other subsets of the data, in that *te* is preferred overall (72% *te*, n=203). Additionally, respondents who use subject pronoun *você* selected more object *você* (30.9% *você,* n=76) than subject pronoun *tu* users did (8.3% *você,* n=3). Furthermore, we also observe that *você* is more likely to be selected when it will allow the respondent to avoid sentence-initial *te* (Context 2, 33.5% *você*, n=94) than when the other choice does not include sentence-initial *te* (Context 1, 22.1% *você*, n=31).

Figure 6. Survey results for items considered in model 4

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

275

To build a model for these data, we used a step-up function followed by an ANOVA comparison to determine which factors to include in the statistical model. The output of these functions suggested that the best-fit model includes reported subject pronoun (*tu*, *você*) and survey item (Context 1, Context 2) as predictors. The results of the logistic regression for model 4 are displayed in Table 3.

Table 3. Logistic regression results for model 4

		Estimate	Std. Error	Z value	Pr(>|z|)
	Intercept	-4.4718	1.1301	-3.3957	7.6e-05***`
Reported Subject Pronoun (Reference level: Tu)	Você (n=236; 87.3%)	2.4147	0.9944	2.428	0.01517*
Item (Reference level: Context 1)	Context 2 (n=282; 50%)	1.0327	0.3876	2.664	0.00772**

As in our previous models, the output of this model shows that respondents preferred the object pronoun *te* over *você* overall. Additionally, in both Contexts 1 and 2, respondents who reported using subject *você* were significantly more likely to select object *você* than those who reported using subject *tu* (p=0.01517). Finally, these results also show that, although respondents preferred *te* overall, they were significantly more likely to select the object pronoun *você* in Context 2—where selecting *te* would result in a sentence-initial clitic—than they were in Context 1 (p=0.00772). This effect is illustrated in Figure 6 above and confirms our hypothesis that the prescriptive norm against sentence-initial clitics can influence speakers to choose *você* rather than sentence-initial *te*. The strength of this norm, however, is undoubtedly much less than its proponents would like it to be.

Model 5: Single vs. double contrast

Model 5 investigated the effect of contrast type on DO choice. Single contrast contexts vary along only one dimension, as in (14) where only the DO differs across the clauses. The contrast in this case is between who the speaker wanted to see ("you") and who she actually saw ("Luana"); the lexical verb remains the same in both cases (*ver*). In double contrast contexts, the contrast varies along two dimensions, as in (15) where both the DO and the lexical verb vary across the clauses. In this case, the speaker "saw Luana," but was "looking for you." Thus, in this second case, the referents of the DO and also the verbs of the respective conjuncts (*ver* versus *procurar*) show contrast.

For this model, we considered two specific survey items:

(14)   Eu vi a Luana, *mas queria te ver/mas queria ver você.*

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*
Revista Linguíʃtica **276**

'I saw you Luana, but I wanted to see you.'

(15)   Eu vi a Luana, *mas te procurava/mas procurava você.*

'I saw you Luana, but I was looking for you.'

Here, we note that object pronoun selection varies according to contrast type. In situations of single contrast, as in (14), *te* and *você* are selected equally by survey respondents. However, in situations of double contrast, like that in (15), survey respondents show a notable preference for *você*, as can be seen in Figure 7.

Figure 7. Survey results for items considered in model 5
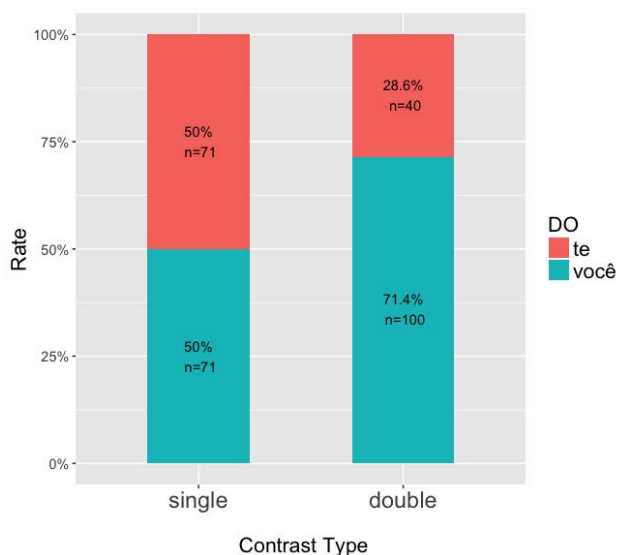
To build a model for these data, we employed a step-up function and ANOVA comparison. The output of these functions showed that the best-fit model for this subset of the data includes reported subject pronoun (*tu*, *você*) and contrast type (single, double) as predictors. The results of the logistic regression for model 5 are displayed in Table 4.

Table 4. Logistic regression results for model 5

		Estimate	**Std. Error**	**Z value**	**Pr(>|z|)**
	Intercept	*-2.1197*	*0.7643*	*-2.773*	*0.0055***
Reported Subject Pronoun (Reference level: Tu)	Você (n=247; 87.6%)	2.9724	0.9072	4.379	1.19e-05***
Contrast type (Reference level: double)	Single (n=142; 50.4%)	-1.4678	0.3735	-3.930	8.50e-05***

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*
Revista Linguíʃtica **277**

The results of this model show that there is a significant difference between single (DO only) and double contrast (DO and lexical verb) contexts. The latter context-type results in significantly more respondent choice of *você* (72.5% versus only 28.5% *te*, n=140) than in the former, where the selection of the two forms is exactly 50% for each (n=142). Although this result needs to be verified with more data and distinct examples, it strongly suggests that the mapping between DO form and contrast is not a binary category, but rather a gradient one. The greater the contrast between the two conjuncts, the more likely speakers are to choose *você* over *te*.

To conclude this section, we note that, for each of the previous models, we performed the same procedures as above for the subset of these data that included only subject pronoun *você* users, in order to investigate the possibility of other significant predictors in the subset that were potentially masked by including both *você* and *tu* users at the same time. However, the output of these models showed the same constraints as the models which included both *tu* and *você* users together.

5. Discussion

The results presented in the previous section show unequivocally that the choice between *te* and *você* in DO function is neither random nor *livre* ('free'). In fact, the five models described above show that this variation is easily explained in terms of contrast, negation, and what is most likely the prescriptively-motivated avoidance of sentence-initial *te*. Additionally, subject pronoun preference was shown to significantly constrain DO choice, with users of *você* as a subject pronoun also selecting *você* as a DO at higher rates across contexts than those who chose *tu* as their preferred subject pronoun. However, despite differences in overall rates, the effects of the significant linguistic constraints found in our analysis have the same directionality for both *você* and *tu* subject pronoun users.
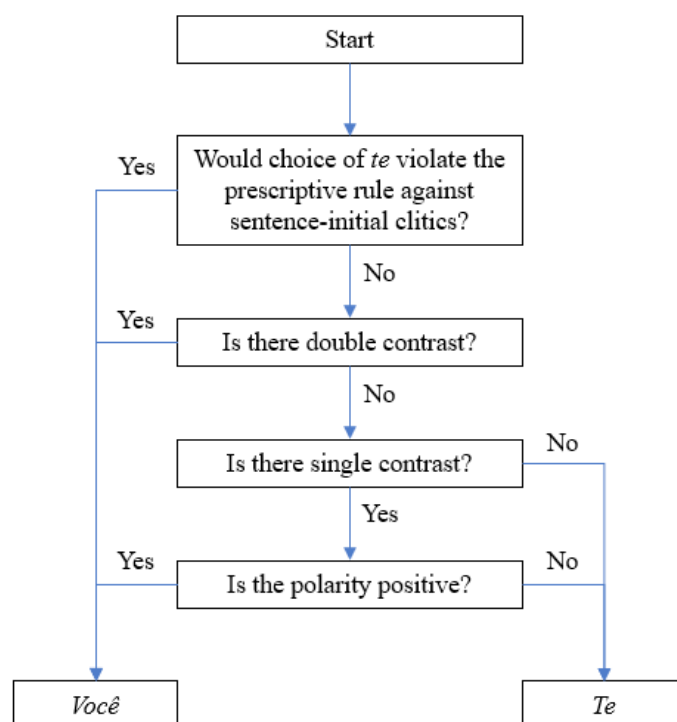
Model 1 showed that *te* is the default DO form in non-contrastive contexts and that this preference is strongest for users whose subject pronoun is *tu*. Thus, in terms of pronominal markedness (Bresnan 2001), the clitic form is the more unmarked form overall, but the strength of its unmarked status is stronger for subject *tu* than for subject *você* speakers. *Você* on the other hand is chosen significantly more often than *te* in contrastive cases of positive polarity, while *te* is the preferred form in cases of negative polarity. This is likely due to the fact that *não* is traditionally considered to form part of the class of proclisis triggers (Cunha and Cintra 2001; Martins 1993; Washington 2015) and *te* is the proclitic variant for 2SG DO pronouns. Unsurprisingly, given the function of *você* as marking

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*
Revista Linguística **278**

contrast and the proclitic-triggering properties of negation, *te* and *você* are chosen at nearly identical rates in negative-polarity contrastive contexts. Finally, as shown in model 3, negation does not have significant effects in non-contrastive contexts where *te* is already the preferred pronoun, nor does the specific negative polarity item (*não* vs. *ninguém*) significantly affect *te/você* choice. Finally, while *te* and *você* are chosen at exactly the same rates when all contexts of single contrast are considered, *você* is chosen more often in cases of double contrast.

In addition to contrast and negation, pressures related to standard language ideology also appear to play a role in speakers' choice of DO forms. The avoidance (albeit only partial) of sentence-initial clitics was shown in model 4 to lead to more selection of *você* in contexts where the choice of *te* would violate prescriptive norms. Model 2 showed that, despite Zampaulo's (2014) findings from corpus data, the number of verbs was not a significant predictor of *te* vs. *você* choice.

Based on the results of these models, then, the decision-making process behind speakers' choice of *te* vs. *você* (most strongly, of course, for speakers who use *você* as their default subject pronoun) can be modeled as a flowchart as below. While the model is probabilistic, not deterministic (cf. Estigarribia 2013:139 for similar comments on clitic doubling in Rioplatense Spanish), it accounts for the major patterns found in our survey results.

Figure 8. Decision-making process for te/você selection.

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista Linguística **279**

Our findings regarding the rates of *te/você* choice in DO function lend themselves to fruitful comparison with a similar process of pronominal selection in so-called *sujeito acusativo* ('accusative subject') constructions. These are cases where *te/você* functions at the same time as the DO of the first verb and the subject of the following one, which may appear either in its infinitival (parallel to English *I saw you/him leave*) or progressive forms (as in English *I saw you/him leaving*) (cf. Bagno 2001, 2009). Examples (16-19) taken from the *Corpus do português* show the use of both *te* and *você* in this construction, as well as the possibility of either the infinitive or progressive forms in the second verb, in this case *fazer* 'to do.'

(16) *Antes eu já **te** vi **fazer** o mesmo com outros.*

'I've already seen you do the same thing to other people.'

(17) *Já **te** vi **fazendo** muito isso.*

'I've already seen you doing that a lot.'

(18) *Antes eu já vi **você fazer** isso no blog muitas vezes.*

'I've seen you do that on the blog a lot of times before.'

(19) *Eu já vi **você fazendo** coisas incríveis.*

'I've already seen you doing incredible things.'

According to prescriptive grammars (e.g. Bechara 1999; see Bagno 2001, 2009 for an overview and descriptivist response), the accusative pronoun *te* is the normative variant in such contexts. However, despite prescriptive norms, the use of subject pronouns or *pronomes retos* such as *eu, você, ele/ela,* etc. is claimed to be more common in this construction in colloquial BP than use of their standard counterparts. In fact, Bagno (2001: 111) describes the use of *eu* in *Deixa eu dizer o que penso disso* ('Let me say what I think about that'), instead of the clitic version *Deixa-me dizer…*, as "following the practice of the vast majority of Brazilian Portuguese speakers, even the highly educated" (our translation). In Figure 9 below, we examine the frequency of the forms *te* and *você*

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*
280

specifically in the *Corpus do português* to test such claims.[12]

Figure 9. Relative frequency of te and você in (infinitival) accusative subject constructions[13]
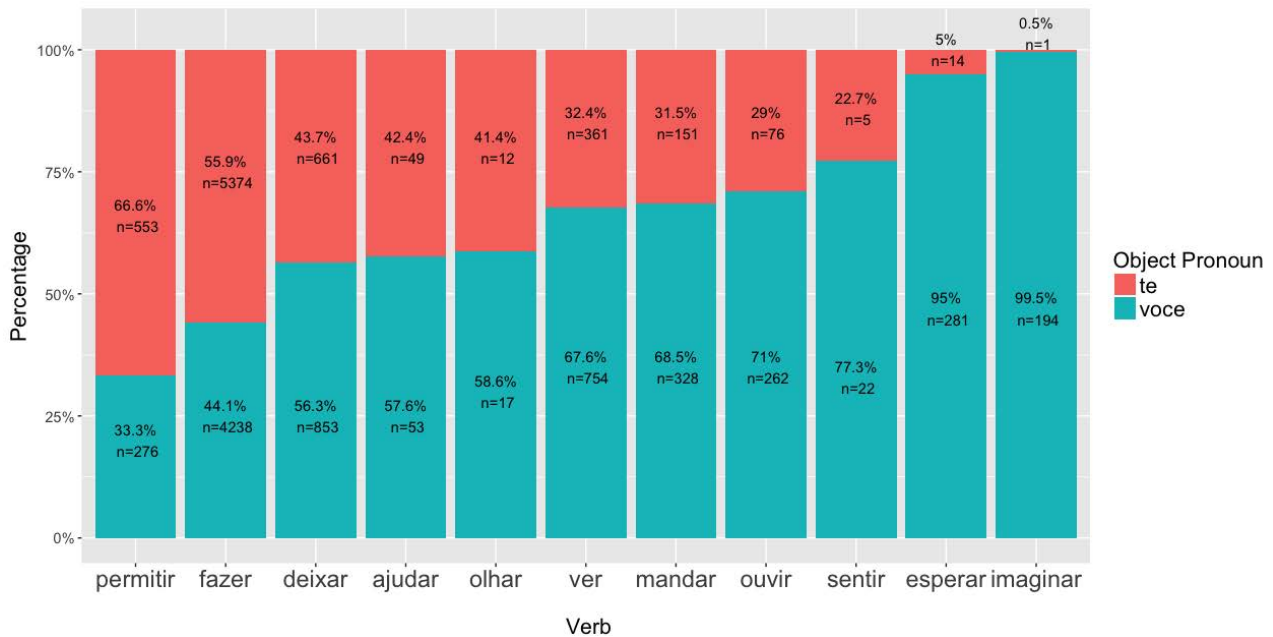
Overall, *te* is very slightly more frequent than *você* in that it is selected 50.2% (7247/14430) of the time. However, as Figure 9 makes apparent, the rate of *te/você* selection is highly variable from one main-clause verb to the next, and the number of tokens corresponding to each verb also varies tremendously. For example, the high rate of *te* selection with *fazer* (55.9%; 5374/9612) suggests that this verb has a much higher degree of entrenchment with *te* than the other verbs, especially since this verb alone makes up 67% (9612/14430) of all of the tokens examined. If *fazer* is removed from the sample, *você* easily becomes the more frequent variant, occurring in 61.1% (2945/4818) of the remaining cases. As the figure reveals, fully nine of the 11 verbs we investigated show more use of *você* than *te* with infinitival constructions, and all of these verbs have a lower frequency than *fazer* (the only other verb that does not follow the dominant pattern with *você* is *permitir*).

With respect to the progressive forms, we performed a similar search for the verbs *deixar, imaginar, ouvir, querer, sentir, ter,* and *ver*.[14]

_____

12   In fact, as Bagno (2001: 112) points out, certain collocations with the subject pronoun such as *deixa eu ver* 'let me see' are so frequent that they are commonly highly reduced (e.g. to *xovê*).

13   The syntax used for the corpus search is as follows. For *te: te [fazer] _v**. For *você: [fazer] você _v**. Brackets indicate that all forms of the verb inside are included in the search, and *_v** returns results for any verb in that position. Furthermore, our search included data only from Brazilian, rather than European Portuguese, and the alternate spellings *voce* and *vc* were included in the count for *você*.

14   For both the infinitival and progressive constructions, there is disagreement among grammars about the

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

281

As in Figure 9 above, the rate of selection of *te* vs. *você* varies appreciably from one verb to another; for example, *você* is selected 92.8% (231/249) of the time with *imaginar* and 66.5% of the time with *ver*, but only in 32.3% (51/158) of cases with *deixar*. Considering all verbs together, however, *você* is clearly more frequent than *te* in progressive constructions with a 2nd person subject, appearing in 67.7% (1091/1611) of all cases, with *te* chosen at a rate of only 32.3% (520/1611). Both of these patterns can be assumed to have an impact on the choice of *você* as the DO pronoun in other contexts, including non-contrastive contexts where there is no necessary discourse-pragmatic motivation behind its choice, since in terms of surface form *você* occurs after the finite verb in both cases. Notice as well that when *você* occurs as a *sujeito acusativo* with both infinitives and gerunds, it necessarily occurs in a switch reference context, since the subject of the main clause is always distinct from the 2SG subject of the subordinate clause (as in examples [16-19] above, where the subject of the main clause is 1SG). Given that contrastive contexts with DOs such as those we have investigated are necessarily also have switch reference, this represents a clear parallel between *sujeito acusativo* contexts where *você* is preferred and contrastive contexts like those we tested in our survey.

The possibility of using *você* as a DO can be contextualized within the broader change toward the fixing of SVX word order in BP, which has had a number of important consequences including the well-known increased rates of overt subjects (cf. Kato and Negrão 2000; Tarallo 1996). This change has also occurred in conjunction with the loss of clitics in different persons and numbers (*o/a* 'him/her'; *nos* 'us'; *lhes* 'them'), which in turn has led to increased use of tonic pronouns (*vocês, ele/ela*, *a gente*) in DO function and postverbal position. In fact, in some cases the only option in spoken BP is now a tonic pronoun, such as the case of 2nd person plural (2PL) *vocês*, which does not show variability with a clitic like that found between 2SG *você* and *te*. Only *me* and *te* are fully viable in present-day spoken BP as DO clitics but, as we show in our results, in the 2SG especially there is abundant variation. Thus, in addition to the discourse-pragmatic function of marking contrast, the paradigmatic pressures of the present-day DO pronominal system are also affecting the choice of 2SG DO pronoun. We have modeled these pressures below in Figure 10, where the tonic DO pronouns associated with the other person/number combinations are depicted as affecting the 2SG (as shown by the solid-line arrows), while the 1SG *me* is the only clitic remaining that is potentially exerting paradigmatic pressure on 2SG expression (as shown by the dashed arrow). These pressures, coupled with the fact that Lamberti et al. (2017) also found that the use of *você* in DO function was not evaluated negatively by BP speakers (while other tonic pronouns in their study, such as 1SG *eu* or 1PL tonic *nós*, were heavily stigmatized as DOs), are all factors that are driving the use of *você*

verbs that may be used. The verbs we consider here, while not an exhaustive list, are among the most common.

even in non-contrastive contexts, where there is clearly a degree of inherent variability with its clitic counterpart *te*.

Figure 10. Paradigmatic pressures on 2SG DO pronoun expression
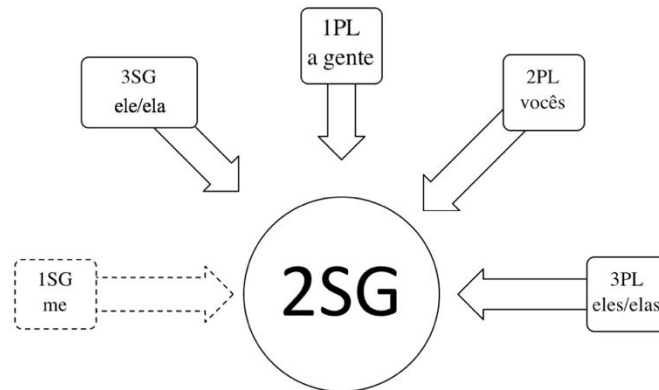
In order to further explore the behavior of subject *tu* users in contrastive contexts, we created a second version of the questionnaire which contained the original stimuli, but with the added option of *tu* used as a DO, as in (20) and (21) (cf. examples 2 and 3 above). This version of the questionnaire was shared on social media with respondents in several Brazilian regions known for use of *tu* as subject. This step was taken in order to investigate the possibility that, instead of *você*, these speakers would actually prefer to use *tu* as the 2SG DO pronoun in contrastive contexts. Such examples are easily found in online sources such as Twitter (e.g. *Eu amo só tu* 'I love only you').

(20) *Eu vi a Luana, mas queria ver tu.*

(21) *Eu vi a Luana, mas procurava tu.*

Results from this version of the questionnaire, completed by participants from the states of Pará, Rio Grande do Sul, and Santa Catarina, show that users of *tu* as subject pronoun may also use it as a DO, but they chose this option sparingly. In fact, *tu* was selected as a DO **only** in cases of contrast: in cases of single contrast (as in [20]; 1 out of 47 respondents), single negated contrast (2/49), and double contrast (as in [21]; 4/48).[15] As alluded to above, Twitter searches confirm that the use of *tu* as a DO is possible in single contrast contexts (e.g. *Amo até cerveja mas não amo tu* 'I even love beer, but I don't love you'), and in non-contrastive contexts as well (e.g. *Ainda amo tu muito* 'I still love you a lot'). However, our participants' responses reflected these possibilities in only very weak fashion, quite possibly due to prescriptive pressures: *tu* (unlike *você*) is stigmatized as a subject

---

15 The number of responses varies from one context to another because a few participants did not respond to all stimuli.

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista Linguíʃtica **283**

pronoun in BP and further extension of it into DO function is presumably even more so.[16]

We must emphasize once again, however, that even though strong pronouns like *você* may be preferred in contrastive contexts, they are clearly not obligatory in such contexts. This variability is parallel to that analyzed by Amaral and Schwenter (2005) for Spanish and European Portuguese, where strong pronouns in contrastive contexts are likewise not obligatory for subjects; other linguistic elements (such as adverbials of different types) can carry out the desired contrast as long as the identity of the subject referent(s) can be resolved.[17] We find a similar result for 2SG DOs in BP: even though tonic *você* is preferred in contrastive contexts it still varies with atonic *te*, and this variation is predictable based on the strength of the contrast being expressed.

The situation we have elucidated here for 2SG DO pronouns in BP echoes comments by Bresnan (2001: 114), who stated that, "In languages with both bound and free pronominals, the free pronoun generally is used for focus (Schwartz 1986) in those contexts where it contrasts with a bound form, but it may nevertheless fill in the gaps in the paradigm of bound pronominals, taking on the non-focus uses of the latter." As our results for the non-contrastive contexts show, *você* can be and is found in non-focus contexts with some frequency, but is more generally preferred for expressing contrastive focus. Thus, we would add to Bresnan's position that it is not only the gap-filling function in which we find *você* but in addition we see inherent variability between this form and the clitic *te*, spurred on by paradigmatic pressures (which in turn may be affected by the fixing of SVX order) that have led to significant leveling of the DO pronoun system in BP in favor of the tonic forms.

Conclusion

Our findings in this study provide contributions both to the more accurate description of the grammar of BP as well as to general linguistic theory. On the one hand, we have shown that, contra many grammatical descriptions (e.g. Perini 2010), *te* and *você* do not vary *livremente* ('freely') or *indiferentemente* ('interchangeably') but rather are sensitive to a clear discourse-pragmatic motivation, according to which *você* is the preferred form in contrastive contexts for many speakers, while *te* is

---

16   According to one of the authors of this paper (Lamberti), a native *tu* speaker from Rio Grande do Sul, another possible strategy would be to place a contrastive pitch accent on the clitic *te*. While clitics are typically considered to be unstressed elements, we know from recent studies of Argentine Spanish (Colantoni and Cuervo 2013; Klassen and Patience 2016) that they can be accented in some cases. In such instances, these forms could be considered weak pronouns instead of clitics in Cardinaletti and Starke's (1999) three-way typology. We leave this possibility open for further research.

17   See Pinillos Chávez and Dickinson (2018) for experimental verification of this for Spanish.

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista Lingüíſtica **284**

preferred elsewhere. However, *você* is also selected, albeit less often, in non-contrastive contexts due to paradigmatic pressures such as the use of tonic pronouns for DOs with other persons/numbers, including 2PL where the nominative form *vocês* is the only productive pronoun, 1PL where *a gente* is much more frequent than clitic *nos* (Lopes 2003; Travis and Silveira 2009; Zilles 2005), and also 3rd person DOs, where human DOs are overwhelmingly encoded by tonic pronouns in the spoken language (e.g. *Vi **ele** ontem no parque* 'I saw him (lit. he) yesterday at the park'; see Schwenter and Silva 2003, Schwenter 2014). Additionally, DO *você* is likely influenced by the syntactically similar *sujeito acusativo* constructions in which tonic pronouns appear post-verbally and fulfill a double grammatical role as both DO of the preceding and subject of the following verb. With only a few exceptions due to high frequency, *você* is the preferred form in *sujeito acusativo* structures. Together, these contexts provide a strong analogical model for *você* as a post-verbal DO, as opposed to the preverbal position of *te*. Finally, *você* as a DO is not subject to the same normative pressures as are other tonic pronouns (*eu, nós, ele, ela*), which are comparatively more socially stigmatized when used as post-verbal DOs (cf. Bagno 2001, 2012, Tesch et al. 2014, Dickinson & Lamberti 2017). While we have not investigated in detail the reasons behind these differences in evaluation, it seems plausible to assume that the widespread acceptance of *você* as the de facto "standard" subject pronoun in BP is at least partially transferred to its use in DO function.

Our findings also inform general linguistic theory by demonstrating that the well-known distinction between clitic/weak vs. strong pronouns (cf. Cardinaletti & Starke 1999, Manzini 2014) is relevant to 2SG DO pronoun choice in BP. That is, speakers prefer a strong form **(***você***)** in contrastive contexts but a maximally weak form (the clitic *te*) in non-contrastive contexts. This finding parallels that of Kaiser (2010) for the strong and weak forms in Estonian, corroborating her view that pronouns cross-linguistically can show sensitivity to contrast. In closely-related Romance varieties, such as European Portuguese and Spanish, contrastive contexts often play host to clitic doubling constructions, in which a tonic pronoun within a prepositional phrase "doubles" the atonic clitic. Indeed, as has been pointed out by Diniz (2007) and others, in Minas Gerais there also exist dialects where the *redobro* ('doubling') of clitics (as in *Eu te amo você*) can be found as a syntactic strategy of reinforcing the DO referent.[18] Thus, even though the details of the formal options differ across languages and dialects, a similar generalized strategy of employing a tonic element can be encountered in contrastive contexts. More broadly, then, our results corroborate the claim that "empirical data from a great number of

---

18   Strangely, however, the studies of this *redobro* construction do not consider the strictly postverbal option with *você* as DO. We have no explanation for this lack of attention, besides the seemingly widespread opinion (even among some linguists) that *te* and *você* are interchangeable as 2SG DO pronouns.

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista Lingüíʃtica **285**

languages show that contrast has a crucial impact on [choice of] linguistic forms" (Molnár 2006: 227).

Further, our results demonstrate that the choice of strong vs. weak/clitic forms is probabilistic, rather than discrete (cf. Bresnan 2007, Wolk et al. 2013). Speakers do not invariably choose *você* in contrastive contexts or *te* in the absence of contrast, but rather show clear quantitative preferences in their choices in speech/writing. The gradience of the pronominal distribution is paralleled by the gradient effects found in the strength of contrast (none vs. single vs. double, with choice of *você* increasing with each increase), as well as, potentially, the presence/absence of the contrastive conjunction *mas* ('but') in contrastive contexts. We can visualize this gradience in the following way, where the form in CAPS is preferred but the non-capitalized form is not; as our survey results show, in single contrast contexts, both forms are selected equally:

No Contrast	**Single Contrast**	**Double Contrast**
TE/você	TE/VOCÊ	te/VOCÊ

In other words, just as strong vs. weak pronoun choice is not categorically determined by contrastive vs. non-contrastive contexts, contrast itself is likewise not binary (Molnár 2006). The more overtly contrastive the context, the more likely BP speakers are to choose *você* over *te*, thereby leading to an isomorphic, but still gradient, relationship between context and form. Again, such differences in our results are suggestive of the gradient effects of contrast on DO pronoun choice and on the realization of DOs more generally.

REFERENCES

Amaral, Patrícia & Scott A. Schwenter. 2005. Contrast and the (non-) occurrence of subject pronouns. In David Eddington (Ed.), *Selected Proceedings of the 7th Hispanic Linguistics Symposium,* 116-127. Somerville, MA: Cascadilla Proceedings Project.

Azevedo, Milton. 2005. *Portuguese: A linguistic introduction.* Cambridge: Cambridge University Press.

Bagno, Marcos. 2001. *Português ou brasileiro? Um convite à pesquisa*. São Paulo: Parábola.

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista Linguística **286**

Bagno, Marcos. 2009. *Não é errado falar assim! Em defesa do português brasileiro.* São Paulo: Parábola.

Bagno, Marcos. 2012. *Gramática pedagógica do português brasileiro*. São Paulo: Parábola.

Bates, Douglas, Martin Maechler, Ben Bolker, & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1-48. doi:10.18637/jss.v067.i01.

Bechara, Evanildo. 1999. *Moderna gramática do português contemporâneo*. Rio de Janeiro: Lucerna.

Bresnan, Joan. 2001. The emergence of the unmarked pronoun. In Géraldine Legendre, Sten Vikner, & Jane Grimshaw (Eds.), *Optimality-theoretic Syntax*, 113-142. Cambridge, MA: MIT Press.

Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (Eds.), *Roots: Linguistics in search of its evidential base*, 77-96. Berlin: Mouton de Gruyter.

Cardinaletti, Anna & Michal Starke. 1999. The typology of structural deficiency. A case study of the three classes of pronouns. In Henk van Riemsdijk (Ed.), *Clitics in the languages of Europe*, 145-233. Berlin: Mouton de Gruyter.

Colantoni, Laura & María Cristina Cuervo. 2013. Clíticos acentuados. In Laura Colantoni & Celeste J. Rodríguez-Louro (Eds.), *Perspectivas teóricas y experimentales sobre el español de la Argentina*, 143-158. Madrid/Frankfurt: Iberoamericana/Vervuert.

Cunha, Celso & Lindley Cintra. 2001. *Nova gramática do português contemporâneo.* Terceira edição revista. Rio de Janeiro: Editora Nova Fronteira.

Davies, Mark & Michael Ferreira. 2006-. Corpus do português: 45 million words, 1300s-1900s. http://www.corpusdoportugues.org.

Dickinson, Kendra V. & Luana Lamberti. 2017. Anaphoric direct objects in Brazilian Portuguese. Paper presented at Hispanic Linguistics Symposium, Texas Tech University.

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista Linguíſtica **287**

Diniz, Carolina Ribeiro. 2007. Eu te amo você: O redobro do clítico do PB sob uma abordagem minimalista. Masters Dissertation, Universidade Federal de Minas Gerais.

Duarte, Maria Eugênia & Jânia M. Ramos. 2015. Variação nas funções acusativa, dativa e reflexiva. In Marco Antonio Martins & Jussara Abraçado (Eds.), *Mapeamento sociolinguístico do português brasileiro*, 173-195. São Paulo: Editora Contexto.

Estigarribia, Bruno. 2013. El modelo de las fuerzas discursivas y el doblado de clíticos rioplatense. *Signo y Seña* 23, 119-142.

Hothorn, Torsten, Peter Buehlmann, Sandrine Dudoit, Annette Molinaro, & Mark Van Der Laan. 2006. Survival Ensembles. *Biostatistics* 7, 355-373.

Kaiser, Elsi. 2010. Effects of contrast on referential form: Investigating the distinction between strong and weak pronouns. *Discourse Processes* 47, 480-509.

Kato, Mary Aizawa & Esmeralda Vallati Negrão (Eds.). 2000. *Brazilian Portugese and the null subject parameter.* Madrid/Frankfurt: Iberoamericana/Vervuert.

Klassen, Gabrielle & Matthew Patience. 2016. Stressed clitics in Argentine Spanish: Which way does the clitic lean? In Alejandro Cuza et al. (Eds.), *Inquiries in Hispanic linguistics: From theory to empirical evidence*, 149-170. Amsterdam: John Benjamins.

Lamberti, Luana, Kendra V. Dickinson & Eleni Christodulelis. 2017. Variable Direct Objects in Brazilian Portuguese: Linguistic Constraints and Social Acceptability. Poster presented at NWAV 46, University of Wisconsin-Madison.

Lara Bermejo, Víctor. 2017. La generalización de *vocês* en el portugués europeo continental y su patrón de difusión geográfica. *Hispanic Research Journal* 18, 93-117.

Lopes, Célia Regina dos Santos. 2003. *A inserção de* a gente *no quadro pronominal do português.* Frankfurt/Madrid: Vervuert/Iberoamericana.

Manzini, Maria Rita. 2014. Grammatical categories: Strong and weak pronouns in Romance. *Lingua* 150, 171-201.

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista Linguística **288**

Martins, Ana Maria. 1993. Clitic placement from Old to Modern European Portuguese. *Historical Linguistics, 1993: Selected Papers from the 11th International Conference on Historical Linguistics*, 295-308. Amsterdam: John Benjamins.

Molnár, Valéria. 2006. On different kinds of contrast. In Valéria Molnár & Susanne Winkler (Eds.), *The Architecture of Focus*, 197-234. Berlin: Mouton de Gruyter.

Monteiro, José Lemos. 1994. *Pronomes pessoais*. Fortaleza: Editora Universidade Federal do Ceará.

Perini, Mário. 2002. *Modern Portuguese: A reference grammar*. New Haven, CT: Yale University Press.

Perini, Mário. 2010. *Gramática do português brasileiro*. São Paulo: Parábola.

Petersen, Carol. 2008. A tripartição pronominal e o estatuto das proformas *cê, ocê* e *você*. *D.E.L.T.A.* 24, 283-308.

Pinillos Chávez, Paloma & Kendra V. Dickinson. 2018. Contrastive focus in Spanish. The role of subject pronouns and stance adverbs. Paper presented at The Ohio State University Congress on Hispanic and Lusophone Linguistics (OSUCHiLL), The Ohio State University.

Preto-Rodas, Richard A. 1972. Some examples of pronominal confusion in modern Peninsular Portuguese. *Hispania* 55, 327-330.

R Core Team. 2017. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.R-project.org/.

Ramos, Jânia. 1996. O uso das formas *você, ocê* e *cê* no dialeto mineiro. In Dermeval da Hora Oliveira (Ed.), *Diversidade lingüística no Brasil*, 43-60. João Pessoa: Idéia Editora.

Schwartz, Linda. 1986. The function of free pronouns. In Urusula Wiesemann (Ed.), *Pronominal Systems*, 405-436. Tübingen: Gunter Narr Verlag.

Schwenter, Scott A. 2014. Two kinds of Differential Object Marking in Portuguese and Spanish. In

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

289

Patrícia Amaral & Ana Carvalho (Eds.), *Portuguese/Spanish interfaces. Diachrony, synchrony, and contact* (Issues in Hispanic and Lusophone Linguistics 1), 237-260. Amsterdam: John Benjamins.

Schwenter, Scott A. & Gláucia Silva. 2003. Anaphoric direct objects in spoken Brazilian Portuguese: Semantics and pragmatics. *Revista Internacional de Lingüística Iberoamericana* 2, 109-133.

Tarallo, Fernando. 1996. Turning different at the turn of the century: 19th century Brazilian Portuguese. In Gregory R. Guy et al. (Eds.), *Towards a social science of language, vol 1: Variation and change in language and society*, 199-220. Amsterdam: John Benjamins.

Tesch, Leila Maria, Lilian Coutinho Yacovenco, & Maria Marta Pereira Scherre. 2014. Variação e mudança na fala e na escrita: caminhos e fronteiras. *Revista (Con)textos Linguísticos* 8, 87-106.

Thomas, Earl. 1974. *A Grammar of Spoken Brazilian Portuguese.* Nashville, TN: Vanderbilt University Press.

Travis, Catherine & Agripino Silveira. 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.

Vitral, Lorenzo. 1996. A forma *cê* e a noção de gramaticalização. *Revista Estudos da Linguagem* 4, 115-124.

Washington, Hannah. 2015. *Variable object clitic placement: Evidence from European and Brazilian Portuguese*. PhD Dissertation, The Ohio State University.

Whitlam, John. 2011. *Modern Brazilian Portuguese grammar*. New York: Routledge.

Wolk, Christoph, Joan Bresnan, Anette Rosenbach, & Benedikt Szmrecsanyi. 2013. Dative and genitive variability in Late Modern English. *Diachronica* 30, 382-419.

Zampaulo, André. 2014. Los pronombres de objeto directo *te* y *você*: variación lingüística en el portugués brasileño. *Cuadernos de Lingüística* 2, 173-193.

Zilles, Ana Maria Stahl. 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.

Rio de Janeiro | Volume 14 | número 2 | p.259-290 | mai.-ago. 2018
Estudos Experimentais do Português | *Experimental Studies on Portuguese*

Revista Lingüística **290**