http://dx.doi.org/10.1590/1678-460X2020360107

DELTA

Articles

The Effect of Semantic and Discourse Features on the Use of Null and Overt Subjects – A Quantitative Study of Third Person Subjects in Brazilian Portuguese

O Efeito de Características Semânticas e Discursivas no Uso de Sujeitos Nulos e Pronominais – Um Estudo Quantitativo dos Sujeitos de Terceira Pessoa no Português do Brasil

Eduardo Correa Soares[1]
Philip Miller[2]
Barbara Hemforth[3]

ABSTRACT

In a corpus study and two acceptability experiments, we investigate whether semantic and discourse features of the antecedents affect the use of null and overt subjects in Brazilian Portuguese. Previous literature has proposed two hypotheses: the Hierarchy of Referentiality Hypothesis

1. Universidade Federal de Santa Catarina. Florianópolis, Santa Catarina – Brasil. http://orcid.org/0000-0002-4526-3299. E-mail: soares_ec@yahoo.com.br.
2. Université de Paris. Paris – França. https://orcid.org/0000-0002-1436-9533. E-mail: philip.miller@u-paris.fr.
3. Centre Nationale de Recherche Scientifique/Université de Paris. Paris – França. https://orcid.org/0000-0001-8371-1323. E-mail: bhemforth@gmail.com.

(Cyrino et al. 2000, inter alia) and the Semantic Gender Hypothesis (Creus & Menuzzi 2004, among others). We show that animacy and specificity affect the relative frequency of null and overt subjects and their acceptability in sentences. We propose an approach that accounts for previously published results as well as our new data in the light of a theory of anaphora resolution (Ariel 1990, etc).

Keywords: *Null Subjects; Animacy; Specificity; Anaphora Resolution.*

RESUMO

Em um estudo de corpus e em dois experimentos de aceitabilidade, investigamos se as propriedades semânticas e discursivas dos antecedentes afetam o uso de sujeitos nulos e pronominais no português brasileiro. A literatura propôs duas hipóteses: a Hipótese da Hierarquia de Referencialidade (Cyrino et al. 2000, entre outros) e a Hipótese do Gênero Semântico (Creus & Menuzzi 2004, e outros). Mostramos que animacidade e especificidade afetam a frequência relativa de sujeitos nulos e pronominais e sua aceitabilidade em sentenças. Propomos uma abordagem que explica os dados publicados anteriormente e nossos próprios novos dados à luz de uma teoria geral da resolução anafórica (Ariel 1990, etc).

Palavras-chave: *Sujeitos Nulos; Animacidade; Especificidade; Resolução Anafórica.*

1. Introduction

The present paper focuses on the use of third person null and overt subjects in Brazilian Portuguese (henceforth, BP). In the mainstream generative literature about null and overt subjects (analyzed in terms of the 'pro-drop parameter'), some authors suggest that BP is a living example of ongoing parametric change (Duarte 1993, 1995, 2000, *inter alia*). These proposals can be broadly outlined as claiming that BP is on a path from a null subject language (like Italian, for instance) toward a language with obligatory phonological expression of grammatical subjects (like English, for example). However, it has long been observed that the decrease in the number of null subjects has not affected third

persons as much as other persons (Negrão 1990, Duarte 1993, 1995, Cyrino et al. 2000). In the literature about BP, two different hypotheses have been proposed to account for this imbalance:

(i) The Hierarchy of Referentiality Hypothesis (HRH): The relative number of null subjects in the third persons is higher than in the others because third persons tend to be lower on a natural scale, the "Hierarchy of Referentiality": third persons NPs have inanimate or non-specific referents (as illustrated in examples 1a and 1b respectively) more frequently than the other discourse persons (Cyrino et al. 2000, Duarte, Mourão & Santos 2012, Duarte, Mourão & Guimarães 2012, Kato & Duarte 2014).

(ii) The Semantic Gender Hypothesis (SGH): Null subjects are preferably used to refer to a specific class of referents, namely those which have a negative value for the feature "Semantic Gender" ([–semantic gender]) (Creus & Menuzzi 2004, Othero & Spinelli 2017, 2019a,b), such as (1a) and (1b), as opposed to (2). This is due to the fact that [+semantic gender] referents are similar to those of the other discourse persons.

(1) a. A casa estava velha. Esse suporte$_1$ caía toda hora ... uma vez $_{-1}$
the house be.IMP.3SG old this support fall.IMP.3SG all time ... one time
caiu na orelha da empregada, $_{-1}$ quase tira a orelha
fall.PST.3SG in.the ear of.the maid, almost take.PRES.3SG the ear
fora ...
out ...
"The house was old. This support$_1$ fell all the time . . . once [it$_1$] fell on the ear of the maid, [it$_1$] almost took her ear off . . . "

b. as pessoas$_1$ comem tanto ... $_{-1}$ comem milho ... paçoca
the people eat.PRES.3PL so.much ... $_{-1}$ eat.PRES.3PL corn ... paçoca
... pamonha ...
... pamonha ...
"People$_1$ eat so much... [they$_1$] eat corn . . . paçoca . . . pamonha . . . "
(NURC-RJ, "Inquiry 011")

(2) Mas a garota$_1$ é novinha. Ela$_1$ ainda faz faculdade ...
But the girl be.PRES.3SG young.DIM. She still do.PRES.3SG college ...
"But the girl$_1$ is young. She$_1$ is still an undergrad . . . "
(NURC-RJ, "Inquiry 003ac")

This paper aims to investigate these two hypotheses combining corpus and experimental data. Several corpus analyses have recently been reported with respect to these two hypotheses (Othero & Spinelli 2017, 2019*a,b*). We carried out a new analysis of the NURC-RJ corpus, originally studied by Duarte (1995, 2000), providing a more fine-grained analysis of the cases in third person singular and plural and using inferential statistics to analyze the generalizability of the results. This corpus study shows that the effect of feature specificity is not the same for the singular and for the plural third persons when the referent of the subject is inanimate. Additionally, we report two experiments in which materials were set up so as to allow specificity and animacy to vary independently, making it possible to evaluate their respective effects on the acceptability of null and overt subjects in a controlled environment.

Our results show that each of these features has a significant effect on the acceptability of null and overt subjects: null subjects are preferred to refer to antecedents that are [–animate] or [–specific]. Thus, our experimental results favor the HRH overall, since the crucial distinction between the HRH and the SGH is that only the former predicts a role for specificity, whereas both theories predict that inanimate referents will favor null subjects. However, the interaction between animacy, specificity and number, found in the corpus analysis, leads us to another interpretation: using a suggestion made in Othero & Spinelli (2019a), we propose a version of the SGH that relies on the morphological exponent of gender and number semantic features. Namely, we argue that the inanimate third person singular is not affected by specificity, because BP singular overt pronouns are the exponent of [–plural] [+semantic gender] while the null is the unmarked option for [–plural] [–semantic gender] antecedents, much like the pronouns "he" and "she" as opposed to "it" in English; on the other hand, third person plural pronouns are the exponent of [+plural] [±semantic gender] in BP. This set up make them better candidates for coreference with [+plural] [–semantic gender] referents with the effect of specificity entering the calculus of the accessibility of the antecedent: null subjects are favored when they have intradiscursive antecedents, *i. e.* non-specific.

This paper is organized as follows. In section 2, we briefly review the diachronic data and assumptions about the change in the use of null

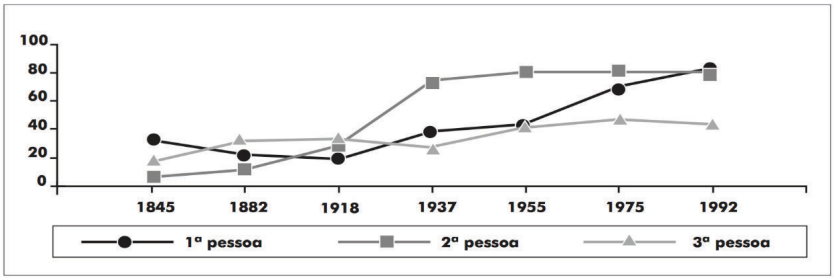
and overt subjects in BP and show the problem with the third persons. In section 3, we present the two competing hypotheses to account for the higher number of null subjects in third persons. In section 4, we insert this discussion in a general theory of anaphora resolution. In section 5, we present and discuss our corpus data. In section 6, we present experimental evidence to test the hypotheses. Finally, in section 7, we discuss both the corpus and experimental evidence with respect to these accounts and put forth our proposal.

2. Null Subjects in Present Day Brazilian Portuguese

The BP inflectional system is substantially impoverished when compared to its previous stages and to other varieties of Portuguese or to other Romance Languages (Italian and Spanish, for instance) (see Duarte 1995, Kato & Negrão 2000, *inter alia*). At the same time, null subjects in present day spoken BP are becoming scarcer (see Duarte 1993, Kato & Negrão 2000, among many others). Much research has observed, however, that the decrease in the number of null subjects is not uniformly distributed across discourse persons. Studying a corpus of oral production in a public school in São Paulo, Negrão (1990) was the first to point out that overt pronouns are less used in the third person than in the others. In a study of popular written plays, Duarte (1993, 1995, 2000) also observed an asymmetry across discourse persons over the period at stake. Duarte (1995) claims that the impoverishment of the inflectional paradigm, along with the deactivation of the "Avoid Pronoun Principle", triggered the decline in the relative percentage of null subjects from 80% in the second quarter of the XIXth century to 26% in the 1990s (see Duarte 1993, *inter alia*).[4] Nonetheless, the decrease in the number of null subjects has a more far-reaching effect on the first and second persons than on the third persons, as plotted in Figure 1 below.

4. The Avoid Pronoun Principle is a principle which states that pronouns should not be used whenever they are not required (Chomsky 1981).

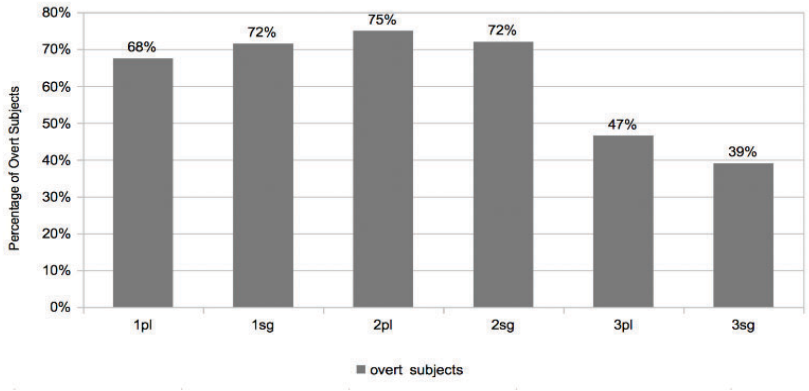
Figure 1 – Percentage of Overt Subjects according to Discourse Person



Duarte 1995, p. 20.

In previous research (Soares, Miller & Hemforth, 2019), we reported a similar observation regarding the distribution of null and overt subjects across discourse persons. Reassessing the NURC-RJ corpus, we found a significantly lower number of overt subjects in third persons than in all other persons. As Figure 2 shows, the data presented in our research overlap with those presented in previous research (Negrão 1990, Duarte 1993, 1995, 2000, Cyrino et al. 2000, Barbosa et al. 2005, Duarte 2012, Kato & Duarte 2014, Duarte 2015, *inter alia*). In our research, however, the data are also sorted into singular and plural: third person subjects (46.5%, *i. e*. 3740 out of 8032 finite clauses) are much more frequently null than the others. Only about half of the third person plural subjects and even fewer of the third person singular subjects are overt. The remaining discourse persons are more frequently overt, with quite similar distributions. To compare preferences for overt and null subjects, we entered the data into a binomial generalized mixed-effects model with a link function ("logit", a logistic regression) with overt pronouns coded as "1" and null subjects coded as "0", with the "interviewed speaker" as a random factor including random slopes for the factor Discourse Person. The maximal model revealed that, taking the condition with the highest number of overt subject pronouns (2nd person plural) as the baseline, overt subjects pronouns occur significantly less frequently for the 3rd discourse persons (for plural, ß:-2.171/SE: 0.3058/z-value: -7.10/p-value:1.25e12 and for singular, ß:-3.396/SE: 0.2874/z-value: -11.818/p-value: < 2e-16) than the baseline (see Soares, 2017 for the full statistical analysis).

Figure 2 – Percentage Overt Subjects in NURC-RJ according to Discourse Persons



Soares 2017, p. 49.

To account for the observed data, Cyrino et al. (2000), Duarte, Mourão & Santos (2012), Duarte, Mourão & Guimarães (2012), Kato & Duarte (2014), Duarte (2015), Duarte & Reis (2018) propose a diachronic path along which the change in BP is taking place. We present this proposal in detail in the next section.

3. Null Subjects and the Features of the Antecedents

3.1. The Hierarchy of Referentiality Hypothesis

Cyrino et al. (2000), Kato et al. (2006), Duarte, Mourão & Santos (2012), Kato & Duarte (2014), Duarte (2015), try to explain the asymmetry in the use of null subjects in BP across different persons and numbers in diachronic terms. They propose that the change progresses over the whole BP system governed by a "Hierarchy of Referentiality", as shown in (3). According to this hierarchy, languages tend to use overt pronouns for picking up more referential entities, that is, those that are higher in the hierarchy.[5]

5. According to (Cyrino et al. 2000, 59), "[+N +human] arguments are the highest in the Referential Hierarchy, while non-arguments [(expletives)] are the lowest. For pronouns,

(3) Hierarchy of Referentiality

non-argument	proposition	[-human]	[+human]
			3rd 2nd 1st person
-specif			+specif
[-ref] <—			—> [+ref]

(4) The Implicational Mapping Hypothesis:
The more referential the subject is, the greater the possibility of it being expressed by a non-null pronoun is.

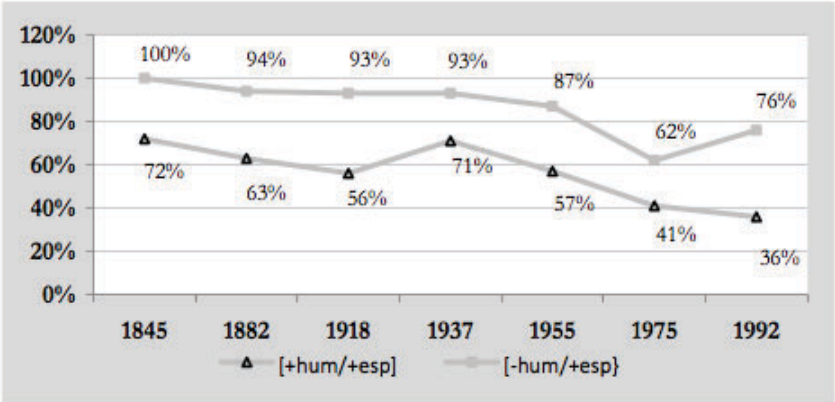
(Cyrino et al. 2000, p.39)

Assumption (4) predicts that more referential subjects are likely to be expressed by overt pronouns. This hypothesis is slightly counterintuitive, since in the literature on anaphora resolution, for example, the correlation is taken to be that the more accessible/salient the antecedent, the less explicit the anaphoric element needs to be, with null items being the least explicit (see section 4 below). This leads to predictions diverging from those made by the Hierarchy of Referentiality. Specifically, much literature assumes that less referential antecedents (e.g. higher order entities, less specific entities, etc.) are inherently less accessible/salient than first order entities, including humans, and specific entities, so that one would expect more null subjects for the latter than for the former (Ariel 1990, Gundel et al. 1993, *inter alia*). See section 4 below for further discussion. Assumption (4), however, offers a possible account for the higher number of third person null subjects in BP, which has long been observed by many researchers. Duarte, Mourão & Santos (2012) sorted the data previously reported by Duarte (1993) and Duarte (1995) into different semantic categories using the features [± human] and [± specific]. In the diachronic data, presented in Figure 3 below, they found a clear imbalance between human and non-human antecedents (presumably animate non-human referents are not common in written plays). However, they did not present the data with respect to specificity because of the low number

since the speaker (*eu* 'I') and the addressee (*você* 'you') are inherently human, first and second person pronouns are the highest in the hierarchy, while third person pronouns referring to a proposition are the lowest, with [-animate] entities in between. The feature [± specific] interacts with all the other features."

of cases. It is worth noticing that the trajectory of change, which appears to be moving unidirectionally toward a lower number of null subjects till 1975, exhibits a different pattern in the last observation, rising from 62% to 76%.

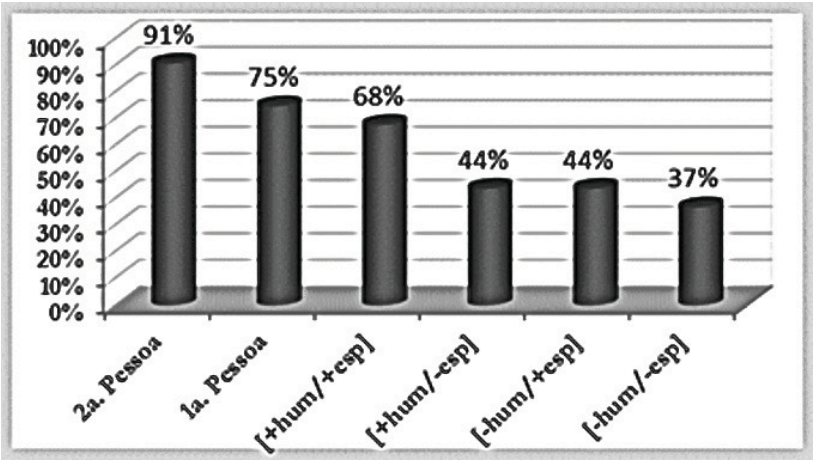
Figure 3 – Percentage of Null Subjects according to Human and (E)specific Features



Duarte & Varejão 2013, p. 106.

With respect to corpus data, in a series of papers, Duarte and colleagues (Duarte, Mourão & Santos 2012, Duarte & Varejão 2013, Duarte 2015, Duarte & Reis 2018) provide a reassessment of the importance of the semantic features to account for the imbalance across discourse persons in BP as regards the change in the relative frequency of null subjects. As summarized in Figure 4, non-human or non-specific referents are retrieved by overt subjects less frequently than human and specific antecedents. Each of these features also seem to have a cumulative effect, since referents that are both non-human and non-specific are the least frequently retrieved by overt subjects.

Figure 4 – Percentage of Overt Subjects according to Discourse Person and Semantic Features



Duarte & Reis 2018, p. 180.

3.2. The Semantic Gender Hypothesis

Extending a proposal to account for the use of null and pronominal objects in BP put forth by Creus & Menuzzi (2004), Othero & Spinelli (2017, 2019a,b) propose a hypothesis according to which the two features proposed by HRH are subsumed under one single feature, namely [± semantic gender], which is a correlate of natural sex.[6] In Table 1, we summarize the semantic/discourse features proposed in the literature with the examples extracted from Creus & Menuzzi (2004). Animacy and Specificity are relevant for the HRH, whereas Semantic Gender is relevant for the Semantic Gender Hypothesis.

6. Othero & Schwanke (2018) define [± semantic gender] as "the feature that distinguishes nouns that refer to sexually marked individuals from nouns that refer to individuals whose sex is not marked or, more precisely, nouns that do or do not denote an apparent sex distinction" (in our translation). They provide the following examples of [+ semantic gender] *homem* "man", *mulher* "woman", *professor* "teacher.MASC" (vs. *professora* "teacher.FEM"), *cachorro* "dog.MASC" (vs. *cadela* "dog.FEM") as opposed to *mesa* "table.FEM" (FEM), *livro* "book.MASC" (MASC), *vítima* "victim" (grammatically feminine, but used to refer to both sexes), *cônjuge* (grammatically masculine, but used to refer to both sexes), *tartaruga* (grammatically feminine, but used to refer to both sexes), which are [– semantic gender]. The notion of [± semantic gender] may be highly disputable. For the sake of this paper, we assume this notion as proposed in the references cited.

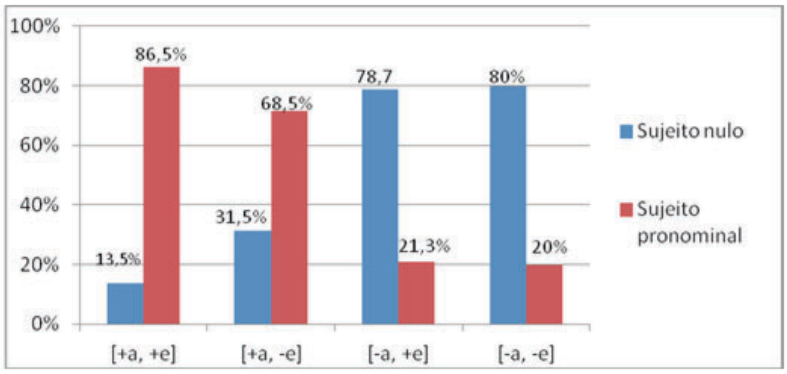
Table 1 – Sample with the relevant features adapted from Creus & Menuzzi (2004, p. 153)

	Examples	Anim/Hum	Spec	SemGen
(A)	"Maria, este senhor, esta mulher"	+	+	+
	Maria, this gentleman, this woman			
(B)	"um menino, todo garoto, qualquer mulher"	+	-	+
	a boy, every boy, any woman			
(C)	"muita gente, toda pessoa, um profissional"	+	-	-
	many people, every person, a professional			
(D)	"essa pedra, este carro, o Rio de Janeiro"	-	+	-
	this stone, this car, Rio de Janeiro			
(E)	"qualquer ávore, uma sala, um poema"	-	-	-
	any tree, a room, a poem			

According to the Semantic Gender Hypothesis [SGH], put forward by Othero & Spinelli (2017, 2019a,b), pronominal subjects are mostly used to retrieve [+semantic gender] antecedents, while null subjects are preferentially used to refer to [–semantic gender]. Though the HRH and the SGH make identical predictions in some cases, they differ in other cases. For instance, cases (A) and (B) are treated as identical under the SGH, whereas they differ in terms of specificity under the HRH. The NP *os cantores* "the singers", for example, is [+ semantic gender], but may be inserted in a context in which it is not specific, as in *Os cantores em geral têm muitos fãs* "Singers in general have many fans" (see Experiment 2, in section 6.3 below). Similarly, for (B) and (C), for (D) and (E) and for (C) and (E). Thus, these cases with divergent predictions can serve to provide evidence adjudicating between the two hypotheses.

Othero & Spinelli (2017, 2019a,b) gathered spoken data in two corpora from a Southern state of Brazil (Rio Grande do Sul): VARSUL-RS and LinguaPOA. They found data that partially support the SGH. We will restrict the discussion to these data for reasons of space and present only one plot that shows figures similar to what we found in our corpus research in section 5 (Figure 5, with a=animate, e=specific).

Figure 5 – Percentage of Null and Pronominal Subjects according to Semantic Features



Othero & Spinelli 2019a, p. 18.

As regards the comparisons proposed above, we can observe that the distribution partially corroborates the SGH, since, as predicted by that hypothesis, there is no difference between specific and non-specific inanimates (examples (D) and (E) in Table 1). The HRH would predict a distinction due to the difference in specificity. As for animate referents, it is not clear whether the difference found between + and – specific animates can be accounted for by the fact that some of these are [+semantic gender] and others [–semantic gender]. Clearly these matters require further investigation.

In this paper, rather than arguing for one of these two analyses, we attempt to provide an account of the data that builds on the SHG and also takes specificity into consideration in the calculus of antecedent salience. In view of building such an analysis, the next section briefly reviews a general theory of anaphora resolution.

4. Anaphora Resolution

As summarized in the previous section, many researchers have found that less referential antecedents are relatively more frequently referred to by null subjects than by overt pronouns in BP (Cyrino et al. 2000, Duarte, Mourão & Santos 2012, Kato & Duarte 2014, Soares 2017, Othero & Spinelli 2019*a,b*, *inter alia*). This fact could be

viewed as counter-evidence for a standard assumption in the literature about anaphora resolution (Ariel 1990, 2001, Gundel et al. 1993, Grosz et al. 1995, Almor 1996, Carminati 2002, among many others): anaphora resolution is guided by a reverse mapping principle between antecedent salience and anaphor explicitness – more salient antecedents are retrieved by less complex and less informative anaphoric forms. Evidence in favor of this hypothesis has come from many different sources, from corpus research to experimental studies (see for example McEnery 2000, Garnham 2001, for overviews).

It turns out that BP is only an apparent counterexample to this principle. To see this, it is necessary to clarify the notion of antecedent salience. In the literature, different definitions and applications of salience have been proposed, which ultimately lead to contradictory predictions. We propose to split the notion of salience into two different subtypes: discourse salience and inherent semantic salience. In this section, we focus on differentiating them. In Section 7 below, these two subtypes will be incorporated into a coherent theory in order to understand how they can provide insights to understand the data in this paper.

Previous research has studied semantic features of the antecedent as relevant for anaphora resolution. Most of them propose a hierarchy such as (5) below, taken from Silverstein (1976):

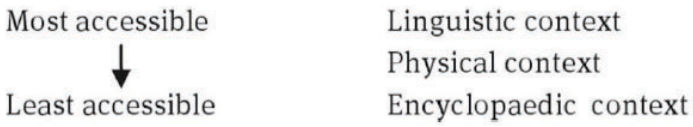
(5) HUMAN > ANIMATE > INANIMATE.

Dahl & Fraurud (1996) show that this hierarchy is a strong predictor for the choices between pronouns and NPs in a Swedish written corpus study: humans are more likely to be pronominalized, while non-humans are more likely to be retrieved by NPs or demonstratives. Bittner (2007) and Gagarina (2007) crossed Animacy with the Syntactic Function of the antecedent in several interpretation experiments with children in German and Russian respectively. Both studies tested null subjects (ungrammatical in German, except for cases of topic drop), personal pronouns and demonstratives. Some of the main findings are that (i) when crossing the semantic feature of animacy with the syntactic feature of grammatical function, no unified notion of prominence can be proposed, (ii) younger children seem to be more sensitive to animacy,

while older children seem to rely more on syntactic salience; and (iii) there is a primacy of syntactic over animacy salience in these languages in the contexts that were tested, since the less complex forms were significantly biased toward the subject, although animacy also exerted effects on many conditions and age groups. However, it is clear that younger children tend to use animacy as the basis for salience, while older children bias their interpretation according to more "linguistic" features (e. g. subjecthood).[7]

As for the discourse-semantic definition of Salience, there are various somewhat different proposals in the literature. The main point for us is that, if salience is taken to be a scale based on ontological properties of the referents, with human referents being the most salient and abstract referents the least, BP data would challenge the predictions of the reverse mapping principle, since null subjects in BP are biased toward lower levels of salience in this sense. But, if salience is understood as a discourse property, as proposed by Ariel (1990), the preferences established for BP null subjects do not contradict the inverse relation mapping. According to Ariel, the salience of the antecedent is defined according to its degree of discursiveness, cf. Figure 6 below:

Figure 6 – Ariel (1990)'s Salience of the Antecedent



Following this approach, many papers have pointed out that the null forms appear to be biased toward discourse internal antecedents,

7. There does not seem to have been any study of the effects of specificity in the literature on anaphora resolution. The closest investigation appears to be research by Carminati (2002), who tested the bound variable behavior of null subjects which co-refer with quantified antecedents, following Montalbetti (1984) (the behavior of null subjects when they are bound by quantified antecedents is a matter of much discussion in the literature see, among others, Grodzinsky & Reinhart 1993 and Reuland 2001). She found a correlation between the use of null subjects in Italian and their interpretation as bound by quantified subject antecedents as proposed by Montalbetti (1984).

while overt forms are freer to refer either deictically or to discursively less salient antecedents (see Mayol 2010, *inter alia*).

Based on Ariel's proposal and on an observation made by Othero & Spinelli (2019a), according to which specificity is a discourse feature (different from animacy, which is a lexical feature), in section 7 we propose to incorporate specificity in the calculus of antecedent salience. At this point, two ideas must be kept in mind: (i) inherent semantic features of the referent (animacy, for instance) do not overlap with discourse sentential salience (topicality, centrality, among others); and (ii) different notions of salience lead to different and sometimes contradictory predictions concerning the inverse correspondence principle of anaphora resolution.

In the next two sections, we present empirical research which provides a clearer view on the data that the HRH and SGH hypotheses purport to account for, allowing a more precise understanding of their relation with general principles of anaphora resolution. In section 5, the results of reassessing the corpus previously studied by Duarte (1995) are reported. These results were corroborated in two experiments, the results of which are reported in section 6.

5. Reassessing Corpus Data

In this section, the results of a reanalysis of the NURC-RJ corpus are reported. A new analysis was carried out for three main reasons: (i) the criteria used to exclude some data in previous research seemed too restrictive and ended up excluding cases that for the purposes of the present paper are crucial; (ii) with new theories and analytical toolkits, such as new statistical packages, more relevant factors and correlations might be discovered (see, for instance, Gries 2015 for a critical point of view on previous corpus studies without inferential statistical analysis and for arguments in favor of using (generalized) linear mixed models in this sort of analysis) and (iii) the amount of data analyzed here was at least three times larger than in previous analyses. Nine interviews carried out in the 70s and nine interviews from the 90s (of which twelve were with the same person during the two relevant periods, that is, six people participate twice), were analyzed. Overall 8032 inflected

clauses in which the subject was either co-referential, deictic or generic were collected. These data were descriptively analyzed in qualitative and quantitative terms. Finally, an inferential analysis was carried out using logistic regressions with the *glmer* function of the *lmer4* package applying the logit linking function with Laplace approximations in the statistical environment R (R Core Team, 2018).[8]

5.1. Analysis

We analyzed the discourse-semantic and inherent features of the antecedents by sorting them into specific vs. non-specific, animate vs. inanimate, singular vs. plural based on the Hierarchy of Referentiality in (3). In this paper we only report the results for non-sentential referents (total: 2882 clauses). The distinction between plural and singular is important, because differently from grammatical gender, grammatical plural usually implies semantic plurality.[9] This feature might (and indeed it does) interact with the other features we are interested in.

Animacy is easy to annotate in BP. Animals are taken to be animate, since they can be combined with almost any predicate typical of animate humans, such as intentional predicates, as *morder* "to bite", or sentience predicates, such as *sentir* "to feel".

The definition of specificity raises more problems. As shown in the literature on the topic (Enç 1991, Abbott 1995, von Heusinger 2002, Falco 2002, Kagan 2006, von Heusinger 2011), despite the notion's intuitive simplicity, it is difficult to come to a consensus on a formal definition of specificity, and it is certainly beyond the scope of this paper to attempt to do so. For the present purposes, the following operational criterion will be used.

8. See Baayen et al. (2008), Jaeger (2008), Bates & Maechler (2009), Bates et al. (2011, 2015) for details on the statistics. For more details about our methodology, see Soares (2017).

9. A reviewer raised the point that arbitrary plural null subjects do not imply semantic plurality. This is probably the case in sentences like *Bateram na porta.* "[They (=someone)] knocked on the door", but they were not included in our sample from NURC-RJ.

(5) Operational criterion for Specificity:
Given a NP denotation α in a predicate β, the denotation of α is specific iff:

$$\forall x \in \| \alpha \| [\| \beta \| (x)] \quad \rightarrow \quad \neg \diamond \exists y \in \| \alpha \| \neg [x \otimes y] \wedge [\| \beta \| (y)]$$

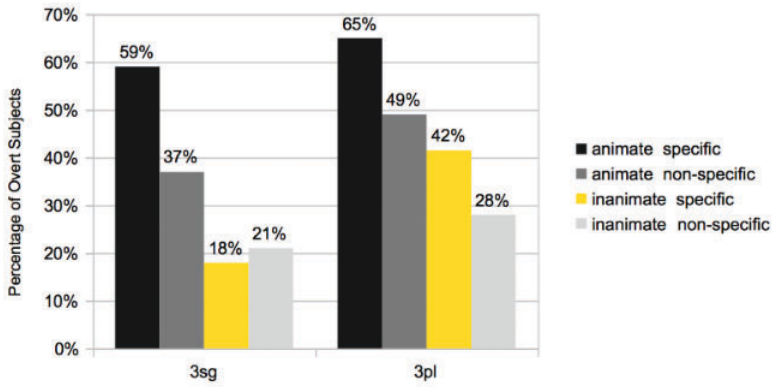
If for any individual *x* that belongs to the denotation α such that the proposition β applies to *x*, it is not possible that there is at least one individual *y*, which does not overlap with *x*, and the proposition β also applies to *y*. In (5), a mereological definition of individual is assumed, and also of complete overlap (see Link 1983, *inter alia*). This definition is a simplification of what was proposed by Kagan (2006).[10] For the purposes of this paper, we use this discourse-logical criterion (examples of applications of this criterion are provided in the Appendix), although we are aware that further research with respect to the discourse-semantic definition of specificity must be carried out.

5.2. Results

In Figure 7 below, the percentages of third person overt subjects divided according to their animacy, specificity and number are plotted.

10. Clearly, taking a binary operational criterion is not the optimal way to understand the effect of specificity on the choice of overt and null subjects. As explained above, specificity is a controversial notion. In particular we will not address the issue of different degrees of specificity. The aim here is to study and establish whether a distinction between non-specific and specific, as a first approximation, can exert any influence on the use of overt and null subjects in BP.

Figure 7 – Percentage Third Person Overt Subjects according Animacy and Specificity



Soares 2017, p. 53.

In Figure 7, there is a clear tendency: animate and specific antecedents are mostly retrieved by overt subjects while inanimate and non-specific antecedents are preferably recovered by null subjects.[11] The third person singular subject is more frequently overt when its antecedent is animate and specific, followed by the animate non-specific antecedent, and inanimate specific and non-specific antecedents are more or less at the same level; in the third person plural, animate specific overt subjects are close in frequency to other discourse persons, followed by animate non-specific subjects, then by inanimate specific subjects with slightly lower frequency and, at the bottom of the scale, inanimate non-specific subjects. The data were fitted in a generalized linear regression model with three fixed factors (Animacy, Specificity and Number) and one random factor (Interviewed Speaker) including random slopes for all fixed factors. Each fixed factor showed a main effect (Number ß:-1.500, SE: 0.516, z-value: -2.907, p-value: 0.00365; Animacy ß:-1.779, SE: 0.5449, z-value: -3.266, p-value: 0.00109; Specificity ß:1.2606, SE: 0.4262, z-value: 2.958, p-value: 0.00310). No interaction was even marginally significant, but the three-way interaction was (ß:-1.198, SE: 0.56797, z-value: -2.111, p-value: 0.03476), which is due to the number of specific inanimate overt

11. The same distribution is found with null objects, as pointed out in much previous research (see Cyrino *et al*.2000, Creus & Menuzzi 2004, *inter alia*).

subjects being lower than non-specific inanimate overt subjects in third person singular.[12]

5.3. Discussion

The results of the corpus research reported here converge in many aspects with those found in previous research: the two features studied have been shown to exert an effect on the data, namely inanimate and non-specific antecedents favor null subjects. Because such antecedents are strongly linked to the third person, these features might explain the data observed in this and in previous corpus studies. As suggested by the quantitative analyses carried out here, the features [–animate] and [–specific] seem to be suitable predictors of null subjects in the current grammar of BP, by being preferentially and sometimes obligatorily realized by an anaphoric null subject. At first sight, our data appear to strongly favor the HRH, since each factor individually showed a main effect. There seems also to be a gradual and cumulative effect of animacy and specificity, as previously found by Duarte, Mourão & Santos (2012) in a study on the same corpus. However, the three-way interaction, which reveals that specificity exerts a significant and different effect on inanimate singular subjects, suggests that part of the story remains to be told. This result aligns with those found by Othero & Spinelli (2019b), supporting at least partially the SGH, according to which no effect of specificity is found when the antecedents are [–semantic gender] (comparison between (D) and (E) in Table 1). We will address these data in section 7 and make a full proposal to account for these unexpected patterns.

In the next section, experimental evidence will be provided in order to test the robustness of the principal observations reported in this section. This procedure aims (i) to ensure that, in a controlled linguistic environment, these individual predictors can also be shown to

12. Additional statistical models with Period (1970s vs 1990s) as fixed factors showed no significant effect (p-value > 0.05). Model comparison using ANOVA revealed a significant advantage of the simplest model presented here (p-value < 0.05). Factors that might co-variate with the interviewed speaker (such as sex, age and level of formal education, for example) were not included in the models because they are presumably included in our random factor Interviewed Speaker.

be valid; (ii) to eliminate possible confounding factors that might have influenced the choice between overt and null subjects in the interviews (for instance, specific syntactic contexts); (iii) to study the intuitions of present day naive BP speakers (since the data reported is mostly from the 1990s) and (iv) to provide evidence from comprehension, rather than from production, that these features play a role in the grammar of anaphoric subjects in BP.

6. Experimental Evidence

Given the results obtained in the corpus research described in the previous section, two experiments were carried out in order to test whether the factors established in the corpus research can be considered individually relevant and to disentangle evidence that corroborates either the HRH or the SGH. In Experiment 1, animate vs. inanimate antecedents were investigated by varying the most accessible antecedent for null and pronominal subjects. The results of Experiment 1 show for the first time the effect of animacy on the use of the null and overt subjects in an experimentally controlled environment. In Experiment 2, we aimed to disentangle the two proposals with respect to specifity: while the HRH predicts effects of specificity across the board (which is not what we found in our corpus research), the SGH predicts that for some cases specificity should not exert any effect. We set up an experiment with nouns that are lexically [+ semantic gender] and manipulate the contextual interpretation to be [± specific]. Our results suggest that specificity affects the acceptability of null and overt subjects in BP, even when the antecedents are lexically [+ semantic gender].

6.1. Methodology

In both experiments reported here, we applied the same methodology: participants were asked to judge the acceptability of the sentences in the relevant context on a Likert scale, cf. Figure 8. They were told to use the full scale according to how *natural* "Normal" or *strange* "Estranha" the answer seemed in the context. In Experiment

1 the range was from 1 to 10; in Experiment 2, from 1 to 5. After judging the sentences' acceptability, the participants were asked about the interpretation of the relevant subject – null or overt – in a closed question task, cf. Figure 9.[13]

Figure 8 – Screen sample – Judgment Task

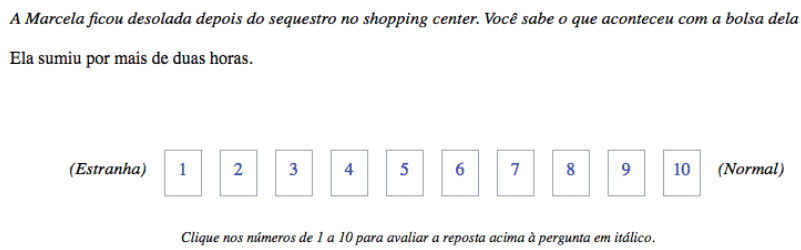
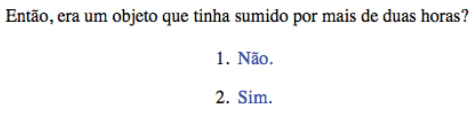


Figure 9 – Screen Sample – Closed Question Task



All participants voluntarily participated in the experiments on the IbexFarm platform (http://spellout.net/ibexfarm, Drummond, 2014). They filled in a basic information form that included a declaration of written consent and had 4 sentences to practice before starting the experiments, which took them around 30 minutes to complete. Data were only stored and analyzed when participants completed the experiment.

Among the items, four perfectly acceptable control sentences were inserted. Four control sentences that violate strong grammatical or pragmatic constraints were inserted at the end of the experiment,

13. The translations of materials in Figures 8 and 9 are respectively the following:
(i) Marcela was devastated after the kidnapping at the mall. Do you know what happened to her purse there?
(ii) It disappeared for more than two hours.
(iii) So was it an object that had disappeared for more than two hours?

in order to ensure that participants were attentive until the end of the experiment and to avoid ceiling effects in the experimental items. Beyond these items, fillers were inserted between other items. They were twice the number of target items.

6.2. Experiment 1 – Inanimate vs Animate Antecedents

This experiment was designed to test whether third person singular null and overt subjects show any preference for inanimate or animate semantic types of antecedents. In all sentences, the subject was informationally and structurally salient both in the context and in the question under discussion. Two binary Factors were tested: overt vs. null subject (Factor Subject) and inanimate vs. animate antecedent (Factor Animacy). The hypotheses were the following: (i) if Animacy plays a role in the use of null subjects in BP (cf. the HRH and SGH, for example), a significant interaction between the conditions is expected; the null subject should be rated more acceptable when referring to an inanimate antecedent and less acceptable in the case of animate antecedents; (ii) if Animacy has a relevant effect on the choice of overt or null subjects, but it is not as predicted by the Referential Hierarchy in (3), the null subject can be better rated when retrieving animate antecedents and the overt when co-referring to inanimate ones (as expected by some versions of the general theory of anaphora resolution presented in section 4); and (iii) if Animacy plays no significant role in the use of null and overt subjects, the overt subject should be preferred regardless of the semantic type of antecedent, since BP is generally taken to favor overt subjects over null subjects in the current stage of the language.[14]

14. As pointed out by Scott Schwenter (pers. communication) the frequency of combination of a given verb with a specific kind of referent might have influenced the results: supposing that a verb, such as *cair* "to fall", is by far more frequent with animate subjects than with inanimate ones, given the general frequency of overt subjects in BP, a collocation overt subject + *cair* could be at stake in the results found in this experiment. We agree that the role of frequency deserves further studied as regards the realization of overt and null subjects with some verbs, but this possible intervening effect is accounted for in the present analysis as a random Factor ("Item") in the mixed-effects model below. In future studies, once the frequency of given combinations is established, "frequency" can be run as main (intervening) Factor in the model.

6.2.1. Experimental Design

Twenty-four items were created, based on verbs that were found with null or overt inanimate subjects in our corpus data. A Google search confirmed that they were used with animate subjects as well. A context sentence was provided, such as (6) below. Following this sentence, an indirect question asking what happened either to an animate or to an inanimate referent was provided, as in Table 4. The answer could have either a null subject or an overt gender-marked subject pronoun, cf. Table 4 (masculine and feminine genders were counterbalanced across items). Afterwards, participants were asked if the subject of the relevant verb was either *uma pessoa* "a person" or *um objeto* "an object", cf. (7) below.

(6) A Maria estava muito irritada depois da reforma no apartamento.
"Maria was very stressed out after the refurbishment of the flat."

Tab. 2 – Stimuli – Experiment 1

	Animacy	Subject	Question	Answer
(A)	animate	null	A – Você sabe o que aconteceu com a colega de quarto dela lá? *A – Do you know what happened to her roommate there?*	B – Caiu da bancada. *B – [She] fell from the stand.*
(B)	inanimate	null	A – Você sabe o que aconteceu com a televisão de quarto dela lá? *A – Do you know what happened to her television there?*	B – Caiu da bancada. *B – [It] fell from the stand.*
(C)	animate	overt	A – Você sabe o que aconteceu com a colega de quarto dela lá? *A – Do you know what happened to her roommate there?*	B – Caiu da bancada. *B – She fell from the stand.*
(D)	inanimate	overt	A – Você sabe o que aconteceu com a televisão de quarto dela lá? *A – Do you know what happened to her television there?*	B – Caiu da bancada. *B – It fell from the stand.*

(7) Então, foi uma pessoa/um objeto que caiu?
"Was it a person/an object that fell?"
a. Sim.
"Yes."
b. Não.
"No."

Based on the results from the corpus study, the empirical predictions were the following: null subjects are preferentially used to retrieve inanimate antecedents (Condition B) over animate antecedents (Condition A); on the other hand, overt subjects are more acceptable when they pick up an animate antecedent (Condition C) than when they refer back to an inanimate one (Condition D). A significant interaction between Factors (Subject and Animacy) is thus expected.

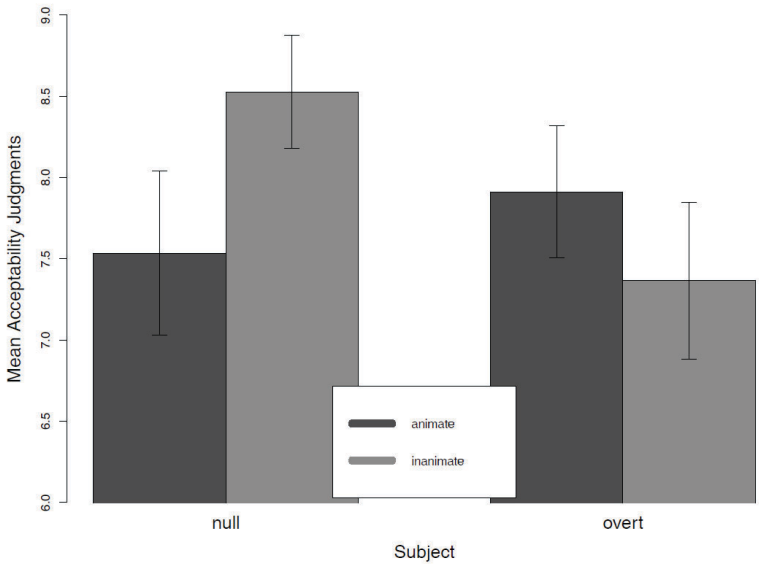
6.2.2. Participants and procedure

Twenty-nine participants from two southern states of Brazil (Rio Grande do Sul and Santa Catarina) took part in this experiment. They were on average 37.1 years old (+ or – 7 years). In the analysis, five participants were discarded because they either scored below 80% in the interpretation task (which had correct and incorrect answers) or rated ungrammatical control sentences above eight. The overall accuracy rate of the participants whose data were analyzed was 92%.

6.2.3. Results

The results of this experiment provide clear corroboration for the findings of the corpus studies and previous literature. As shown in Figure 10 below, the null subject is preferred when referring to an inanimate antecedent (averaging 8.5/10), while the overt subject is less acceptable (mean: 7.5/10). On the other hand, the overt subject is rated as more acceptable when the antecedent is animate (mean: 8/10) than when the antecedent is inanimate (7.45/10).

Figure 10 – Mean Acceptability Judgments according to Factors Animacy and Subject



As for the inferential statistical analysis, the data were analyzed using a log-linear mixed-effects model containing two Factors (Animacy and Subject) with two levels each and random effects (Participants and Items) including random slopes for all fixed factors (Barr et al. 2013). The full model is summarized in Table 5 below. The interaction between both Factors is significant.[15]

Table 3 – Log-linear mixed-effects model for Experiment 1

Factors	Estimate	Std. Error	T-value	Pr(>|t|)
(Intercept)	7.6099	0.3580	21.256	< 2e-16***
Subject	0.4313	0.2974	1.451	0.16165
Animacy	0.7787	0.4316	1.804	0.08397.
Subject:Animacy	-1.2624	0.4149	-3.043	0.00517**

15. The Factor Animacy is marginally significant: inanimate antecedents were generally judged slightly more acceptable than animate antecedents. This is irrelevant to the question at hand and is most likely simply due to the methods used for constructing the materials, where items were constructed on the basis of naturally occurring cases with inanimate antecedents.

6.2.4. Discussion

As in the corpus study reported in Section 5, in Experiment 1 Animacy influences the acceptability of overt or null subjects. The statistical analysis of Animacy turned out to be significant in two different approaches and methodologies. These results explain relative frequency of null subjects across discourse persons: the higher number of null subjects in the third persons is due to an inherent semantic feature of the antecedent, as predicted by the HRH and the SGH. We provided experimental evidence in present day BP (data collected in 2016-17) that corroborates the importance of this feature. However, the controversial feature [± specificity], which may be able to differentiate the hypotheses, remains to be investigated. The next section will focus on testing predictions made by each of these hypotheses with respect to the feature [± specificity], which has also been shown to be relevant in our corpus study.

6.3. Experiment 2 – Specific vs Non-Specific Antecedents

Taking into consideration the results obtained in the corpus research and the proposals made in the literature, Experiment 2 was designed to test whether specificity exerts effects on the acceptability of null and overt subjects. Especially, we controlled the feature [± semantic gender], by making certain that all nouns used in this experiment were [+ semantic gender] following the criteria proposed in Creus & Menuzzi (2004), Othero & Schwanke (2018), Othero & Spinelli (2019a,b). Two experimental Factors were tested: null vs. overt subject (Factor Subject) and specific vs. nonspecific (Factor Specificity) in a two by two design. The hypotheses are the following: (i) if the HRH accounts for the intuitions of BP speakers on the use of null and overt subjects, a significant interaction between Factor Subject and Factor Specificity will come up: null subjects retrieving non-specific antecedents will be judged better than those referring to specific antecedents, and overt subjects referring to specific antecedents will be preferred to those retrieving non-specific antecedents; (ii) if the SGH accounts better for the data, no interaction is expected and overt subjects are expected to be significantly better than null subjects, because all antecedents were [+ semantic gender]; and (iii) if general principles of anaphora resolution account for the data without any

effect of specificity, null subjects will be overall judged better than overt subjects without any interaction with Specificity, since the antecedents are highly salient in the context. The main effect of Specificity (without interactions) may be a side-effect of our materials, given that we are manipulating the contextual interpretation with verbal tense.

6.3.1. Experimental Design

Twenty-eight items were created for this experiment. As in the previous experiment, half of them were masculine, but in this case we used only plural nouns in order to easily manipulate specificity without changing the grammatical properties (singular vs. plural, for instance) of the noun (which might bias the results). Participants were presented with a sequence of two sentences, a context and a target sentence, cf. Table 4. In the context sentence, we manipulate the verb semantics (modal vs. episodic) to obtain a non-specific (A and C) vs. specific (B and D) reading for the NPs (Factor Specificity). The target sentence started with a null subject (A and B) or an overt pronoun (C and D) (Factor Subject). As in Experiment 1, after judging the sequences composed of these two sentences, participants answered to an interpretation task, directing their attention to the relevant referents, cf. (8).

Tab. 4 – Stimuli – Experiment 2

	Specificity	Subject	Context	Target
(A)	nospec	null	Os cantores devem ser reconhecidos pelo talento. *The singers have to be recognized by their talent.*	Podem ser presenteados no palco. *[They] may be presented [with a gift] on the stage.*
(B)	spec	null	Os cantores tinham sido reconhecidos pelo talento. *The singers had been recognized by their talent.*	Podiam ser presenteados no palco. *[They] might be presented [with a gift] on the stage.*
(C)	nospec	pro	Os cantores devem ser reconhecidos pelo talento. *A – Do you know what happened to her roommate there?*	Eles podem ser presenteados no palco. *They may be presented [with a gift] on the stage.*
(D)	spec	pro	Os cantores tinham sido reconhecidos pelo talento. *A – Do you know what happened to her television there?*	Eles podiam ser presenteados no palco. *They might be presented [with a gift] on the stage.*

(8) Quem deve ser/tinha sido reconhecido pelo talento?
"Who must be/had been recognized by their talent?"

a. Cantores.
"Singers."
b. Dançarinos.
"Dancers."

Based on the results from the corpus study, the empirical predictions were the following: (i) if the results found in the corpus are related to the specificity of the antecedent, we expect a significant interaction: Conditions A and D will be better rated than B and C; (ii) if the effect was due to the fact that in the corpus analysis we did not differentiate the animate antecedents that were [+ semantic gender] from those which were [– semantic gender], in the controlled environment where all of them are [+ semantic gender], no interaction between these factors should come up and overall overt pronominal subjects will be judged better than null subjects; finally (iii) if other effects due to discourse prominence were influencing our corpus data, in the controlled environment where all antecedents are topics and highly salient, null subjects will be overall better and no significant interaction is expected.

6.3.2. Participants and procedure

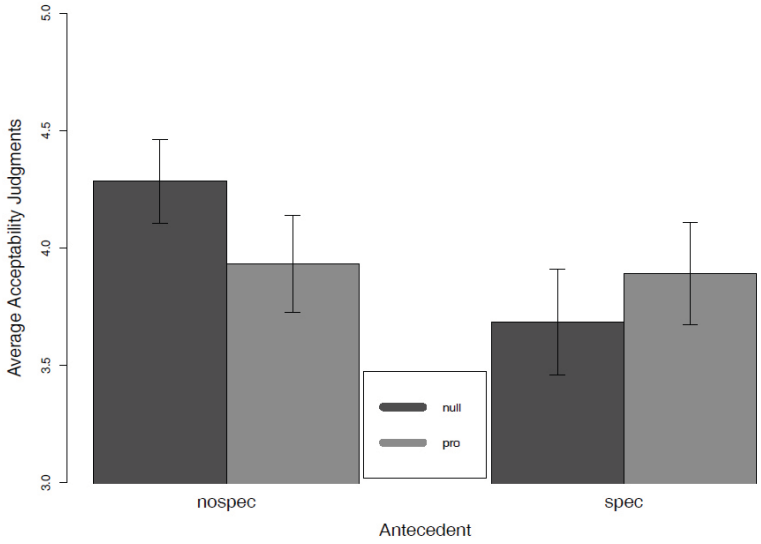
Twenty-seven participants from southern states of Brazil (Rio Grande do Sul and Santa Catarina) took part in this experiment. They were on average 27 years old (range ± 8 years). In the analysis, three participants were discarded because they rated ungrammatical control sentences above four. The Likert scale in this experiment went from 1 (Estranha) to 5 (Normal) in this experiment. Otherwise, the procedure was the same as in the previous experiment.

6.3.3. Results

The results summarized in the Figure 11 below corroborate the relevance of Specificity on the acceptability of null subjects.

While pronominal subjects were rated at the same level (mean 3.9), null subjects were considered better when referring to nonspecific antecedents (mean: 4.3) than when retrieving specific antecedents (mean: 3.7).

Figure 11 – Mean Acceptability Judgments according to Factors Specificity and Subject



As for the inferential statistical analysis, as in the previously reported experiment, the judgments were entered in a linear mixed-effects model with two fixed factors (Subject and Specificity) and two random factors (Participants and Items) including random slopes for all fixed factors (Barr et al. 2013). The maximal model is summarized in Table 5 below.

Tab. 5 – Log-linear mixed-effects model for Judgments in Experiment 2

Factors	Estimate	Std. Error	T-value	Pr(>|t|)
(Intercept)	4.2417	0.1531	27.707	< 2e-16***
Subject	-0.2753	0.1549	-1.777	0.08612.
Specifity	-0.5596	0.1695	-3.301	0.00297**
Subject: Specifity	0.5116	0.2354	2.173	0.04020*

As observed in Table 5, the main Factor Subject is marginally significant (a slight preference for null subjects over pronominal subjects) and the main Factor Specificity is significant. Crucially, the interaction between both fixed factors (Subject and Specificity) is significant.

6.3.4. Discussion

The results of Experiment 2 show a significant effect of the interaction between Specificity and Subject. This means that Specificity has an effect on the choice between null and overt subjects, favoring null subjects retrieving non-specific antecedents and overt subjects when the antecedent is specific. Also, the main factor Subject, which is marginally significant, suggests that an effect of anaphora resolution is also present in our data: the high salience of the antecedent favors null subjects. It seems that in the contexts where there is only one potential antecedent for anaphoric subjects, null subjects are preferred. These results call for an analysis based on the HRH with possibly some insights from anaphora resolution. However, rather than discarding the SGH, we will advance a proposal that tries to make sense of all data presented in this paper, by combining the HRH, the SGH and a general theory of anaphora resolution.

7. General Discussion

The imbalance in the distribution of third person overt and null subjects in BP seems to be explained by the features of the antecedents.[16] The results of testing animacy in both corpus and experiments in a diachronic and synchronic fashion converge towards the conclusion that animacy plays a decisive role in the use and acceptability of null

16. The HRH is proposed as a predictive directional theory of language change (Cyrino et al. 2000, Kato & Duarte2014), that is, the path by which the change is affecting the whole system. Considering the corpus data from two periods and the significant results from Experiment 1 and 2 (whose participants speak present day southern BP), the hypothesis suggested in this paper is that the features have a synchronic effect on the choice between overt and null subjects in BP, besides its possible diachronic effect. We are not going to pursue a diachronic theory for the data analyzed and presented here.

and overt null subjects in BP. At first sight, this is unexpected for some approaches based on a general theory of anaphora resolution (see Silverstein 1976, Dahl & Fraurud 1996, among others). We propose, though, that animacy is not a factor in the calculus of anaphora resolution in BP. We assume an idea from the SGH, according to which in BP the overt pronominal alternatives *ele* "he" and *ela* "she" are associated with their respective semantic genders as a lexical property, while null subjects are semantically unmarked alternatives. Therefore, pronominal overt and null subjects are not "competing" in the same conditions when the antecedent does not have a semantic gender, since overt pronominal subjects appear to be (or on the way to be) "specialized", similarly to the English pronouns "he" and "she". This proposal explains the effect of animacy that is found across the board in BP, because singular overt pronouns are avoided when the antecedent is [- semantic gender] and a default option (a null subject) is used.

The picture is slightly different with respect to the plural versions of the overt pronouns *eles* "they.MASC" and *elas* "they.FEM". As suggested by Othero & Spinelli (2019b), who attribute the idea to Sergio Menuzzi, these pronouns are the morphological exponent of the plural number feature, which, differently from gender, is a grammatical feature that mostly corresponds to a semantic feature. According to them, when the antecedent is third person plural, more overt pronouns are found. Our corpus research shows exactly the same pattern (see Table 2). Third person plural overt subjects are relatively more frequent than singular ones across the board (animate or inanimate; specific or non-specific) (see Figures 2 and 7). So, to conclude, in present-day BP the pronominal system has four morphological exponents: "*ele*/*ela*" for semantic gender, as proposed by the SGH, and "*eles*/*elas*" for semantic plural, as suggested by Menuzzi, Othero and Spinelli.

Where is thus the effect of specificity? Recall the three-way interaction found in our corpus research: specificity modulates the preference for overt subjects when the antecedents are [+animate] and [± plural] and when they are [+ plural] and [± animate], but not when they are [– plural] and [– animate]. That is, specificity plays a role when a BP pronoun, which is a morphological exponent of the semantic features [± plural] or [± semantic gender], is compatible with the antecedent. In such contexts, there is competition, because both

options are compatible: null subjects are preferred when retrieving non-specific antecedents and overt subjects are preferred otherwise. Recall also that, in Experiment 2, all our nouns were lexically marked [+semantic gender] and [+ plural]. These results suggest that the effect of specificity is restricted to cases where both options are possible. Assuming thus that the pronominal system in BP has a structure as described above, we propose that the effect of specificity is explained by standard anaphora resolution calculus, as in other Romance languages. As pointed out by Othero & Spinelli (2019a), specificity is a discourse property, not attached to a specific semantic class, but derived from the meaning of the sentence in the context. The natural locus for this feature is thus in the discourse properties that guide anaphora resolution.

As pointed out before, the findings in the literature on BP may be surprising for some of the theories of anaphora resolution based on the reverse mapping hypothesis. Depending on the concept of "salience" of the antecedent taken into account and its empirical coverage, non-specific antecedents should be taken to be at lower levels of salience scales. The fact that they are retrieved by less complex forms provides counterevidence against an anaphora-resolution-based approach for null subjects in BP and against the universality of the notions of salience/prominence generally accepted in the relevant literature. However, as predicted by Ariel (1990)'s notion of Accessibility, the linguistically relevant notion of salience must be established within the discourse. As shown in section 4, the highest level of salience is that of linguistically/discourse accessible antecedents (see Figure 6). Accepting this notion of salience, non-specific antecedents are expected to be at the top of their respective hierarchy, since they are exclusively intradiscursive entities (that is, they do not necessarily have reference in the physical context or in encyclopaedic context). This theory accounts for the data observed in this paper, since in the context of competition (when both forms are competing in the same conditions), non-specific antecedents are retrieved by the less complex form (the null subjects), while specific antecedents are more frequently retrieved by overt subjects. The effect of specificity is thus reduced to standard anaphora resolution calculus and has already been shown to exert effects on how BP speakers resolve anaphora (see Soares 2017).

References

ABBOTT, Barbara. 1995. 'Some Remarks on Specificity', *Linguistic Inquiry* **26(2)**: 341–347.

ALMOR, Amit. 1996. *Noun-phrase Anaphora and Focus*: *The Informational Load Hypothesis*. PhD thesis, Brown University, Los Angeles, CA, USA.

ARIEL, Mira. 1990. *Accessing Noun-Phrase Antecedents*. Routledge, London, UK.

ARIEL, Mira. 2001. Accessibility Theory: An Overview. In: SANDERS, Ted, J. SCHLIPEROORD, Joost; SPOOREN, Wilbert. (Eds.). *Text representation*. Springer Verlag. p. 29–87.

BAAYEN, R. Harald; DAVIDSON, Debra J.; BATES, Douglas M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* **59**: 390–412.

BARBOSA, Maria Pilar; DUARTE, Maria Eugênia L.; KATO, Mary A. 2005. Null Subjects in European and Brazilian Portuguese. *Journal of Portuguese Linguistics* **4(2)**: 11–52.

BARR, Dale J.; LEVY, Roger; SCHEEPERS, Christopher; TILY, Harry J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* **68(3)**: 255–278.

BATES, Douglas; MAECHLER, Martin. 2009. Lme4: linear mixed-effects models using S4 classes. R package version 0.999375-27.

BATES, Douglas; MAECHLER, Martin; BOLKER, Ben. 2011. Lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-41. http://CRAN.R-Project.org/package=lme4.

BATES, Douglas; MAECHLER, Martin; BOLKER, Ben; WALKER, Steve. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67(1)**: 1–48.

BITTNER, Dagmar. 2007. Influence of animacy and grammatical role on production and comprehension of intersentential pronouns in German L1-acquisition. In: BITTNER, Dagmar; GAGARINA, Natalia (Eds.) *Intersentential pronominal reference in child and adult language: Proceedings of the Conference on Intersentential Pronominal Reference in Child and Adult Language* (ZAS Papers in Linguistics 48)', ZAS, Berlin , Germany. p. 57–85.

CARMINATI, Maria N. 2002. *The processing of Italian subject pronouns*. PhD thesis, University of Massachussets, Cambridge, MA, USA, retrieved from http://scholarworks.umass.edu/dissertations/AAI3039345.

CHOMSKY, Noam. 1981. *Lectures on Government and Binding: The Pisa Lectures.* Foris, Dordrecht, Netherlands.

CREUS, Suzana; MENUZZI, Sergio. 2004. O papel do gênero na alternância entre objeto nulo e pronome pleno em português brasileiro. *Revista da ABRALIN* **3(1-2)**.

CYRINO, Sonia M. L.; DUARTE, Maria Eugênia L.; KATO, Mary A. 2000. Visible subjects and invisible clitics in Brazilian Portuguese. In: KATO, Mary A.; NEGRÃO, Esmeralda V. (Eds). 2000. p. 55–104.

DAHL, Östen; FRAURUD, Kari. 1996. Animacy in Grammar and Discourse. In: FRETHEIM, Thorstein; GUNDEL, Jeanette K. (Eds.) *Reference and Referent Accessibility*. Cambridge University Press, Pragmatics and Beyond New Series 38. p. 47–64.

DUARTE, Maria Eugênia L. 1993. Do pronome nulo ao pronome pleno: a trajetória do sujeito no português do Brasil In: ROBERTS, Ian; KATO, Mary A. (Eds.) *Português brasileiro: Uma viagem diacrônica. Homenagem a Fernando Tarallo.* Editora da Unicamp, Campinas, Brazil. p. 107–129.

DUARTE, Maria Eugênia L. 1995. *A Perda do Princípio "Evite Pronome" no Português Brasileiro*. PhD thesis, Universidade Estadual de Campinas, Campinas, SP, Brazil.

DUARTE, Maria Eugênia L. 2000. The loss of the 'avoid pronoun' principle in Brazilian Portuguese. In: KATO, Mary A.; NEGRÃO, Esmeralda V. (Eds). p. 17–36.

DUARTE, Maria Eugênia L. 2012. *O sujeito em peças de teatro (1833-1992): Estudos diacrônicos*. Parábola Editorial, São Paulo, SP, Brazil.

DUARTE, Maria Eugênia L. 2015. Avanço no estudo da mudança sintática associando a teoria da variação e mudança e a teoria de Princípios e Parâmetros. *Cadernos de Estudos Lingüísticos* **57(1)**.

DUARTE, Maria Eugênia L.; MOURÃO, Gabriela; GUIMARÃES, Luan. 2012. A retomada dos sujeitos proposicionais: categoria vazia ou demonstrativo? In: DUARTE, Maria Eugênia L. p. 69–82.

DUARTE, Maria Eugênia L.; MOURÃO, Gabriela; SANTOS, Heitor. 2012. Os sujeitos de terceira pessoa: revisitando Duarte 1993. In: DUARTE, Maria Eugênia L. p. 21–44.

DUARTE, Maria Eugênia L.; REIS, Eduardo Patrick R. d. 2018. Revisitando o sujeito pronominal vinte anos depois. *ReVEL* **16(30)**: 173–197.

DUARTE, Maria Eugênia L.; VAREJÃO, Filomena. 2013. Null subjects and agreement marks in European and Brazilian Portuguese. *Journal of Portuguese Linguistics* **12(2)**: 101–123.

DRUMMOND, Alex. 2014. *Ibex farm*. Online server: http://spellout.net/ibexfarm.

ENÇ, Mürvet. 1991. The semantics of specificity. *Linguistic Inquiry* **22(1)**: 1–25.

FALCO, Michelangelo. 2002. Specificity: the syntax/semantics mapping: A research project. Ms. Center for Mind/Brain Sciences (CIMeC), University of Trento.

GAGARINA, Natalia. 2007. The hare hugs the rabbit. He is white ... Who is white? Pronominal anaphora in Russian. In: BITTNER, Dagmar; GAGARINA, Natalia (Eds.) *Intersentential pronominal reference in child and adult language : Proceedings of the Conference on Intersentential Pronominal Reference in Child and Adult Language* (ZAS Papers in Linguistics 48)', ZAS, Berlin , Germany. p. 57–85.

GARNHAM, Alan. 2001. *Mental Models and the Interpretation of Anaphora*. Essays in Cognitive Psychology, Psychology Press, Hove, East Sussex, UK.

GRIES, Stefan T. 2015. The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora* **10(1)**: 95-125.

GRODZINSKY, Yosef; REINHART, Tania. 1993. The Innateness of Binding and Coreference. *Linguistic Inquiry* **24(1)**: 69–101.

GROSZ, Barbara J.; WEINSTEIN, Scott; JOSHI, Aravind K. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* **21**: 203–225.

GUNDEL, Jeanette; HEDBERG, Nancy; ZACHARSKI, Ron. 1993. Cognitive status and the form of referring expressions in discourse. *Language and Cognitive Processes* **69**: 274–307.

JAEGER, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* **59(4)**: 434–446.

KAGAN, Olga. 2006. Specificity as speaker identifiability. In: GYURIS, Beáta; KÁLMÁN, László; PIÑÓN, Christopher; VARASDI, Károly. (Eds.). *Proceedings of the ninth symposium on Logic and Language*. Eötvös Loránd University, Besenyotelek, Hungary.

KATO, Mary A.; DUARTE, Maria Eugênia L. 2014. *Restrições na distribuição de sujeitos nulos no Português Brasileiro*. *Veredas* **18(1)**: 1–24.

KATO, Mary A.; NEGRÃO, Esmeralda V. 2000. *Brazilian Portuguese and the Null Subject Parameter*. Vervuert-Iberoamericana, Frankfurt, Germany.

LINK, Godehard. 1983. The logical analysis of plurals and mass terms. In: BÄUERLE, Rainer; SCHWARZE, Christoph; VON STECHOW, Arnin. (Eds.) Meaning, use, and interpretation of language. Academic Press, Berlin, Germany New York, USA. p. 302–323.

MAYOL, Laia. 2010. Redefining salience and the Position of Antecedent Hypothesis: a study of Catalan pronouns. In: *University of Pennsylvania Working Papers in Linguistics*. University of Pennsylvania, Pennsylvania, USA.

MCENERY, Tony. 2000. Corpus-based and computational approaches to discourse anaphora. John Benjamins, Cambridge, MA, USA.

MONTALBETTI, Mario. 1984. *After Binding Properties. On the Interpretation of Pronouns*. PhD thesis, MIT, Massachusetts, MA, USA.

NEGRÃO, Esmeralda V. 1990. *A distribuição e a interpretação de pronomes na fala de crianças da escola pública*. (Ms.) University of São Paulo, São Paulo.

OTHERO, Gabriel A.; SCHWANKE, Carolina. 2018. Retomadas anafóricas de objeto direto em português brasileiro escrito. *Revista de Estudos da Linguagem* (UFMG) **26(1)**.

OTHERO, Gabriel A.; SPINELLI, Ana C. 2017. Sujeito Expresso e Oculto em Peças Teatrais Cariocas do Início do Século XXI (e Sua Relação com o Objeto Nulo em PB). Talk given at Seminários de Teoria e Análise Linguística, UFRGS.

OTHERO, Gabriel A.; SPINELLI, Ana C. 2019a. Sujeito pronominal expresso e nulo no começo do séc. XXI (e sua relação com o objeto nulo em PB). *Domínios De Lingu@gem* **13(1)**: 7–33.

_____. 2019b. Um Tratamento Unificado da Omissão e da Expressão de Sujeitos e Objetos Diretos Pronominais de 3a pessoa em Português Brasileiro. *Cadernos de Estudos Linguísticos* **61(1)**: 1–30.

R CORE TEAM. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

REULAND, Eric. 2001. Primitives of Binding. *Linguistic Inquiry* **32(2)**: 439–492.

SILVERSTEIN, Michael. 1976. Hierarchy of features and ergativity. In: DIXON, Robert M. W. (Ed.). Grammatical categories in Australian languages. UMOP, Australian Institute of Aboriginal Studies, Canberra, AUS. p. 112–171.

SOARES, Eduardo C. 2017. Anaphors in discourse: anaphoric subjects in Brazilian Portuguese. PhD Thesis. University Sorbonne Paris City, Paris, France.

SOARES, Eduardo C.; MILLER, Philip H.; HEMFORTH, Barbara. 2019. The effect of verbal agreement marking on the use of null and overt subjects: a quantitative study of first person singular in Brazilian Portuguese. *Fórum Linguístico*, v. 16, p. 3579-3600.

VON HEUSINGER, Klaus. 2002. Specificity and definiteness in sentence and discourse structure. *Journal of Semantics* **19**: 254–274.

VON HEUSINGER, Klaus. 2011. Specificity. In: VON HEUSINGER, Klaus; MAIENBORN, Claudia; PORTNER, Paul (Eds.). Semantics: An International Handbook of Natural Language Meaning. Vol. 2, Mouton de Gruyter, Berlin, Germany. p. 11–34.

Recebido em: 16/06/2019
Aprovado em: 05/03/2020

Acknowledgements

Research reported in this paper was mostly carried out while the first author was a PhD student at the University of Paris VII Denis Diderot supervised by the other two authors and Sergio Menuzzi. Over this period, this study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 (trial number: 0628/14-0). Some parts of this text first appeared in Soares (2017); the conclusions and crucial evidence (Experiment 2) presented in this paper had not been published before. The final version of the paper was written while the first author was holding a post-doc position in the Federal University of Santa Catarina in the National Program of Post-Doctorate (PNPD) of the CAPES (trial number: 88882.316438/2019-01 up to February, 2019 and 88887.466656/2019-00 from November, 2019 on). We are thankful to the labs where this study was developed (the CLILLAC-ARP and the LLF at University of Paris VII Denis Diderot), to the LabLing at UFSC, and to their members for their support. We are also thankful to the researchers who made comments on preliminary versions of this paper: Jeffrey Runner, Scott Schwenter, Sergio Menuzzi, Gabriel Othero and Anne Abeillé, and two anonymous reviewers for DELTA. Finally, we are thankful to the participants who answered the online questionnaires.

Appendix

Table 6 – Sample of Specific and Non-Specific Antecedents in the NURC-RJ Corpus

Class	Spec	Example
quantifier/ quantified NPs	-	[tudo] foi penhorado "*Everything* was pawned"
indefinite NPs	-	[uma pessoa que tá querendo fazer eletrônica] vai ter cálculo. "*a person who wants to do Eletronics* will have Calculus."
mass nouns	-	[barulho como existe no Rio de Janeiro] eu acho que dificilmente se encontrará em outra cidade "*noise as it exists in Rio de Janeiro* I think that one will hardly find in any other city."
plural NPs	-	há [professores] dando doze horas diárias "There are *teachers* giving twelve daily hours [of classes]"
definite NPs	-	[O indivíduo] tinha até uma escala profissional. "*the individual* had even a professional scale"
negative NPs	-	[ninguém] quer pensar nisso "*nobody* wants to think about this"
negative NPs	+	eu não peguei [nenhuma dessas professoras] "I haven't had *any of these teachers*"
proper names/ definite descriptions	+	Conheço mais ou menos [o sindicato dos professores] "I know more or less *the teacher's union*"
indefinite NPs	+	tenho parente inclusive nessa situação, que é [um indivíduo que trabalhava com mecânica de automóveis] "I have a relative in this situation, who is *an individual that worked on auto mechanics*"
quantified NPs	+	[todos os cursos que anunciavam no Diário de Notícias] receberam a comunicação de que tinham que comparecer lá "*all courses that advertise in the Diário de Notícias* received the notification that [they] had to go/attend there."