

Referring to discourse participants in Ibero-Romance languages

Edited by

Pekka Posio

Peter Herbeck

Open Romance Linguistics 4



Open Romance Linguistics

Editors:

Lorenzo Filipponio (HU Berlin)

Richard WALTEREIT (HU Berlin)

Esme Winter-Froemel (Julius-Maximilians-Universität Würzburg)

Anne C. WOLFSGRUBER (HU Berlin)

In this series:

1. Kocher, Anna. Complementizers on edge: On the boundaries between syntax and pragmatics in Ibero-Romance.
2. Bullock, Barbara E. Cinzia Russi & Almeida Jacqueline Toribio (eds.). A half century of Romance linguistics: Selected proceedings of the 50th Linguistic Symposium on Romance Languages.
3. Fliessbach, Jan. The intonation of expectations: On marked declaratives, exclamatives, and discourse particles in Castilian Spanish.
4. Posio, Pekka & Peter Herbeck (eds). Referring to discourse participants in Ibero-Romance languages.

ISSN (print): 2940-1097

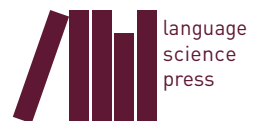
ISSN (electronic): 2940-1100

Referring to discourse participants in Ibero-Romance languages

Edited by

Pekka Posio

Peter Herbeck



Pekka Posio & Peter Herbeck (eds.). 2023. *Referring to discourse participants in Ibero-Romance languages* (Open Romance Linguistics 4). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/376>

© 2023, the authors

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/>

ISBN: 978-3-96110-416-1 (Digital)

978-3-98554-074-7 (Hardcover)

ISSN (print): 2940-1097

ISSN (electronic): 2940-1100

DOI: 10.5281/zenodo.8123675

Source code available from www.github.com/langsci/376

Errata: paperhive.org/documents/remote?type=langsci&id=376

Cover and concept of design: Ulrike Harbort

Typesetting: Helene A. von Martens

Proofreading: Amir Ghorbanpour, Elliott Pearl, Georgios Vardakis, Janina Rado,

Jeroen van de Weijer, Lachlan Mackenzie, Neneng Sri, Rebecca Madlener

Fonts: Libertinus, Arimo, DejaVu Sans Mono

Typesetting software: $\text{X}_{\text{L}}\text{A}_{\text{T}}\text{E}_{\text{X}}$

Language Science Press

xHain

Grünberger Str. 16

10243 Berlin, Germany

<http://langsci-press.org>

Storage and cataloguing done by FU Berlin

Freie Universität  Berlin

This book is dedicated to the memory of
Kimberly Geeslin (1971–2023), whose research and
kindness will continue to inspire linguists across the globe.

Contents

- 1 Referring to discourse participants in Ibero-Romance languages:
Introduction
Peter Herbeck & Pekka Posio 1

- I Variable expression of subjects and objects

- 2 A cross dialectal comparison of first person singular subject pronoun
expression in Southern Arizona and Southeast Texas
Ryan M. Bessett 25

- 3 Variable indirect object pronoun expression: A usage-based analysis
of Galician and Spanish
Javier Rivas & Esther L. Brown 41

- 4 Variable subject expression in second language acquisition: The role
of perseveration
Kimberly L. Geeslin, Thomas Goebel-Mahrle, Jingyi Guo & Bret Linford 69

- II Between personal and impersonal

- 5 Portuguese *a pessoa* and *uma pessoa*: Emerging inclusive
impersonals
Eduardo Tadeu Roque Amaral & Wiltrud Mihatsch 107

- 6 Personal uses of impersonalizing strategies: Hybrid constructions
with *a gente* and *se* in rural Madeiran Portuguese varieties
Yoselin Henriques Pestana 147

- 7 On pronominal uses of *geral* in Brazilian Portuguese
Juanito Ornelas de Avelar 177

Contents

8	Toward the prevalence of a personal use of impersonal <i>uno</i> in Colombian Spanish	
	Rafael Orozco, Luz Marcela Hurtado & Marianne Dieck	203
9	Discourse participants in impersonal constructions: The case of first and second person object pronouns with Spanish non-anaphoric third person plural subjects	
	Émeline Pierre & Barbara De Cock	237
III Referring to the addressee		
10	Referring to discourse participants in European Portuguese: The form of address <i>o senhor</i>	
	Isabel Margarida Duarte & Maria Aldina Marques	273
11	Variation and change in reference to discourse participants in Catalan parliamentary debate (1932–1938 and 1980–2020)	
	Neus Nogué-Serrano & Lluís Payrató	307
	Index	345

Chapter 1

Referring to discourse participants in Ibero-Romance languages: Introduction

Peter Herbeck

University of Vienna & University of Wuppertal

Pekka Posio

University of Helsinki

1 Introducing reference to discourse participants

The current volume aims at presenting a panoramic view of recent advances in the study of reference to discourse participants in Ibero-Romance languages and to search for connections between phenomena that have previously been studied in isolation. It brings together contributions on person reference in Ibero-Romance languages that go beyond the well-established field of study focusing on the expression vs. non-expression of subject pronouns. Several corpus studies on Ibero-Romance languages have shown that the phenomena affecting the expression of subject pronouns transcend the traditionally established factors like morphological ambiguity of the verb, emphasis, contrast, and topic continuity. Besides additional factors like tense/aspect/mood marking, subjectivity, and the degrees of fixation of subject-verb combinations with particular verb forms (see, e.g., Carvalho et al. 2015 and Posio 2018 for an overview), it has been demonstrated that grammatical person is one of the most significant variables affecting subject expression, and that the factors influencing the expression of anaphoric third person subjects differ from those conditioning deictic first- and second-person subjects. The question of overt/covert alternations in syntax has also been



addressed in research focusing on the encoding of impersonal reference in languages with variable subject expression, such as new pronoun-like referential devices emerging from noun phrases, and the uses of impersonal constructions like passives and impersonals formed with the reflexive clitic *se* to express personal reference. These devices allow speaker-inclusive and/or addressee-inclusive interpretations to different degrees, depending on the variety and the type of nominal expression encoding impersonality. Moreover, while approaches to variable pronoun expression have traditionally focused on nominative subjects, recent research has opened new avenues to studying to what extent the same or different factors affect the expression of other forms such as direct (accusative) objects and indirect (dative) objects encoding experiencers and recipients.

This introductory chapter provides an overview of the topics of this volume and emphasizes the relevance of the studies included against the background of previous research on pronouns and other referential devices. We use this latter term, as suggested by Kibrik (2011), to subsume bound person marking morphemes, pronouns and noun phrases used to establish reference, in order to highlight the fact that the boundaries of the traditional categories “pronoun” and “noun phrase” are flexible and some semi-grammaticalized items display properties of both (e.g., Portuguese *a gente* ‘the people; we’ and *a/uma pessoa* ‘the/a person; one’). By focusing on discourse participants, i.e., the speaker(s) and the addressee(s), this volume takes into account the findings of previous research regarding the similarities between these categories as well as differences with regard to the third person.

For instance, while third-person referents need to be introduced into the discourse and the choice of referential devices (null pronouns, agreement morphemes, weak pronouns, demonstratives, strong pronouns, definite and indefinite NPs, lexical expressions) referring to them is therefore heavily influenced by information-structural categories, such as referentiality, topic continuity, and accessibility (Givón 1983, Levinson 1987, Ariel 1990, Gundel et al. 1993, among many others), discourse participants are, in principle, always accessible by virtue of being present in the communicative situation. To use the file card metaphor (Reinhart 1981, Heim 1982), the cards for the speaker and addressee are always on top of the file and, thus, available as topics (Erteschik-Shir 2007: 45–46). For first-person pronominal forms, it has therefore been argued that, apart from topic and referential continuity, factors in determining subject pronoun expression include subjectivity and the epistemic stance of the speaker (see, e.g., Enríquez 1984, Aijón Oliva & Serrano 2010, Posio 2011, Herbeck 2021), probably depending on the type of verb lexeme or individual verb forms. For address forms in morphological second and third person, use of a subject pronoun or noun phrase is governed by factors related to formality of the discourse, (positive and negative)

politeness and (inter)personal relations between the interlocutors (see De Jonge & Nieuwenhuijsen 2012 and Uber 2016 for an overview and references).

However, it is also important to note that grammatical and referential categories do not always coincide: canonically, speakers refer to themselves with first-person pronouns and to their addressees by second-person pronouns, but in practice, there is ample variation in the choice of referential devices. The use of noun phrases formally in the third person to refer to first and second person entities is a well-known case, as is the emergence of formally third-person expressions (like the Spanish impersonal pronoun *uno* ‘one’ or the Portuguese noun phrases *a pessoa* ‘the person’; Orozco et al. 2023, Amaral & Mihatsch 2023 [this volume]) that establish reference to the speaker, either through generalization involving the speaker or directly as a first-person reference. Another example of such a development is the personal use of *a gente* ‘the people’ that has developed into an impersonalization strategy and, subsequently, into a first-person plural pronoun, in particular in Brazilian Portuguese (e.g., Lopes 2004, Zilles 2005), or the appearance of “new” address forms based on noun phrases in European Portuguese (see §4).

This volume is structured into three thematic blocks addressing the before-mentioned topics: Part I, *Variable expression of subjects and objects* contains the contributions from Ryan Bessett examining first-person singular (1SG) subject expression in two varieties of Spanish spoken in the United States, Esther Brown and Javier Rivas studying the expression of first- and third-person indirect object pronouns in Spanish and Galician, and Kimberly Geeslin, Tom Goebel-Mahrle, Jingyi Guo, and Bret Linford, whose study focuses on the role of perseveration in the acquisition of variable subject expression in Spanish as a second language.

Part II, *Between personal and impersonal*, contains the papers by Eduardo Amaral and Wiltrud Mihatsch who study emerging impersonal constructions with the lexeme *pessoa* ‘person’ in Portuguese, Yoselin Henriques Pestana’s paper on personal uses of impersonal constructions in rural Madeiran Portuguese, Juanito Ornelas de Avelar’s paper on the pronominal uses of the word *geral* ‘general’ in Brazilian Portuguese, the contribution by Rafael Orozco, Luz Marcela Hurtado and Marianne Dieck who study the personal uses of the impersonal pronoun *uno* ‘one’ in Colombian Spanish, and the paper by Émeline Pierre & Barbara de Cock on the use of object discourse participant pronouns in a third person plural impersonal construction.

Part III, *Referring to the addressee*, contains Aldina Marques and Isabel Margarida Duarte’s study of the address form *senhor* and its variants in European Portuguese and Neus Nogué Serrano and Lluís Payrató’s paper on changing address forms in Catalan parliamentary discourse.

With this selection of papers, our intention is not only to shed light on the different ways of referring to discourse participants in Ibero-Romance languages, but also to open new perspectives to phenomena related to person reference and inspire future research on reference and discourse. This introductory chapter provides an overview about the state-of-the-art of the three thematic blocks and outlines the research questions addressed in the contributions of the current volume.

2 Variable expression of subjects and objects

The variable expression of subjects has received considerable attention in generative syntax as well as functional and constructional approaches. While in generative studies of the so-called *pro*-drop parameter (Chomsky 1981, 1982, Rizzi 1982, Solà 1992, Barbosa 1995), the focus has been on the formal properties of a language system that make null pronouns possible, functional approaches have focused on the conditions under which subjects are expressed in actual language use (see, e.g., Enríquez 1984, Bentivoglio 1987, Cameron 1993, Morales 1997, Otheguy et al. 2007, Travis & Torres Cacoullos 2012, Posio 2011, 2018, among many others). Thus, the following represent some of the leading questions in studies approaching variable subject expression:

1. Which factors determine subject pronoun use and what is the internal ranking of these factors?
2. How does usage frequency (e.g., of verbs and verb forms) relate to the expression of pronominal subjects?
3. Is subject expression governed by same or different factors across languages and language varieties?

With respect to the first question, there is a vast body of research examining subject expression from various theoretical and methodological perspectives. The following non-exhaustive list presents the most prominent factors affecting subject expression that have been singled out in previous studies:

1. (morphological) ambiguity (e.g., Hochberg 1986) emphasis and contrast (e.g., Rigau 1989, Luján 1999, Mayol 2010)
2. referential and/or topic continuity (e.g., Bentivoglio 1983)

3. switch reference (e.g., Cameron 1993, Silva-Corvalán 2001, Travis & Torres Cacoullos 2012)
4. person/number; tense, aspect, mood; verb type (e.g., Enríquez 1984, Morales 1997, Silva-Corvalán 2001, Otheguy et al. 2007, Orozco & Guy 2008, Posio 2011)

A common trend observed in most studies is that topic and referential continuity favor null subjects, while shifting topics and switch reference favor the presence of overt subject pronouns. With respect to morphological ambiguity, one hypothesis that has been discussed is whether syncretic verb forms trigger higher subject expression rates than non-syncretic ones, i.e., the so-called “functional hypothesis” (Hochberg 1986). However, the results of different studies are not homogenous, some studies supporting and others not providing direct evidence for the relevance of syncretism between first, second and third person (see Posio 2018 for discussion). With respect to emphasis and contrast, several studies claim that strong subject pronouns encode different flavors of these notions. Thus, Rigau (1989) makes a distinction between weakly and strongly emphasized strong pronouns and Mayol (2010) claims that strong pronouns encode different types of contrastive topics.

Tense, aspect and mood marking on verbs is closely related to the factor of morphological ambiguity: certain verb forms have ambiguous person-marking morphology in first-person singular (1SG) and third-person singular (3SG), e.g., in Catalan, European Portuguese, and Spanish. Thus, imperfect, conditional, and subjunctive paradigms display a syncretism between 1SG and 3SG in these languages. Furthermore, phonological processes lead to a higher number of ambiguous verb endings in some varieties of the Ibero-Romance languages, which in turn might influence subject expression rates. Hochberg (1986) investigated how word-final /s/ deletion correlated with subject expression in Puerto Rico Spanish, observing that ambiguity between second-person singular (2SG) and 3SG verb forms, e.g., in the indicative present, or between 1SG, 2SG, and 3SG, e.g., in indicative imperfect correlated with higher subject expression frequencies. More recently, Herbeck (2022) has observed that some Valencian Catalan varieties display high frequency of 1SG subject pronouns in the present perfect, in which the 1SG and 3SG auxiliary has the same form. However, the relevance of the functional hypothesis is not confirmed by some studies with different type of data (Ranson 1991, Cameron 1992, 1993, cf. Posio 2018: 290). Silva-Corvalán (2001), on the contrary, argues that the discourse function of TAM markings (i.e., event foregrounding vs. backgrounding), rather than their surface ambiguity, is the decisive factor

explaining variable subject expression in different tenses and moods. However, Brazilian Portuguese provides strong evidence in favor of the role of morphological syncretism. Brazilian Portuguese has developed from a consistent to a partial *pro*-drop language (in the sense of Holmberg 2005; see Kato 1999, Barbosa 2009; collected works in Kato & Negrão 2000). Interestingly, the loss of some properties of a (consistent) null subject language and syncretism in verb morphology has consequences, not only for the expression of referential subjects, but also for the encoding of impersonal (see, e.g., Carvalho 2019 for impersonal 3SG null subjects) and personal subjects (e.g. *a gente* referring to first-person plural [1PL]; see Lopes 2004).

In the case of person/number, one important question is whether subject expression behaves similarly in different grammatical persons and numbers. Thus, the difference between deictic first and second person and discourse anaphoric third person might affect the relevance or weight of factors such as topic continuity, switch reference and (morphological) ambiguity. In fact, the study of only one grammatical person at a time has turned out to be a very fruitful approach (see Shin & Otheguy 2005, Travis & Torres Cacoullos 2012, Shin 2014, Bessett 2023 [this volume] for Spanish; Herbeck 2022 for Valencian Catalan). This issue takes us to one of the main issues of the present volume: the question of what the factors governing subject expression are in the case of reference to discourse participants and whether they differ from those that have been observed to hold for discourse anaphoric persons. In fact, for devices expressing reference to the speaker, it has been argued that, rather than continuity, subjectivity is a major factor for the use of an implicit or explicit subject pronoun (see, e.g., Enríquez 1984, Aijón Oliva & Serrano 2010, Posio 2011, 2018, Hennemann 2016, Herbeck 2021).

However, the influence of grammatical person cannot be considered in isolation but must be examined together with verb type and specific verb forms, taking into account that these might not behave uniformly in different varieties and languages. This brings us to the points (ii) and (iii) above: Subject expression has been argued to show considerable variation if specific verb forms (and not only verb types) are considered. With respect to verb type and semantics, it has frequently been observed that verbs of cognition have particularly high 1SG subject expression rates in Spanish (e.g., Enríquez 1984, Morales 1997, Posio 2011). However, the issue is complex because the group of verbs of mental activity is not homogenous, some verbs of cognition favoring 1SG subject expression (e.g., [*yo*] *creo* ‘I think’), others disfavoring it in Peninsular Spanish (e.g. [*yo*] *sé* ‘I know’; Herbeck 2021). Furthermore, Aijón Oliva & Serrano (2010) argue that [*yo*] *creo* has higher subject expression rates when used as a verb expressing

the personal opinion of the speaker and lower rates if it has a mere epistemic function. This mirrors Enríquez's (1984) classification of verbs of cognition into verbs expressing a mental state/activity and those expressing opinions: the latter class has the highest 1SG subject expression rates in her study. The question thus arises whether verb semantics, the function of a verb in context (expression of epistemicity, evidentiality, subjectivity, opinions, etc.), the concrete verb form, or a combination of all these factors affects the expression of pronouns referring to discourse participants.

The question of semantic factors influencing expression of pronouns is especially interesting if a cross-linguistic perspective is integrated into the picture. As Posio (2018) notes, there is considerable variation with respect to which verbs and verb forms trigger high and low subject pronoun rates in different Ibero-Romance languages and varieties. For example, with *decir* 'say', subject expression has been observed to have a low frequency in Peninsular Spanish (e.g., Posio 2013, 2014), but a high frequency in Colombian Spanish (Orozco 2015: 25). Furthermore, Posio (2013, 2014) observes in his study of Peninsular Spanish and European Portuguese that 1SG subject expression is particularly high with the verb form *creo* 'I think' in the former language, while the verb form *digo* 'I say' has low to average expression rates. In European Portuguese, on the contrary, the equivalent form *digo* 'I say' has high subject expression rates, but the verb form *acho* 'I think' is associated with low rates. In a similar vein, Herbeck (2022) observes that 1SG subject pronouns are frequently expressed with the verb *dir* 'say' in Valencian Catalan varieties, while *creure* 'think' has only average rates of 1SG subject expression, unlike in Peninsular Spanish. Posio (2013, 2014, 2015) argues that certain verb forms and their subject pronouns have become formulaic sequences due to high frequency of use in determined discourse context and correlated routinization. The degree of fixation of a 1SG or 2SG subject pronoun + verb sequence might differ across languages and varieties, the sequence [*yo*] *creo* showing a higher degree of grammaticalization and conventionalization in Spanish than the corresponding sequence in EP. This raises the question of the relation between frequency, the expression of reference to discourse participants, and verb forms encoding epistemic stance, personal opinions and speech acts – notions that are particularly relevant in speaker/hearer interactions.

Lastly, while subject expression (as null or strong pronouns) is a topic that has received considerable attention from both theoretic and data-based perspectives, the question of whether variable object pronoun expression as weak or strong pronouns is governed by the same or different factors as subject pronoun expression still awaits further research. This issue is particularly relevant for dative experiencers that have been argued to display several subject properties (see

Masullo 1993). Thus, with psych-verbs selecting a dative experiencer, the question is whether the same factors govern expression of strong speaker/addressee pronouns vs. clitic ones as with nominative subjects and with dative objects (see Rivas & Brown 2023 [this volume] for related discussion). The issue is interesting in the light of functional approaches positing that certain scales, such as topicality, animacy, and grammatical function (see Givón 1983 and related work) affect the choice of referential device. Thus, it has been argued that topic continuity favors subject omission (Givón 1983, Bentivoglio 1983, among many others). Topicality is in turn favored with animate referents (if compared with inanimates) and with nominative subjects if compared with dative and accusative NPs (see Givón 1983). While dative experiencers are highly animate and, thus, predicted to be high on the scale of topicality, they display case marking typically associated with lower topicality than nominative subjects.

The thematic part *Variable expression of subjects and objects* of the present volume deals with several questions discussed above. First, the paper by Rivas & Brown (2023 [this volume]) addresses the issue whether object expression is affected by same or different factors as subject expression. They analyze which factors influence weak (e.g., *me*, *te*) and strong (e.g., *a mí*, *a ti*) indirect object pronoun expression in Spanish and Galician, offering insights into two closely related Ibero-Romance languages. The authors show that while expressing the object as a strong pronoun is more frequent in Spanish than in Galician, it is affected by the same syntactic, discourse, and interactional factors in the two languages. Thus, their study indicates that expression of 1SG object pronouns is favored in utterance initial position, in constructions with *gustar*-type verbs, when primed by previous mention, and in non-continuous contexts.

The next paper of this part by Bessett (2023 [this volume]) asks whether reference to the speaker presents differences between two geographically close varieties of Spanish. Bessett examines 1SG subject expression in the Spanish spoken in two communities located on the border between Mexico and the US: Southern Arizona and Southeast Texas. The quantitative results of the study show that 1SG subject expression occurs with similar rates in the two samples representing the two varieties and is conditioned by similar factors (switch reference; clause type; tense, mood, and aspect; and whether the verb is reflexive or not). In addition to contributing to the general discussion on factors affecting subject expression, the paper provides new data from these borderland varieties of Spanish.

Lastly, while the role of perseveration and priming has been studied in the context of subject expression (see, e.g., Travis & Torres Cacoulios 2018), its role in the acquisition of subject expression by L2 speakers of Spanish still awaits wider

research. The paper by Geeslin et al. (2023 [this volume]) investigates the acquisition of subject expression in Spanish as a second language by native speakers of English. In particular, their study focuses on how far perseveration, i.e., the use of the same subject form that has been used in the context preceding the target form, influences subject expression across several proficiency levels. This factor is examined in interaction with other linguistic factors (form of the prime – null, overt, or lexical noun phrase – switch reference, gender continuity, TAM continuity) in a written contextualized preference task. Focusing on third person subjects, the study sheds new light on the role that the form of the previous mention of the referent of a (null or overt) subject plays in conditioning subject form in second language acquisition.

3 Between personal and impersonal

During the last decade there has been an increasing interest towards constructions expressing impersonal, generic, or vague reference to human participants within functional-typological linguistics (see, e.g., the papers in Malchukov & Siewierska 2011). The notion of impersonality is used in various ways in linguistics. Traditionally, Ibero-Romance linguistics has distinguished between syntactic and semantic impersonality (Fernández Soriano & Táboas Baylín 1999: 1723). Syntactically impersonal constructions can be defined as those where the verb does not agree with a subject or the overt subject is lacking completely, as is the case with meteorological verbs (e.g., Spanish *llueve* ‘it rains’) or existential verbs (e.g., Portuguese *há ovos no frigorífico* ‘there are eggs in the fridge’). Impersonal constructions with expletive or “dummy” subjects are rare in Ibero-Romance varieties, although they are found in varieties of Portuguese (Carrilho 2005) and Dominican Spanish (Toribio 2000).

Semantically impersonal constructions can be defined as those where the subject argument is reduced in referentiality: there is a subject in the verbal construction, but it is either non-canonical in the sense that it does not coincide with the agent of the depicted action (e.g., Spanish “passive” *se*-constructions like *se venden coches* ‘cars are sold’ where *coches* ‘cars’ is formally the subject) or it is referentially vague in the sense that it does not point at any particular participant (e.g., Spanish *uno* ‘one’, as in *uno no sabe qué hacer* ‘one doesn’t know what to do’). In the latter case, typical sources of impersonality are personal pronouns used for non-canonical reference: for instance, 2SG pronouns and verb forms are used in many languages to refer vaguely to ‘anyone in general or in a given situation’, and the third person plural is often used to refer vaguely to ‘people in

general'. A further distinction can be made between generic and episodic readings of impersonal subjects: for instance, impersonal second person singulars can, in most cases, only be used in generic sentences that are not anchored into any specific point in time or place, while third person plurals are found in both generic and episodic sentences, i.e., sentences referring to actions taking place in a given time and space. In formal approaches, a similar distinction is made between "generic" and "arbitrary" pronouns, i.e., those referring to 'anyone'.

The subtype of impersonal constructions that is of interest to the current volume has been referred to as reference impersonals or R-impersonals (Siewierska & Papastathi 2011, Malchukov & Ogawa 2011, Malchukov & Siewierska 2011) or human impersonals (Cabredo Hofherr 2008), highlighting the fact that the source of impersonality is a reduction in the referentiality of the subject argument in these constructions. For a more elaborate typology of human impersonals, see Gast & van der Auwera (2013). Crucially, the subject (or agent) of the depicted action is always human (or at least construed as human), and non-animate or non-human animate participants are not acceptable without very specific context (e.g., Portuguese *?/*ladra-se muito à noite* 'one barks a lot at night', where the intended referent would be the neighborhood dogs). What makes these constructions interesting for the current volume is the two-way relationship between personal pronouns and other referential devices used to refer to discourse participants: not only do deictic pronouns like the second-person singular acquire generic and impersonal uses, but originally impersonal forms like the *se*-constructions or Portuguese *a gente* and *a pessoa* (see Henriques Pestana 2023 [this volume], and Amaral & Mihatsch 2023 [this volume]) also develop uses where their primary referential range is the speaker or a group including the speaker.

The reduction of referentiality found in human impersonal constructions does not mean that their reference is completely arbitrary: rather, the choice of the human impersonal construction, as well as contextual elements such as locative expressions, typically restrict the scope of possible referents of the constructions (i.e., their referential range; Posio & Vilkkuna 2013). Thus, human impersonals deriving from personal pronouns such as 2SG or the third person plural (3PL) typically maintain part of their "original" referential properties. For instance, 2SG used impersonally implies that the intended referent is singular and may coincide with the speaker or the addressee, whereas impersonal 3PL, at least in most cases, exclude both speech act persons from their referential range.

Since Ibero-Romance languages display a wide range of human impersonal constructions – including reflexive-based *se*-constructions, *one*-impersonals like Spanish *uno*, pronoun-based like 2SG and 3PL, and noun-based like Portuguese

a pessoa – the constructions are specialized to express different kinds of referential range. For instance, non-anaphoric 3PL with no expressed subjects – the topic of Pierre & De Cock’s paper (2023 [this volume]) – generally expresses a referential range that excludes the speaker and the addressee. However, 1SG and 2SG pronouns can occur in these constructions as direct or indirect objects. Pierre & De Cock argue that the referential vagueness of the subject of these constructions makes the object arguments more prominent.

In the European language area, there is a widespread construction type known as *man*-impersonals, i.e., human impersonal pronouns derived from the word meaning ‘man’ that are found in most Germanic languages and French. As pointed out by Giacalone Ramat & Sansò (2007) and Siewierska & Papastathi (2011), these constructions are found in languages with obligatory subject expression, whereas so-called null subject or *pro*-drop languages are less prone to develop such constructions. Thus, *man*-impersonals are not found in present-day Spanish, Portuguese, or Italian, although they have existed in earlier stages of these languages (Giacalone Ramat & Sansò 2007). Portuguese is an interesting exception to this typological tendency, as it does present a range of constructions based on the noun *pessoa* ‘person’ that have developed uses akin to *man*-impersonals (Duarte & Marques 2014, Posio 2017, 2021, Amaral & Mihatsch 2019, 2023 [this volume]). The Portuguese *a gente* construction deriving from the noun phrase meaning ‘the people’ is another example of a human impersonal construction similar to *man*- constructions. This construction has now become the predominant choice of referential device used for the first-person plural in varieties of Portuguese (in particular Brazilian Portuguese), while in European Portuguese it remains ambiguous between personal and impersonal readings (Posio 2012).

The development of noun-based constructions like *a gente* and the *pessoa* constructions has been previously attributed to the high number of expressed pronominal subjects in Portuguese (in particular Brazilian, but also European variety; Posio 2021). Interestingly, Portuguese seems to be particularly disposed to develop “new” impersonal constructions and referential devices from noun phrase constructions.

The thematic part *between impersonal and personal* addresses the above-mentioned research questions from different angles. Amaral & Mihatsch (2023 [this volume]), Henriques Pestana (2023 [this volume]), and Ornelas de Avelar (2023 [this volume]) examine dialectal and/or informal data from different varieties of Portuguese to study the emergence of “new” impersonal pronouns. Another interesting question is why some impersonal constructions tend to acquire personal uses in discourse: this question is addressed in the papers by Henriques Pestana (2023 [this volume]) and Orozco et al. (2023 [this volume]). A case in

point about the emerging impersonal pronouns in Brazilian Portuguese is the *geral* construction discussed by Ornelas de Avelar, where the adjective meaning ‘general’ has been repurposed as a human impersonal subject. Henriques Pestana (2023 [this volume]) explores the *a gente* construction which is used in Madeiran Portuguese together with the morpheme *se* that also expresses impersonality in what Henriques calls hybrid constructions. These constructions, Henriques argues, are an example of impersonal constructions developing personal uses. A similar development can be observed with the *pessoa* constructions studied by Amaral & Mihatsch that are often used to refer to the speaker, although simultaneously expressing generalizations or mitigation. Orozco et al. (2023 [this volume]) also look into a development from impersonal to personal, but in Colombian Spanish, where the human impersonal pronoun *uno* has developed personal uses to the extent that it can be considered a variant of the 1SG pronoun *yo* ‘I’.

Lastly, the paper by Pierre & De Cock (2023 [this volume]) investigates a configuration where an impersonal third-person plural subject co-occurs with a referential object pronoun. Their paper investigates the use of object discourse participant pronouns in impersonal third person plural constructions (e.g. *me han criticado* ‘they have criticized me’). Given that third-person plurals have been analyzed as an agent-defocusing mechanism, the authors examine to what extent the higher referentiality of the first or second person object pronoun, contrasting with the lower referentiality of the subject, affects the conceptualization of the whole construction. The authors offer a quantitative and qualitative analysis of different types of corpus data by means of which they investigate different factors influencing the use of the examined construction, such as topic continuity, verb type, and different types of register.

4 Reference to the addressee

The variable use of address forms is a widely studied topic in Ibero-Romance linguistics (see, e.g., Hummel et al. 2010 for Spanish), and different geographical and social varieties of Ibero-Romance languages display a wide range of address systems ranging from only one address pronoun referring to singular addressees, e.g., *ustedeo* in different regions of Central America (see Moser 2006, Quesada Pacheco 2010), to tripartite pronominal systems, e.g., the use of *tú*, *vos* and *usted* in Uruguayan Spanish (Steffen 2010), or the use of *tu*, *vós*, and *vostè* in Catalan (see Robinson 1980, Todolí 2006, Nogué 2022, GIEC 2022: 8.2.2), and complex nominal and pronominal systems comprising address pronouns as well as the use of proper names and honorific nominal forms of address, as is the case in European Portuguese (see, e.g., Allen 2019).

Since Brown & Gilman's (1960) seminal work, the use of different forms of address has been related to the notions of power and solidarity that hold between two (or more) interlocutors to different degrees. For example, in the 2SG pronoun *tú* in Peninsular Spanish expresses intimacy and solidarity in reciprocal uses in which the level of power is equal, while in non-reciprocal uses it may express condescendence (cf. Uber 2016: 622). Brown & Levinson (1972) and García (1992) use the notions of positive and negative politeness to account for the choice of address forms. Thus, use of informal address forms like *tú* in Peninsular Spanish can be considered a form of positive politeness, showing affection and approval, while use of formal address forms such as Peninsular Spanish *usted* correlates with negative politeness, i.e., showing respect and keeping distance (cf. Uber 2016: 622).

On a morpho-syntactic level, address forms show interesting patterns of agreement mismatches: the pronoun *usted* and its plural *ustedes* in Spanish are morpho-syntactically third person forms used to refer to the addressee. However, there are varieties of Spanish in which 3PL *ustedes* is used with second person plural (2PL) inflection (De Jonge & Nieuwenhuijsen 2012: 254). Furthermore, in several varieties that use the 2SG pronoun *vos*, different systems of verb inflection can be found: the use of *vos* with the 2SG inflection (called "mixed pronominal *voseo*", e.g., *vos no puedes* 'you cannot') or the use of *tú* with the verbal inflection of *vos* ("mixed verbal *voseo*", e.g., *tú no podés* 'you cannot'; cf. De Jonge & Nieuwenhuijsen 2012: 256–257). In some systems that make use of *tú*, *vos*, as well as *usted*, a functional partition can be observed: for example, in Uruguay, *vos* is used as an intimate and confidential address form and *tú* as an informal but less intimate address form (cf. De Jonge & Nieuwenhuijsen 2012: 258, Hualde et al. 2010: 329). Furthermore, historical, political and social factors may intervene in the choice of address forms in different varieties (as, e.g., in Nicaragua; see Lipski 1994: 159ff).

In several varieties of Catalan, a tripartite address system can also be found: *tu* which is morpho-syntactically a 2SG pronoun, refers to the addressee and it agrees with a 2SG verb (*tu ho saps* '(you) know it'). Use of this form indicates a degree of intimacy (cf. Robinson 1980) between the interlocutors, similarly to *tú* in Spanish. The form *vostè* in singular and *vostès* in plural are interpreted referentially as second person, referring to the addressee(s), but they are morpho-syntactically third person, agreeing with a third person verb form. The address pronoun can also be omitted (e.g., (*Vostè*) *ho sap millor* 'You know it best', GIEC 2022: 8.2.2). This pattern stems from the fact that *vostè* originates from the noun phrase *vostra mercè* 'your grace' and thus behaves like a third person noun phrase

with respect to agreement patterns. Use of these forms is related to a lower degree of intimacy between the interlocutors (Robinson 1980), to politeness and a certain degree of social distancing and formality of the speech act (GIEC 2022: 8.2.2)). Apart from these second person and third person address forms, there is a third form, *vós*, which is used to refer to a 2SG addressee but triggers 2PL agreement on the verb, e.g., *Què en penseu (vós)?* ‘What do (you [2SG]) think [2PL] of it?’ (GIEC 2022: 8.2.2). As mentioned in the GIEC (2022: 8.2.2), traditionally the use of *vós* indicated “cordial and friendly respect” [our translation] and was used to address elderly interlocutors. In colloquial speech, the use of *vós* is decreasing, but it is very common in juridical and administrative language (GIEC 2022: 8.2.2; see also Nogué 2022 for discussion). This form is associated with a lower degree of distancing than the form *vostè*. Thus, in the varieties in which the tripartite system is used, there seems to be a functional partitioning not only between [+/-intimate] forms, but furthermore, between the two [-intimate] forms *vós* and *vostè*. However, some varieties have abandoned the use of a tripartite system, as for example, in Northern Catalonia, where *vostè* is not used anymore, in the Comunitat Valenciana, where *vós* has fallen into disuse; in other regions of Catalonia, the tripartite system is characterized as unstable (cf. Robinson 1980, Nogué 2022).

Within personal pronoun paradigms, address forms are most open to variation and the introduction of new forms deriving from nominal sources. Some well-known cases of such “new” pronouns are the development of *vuestra merced* > *usted* ‘you-singular/formal’ in Spanish as well as the creation of plural forms like *vos* ‘you-PL’ + *otros* ‘others’ > *vosotros* ‘you-plural/informal’ through univerbation (Lapesa 1981 [1942]: 259, 392). As is the case with impersonal constructions, Portuguese is particularly prone to the apparition of “new” address forms based on noun phrases such as *o senhor* ‘the sir’, *a doutora* ‘the doctor’ or combinations of nominal forms of address and proper names (e.g., *a doutora Maria* ‘the doctor Maria’). The complexity of the European Portuguese address system is described in terms of a tripartite categorization into pronominal, nominal and verbal address forms (Cintra 1972, Duarte & Marques 2023 [this volume]). These forms are used to encode different levels of proximity, familiarity and deference that are difficult to formalize or describe in terms of T/V distinctions (Brown & Gilman 1960), as the interpretation of each form depends heavily on sociolinguistic and socio-situational factors. Carreira (2005) considers that the European Portuguese address system provides various ways to encode indirectness and negative politeness. It is also interesting to note that the avoidance of nominal and pronominal address forms is a common strategy: the use of 3SG verb forms

without expressed subjects is a way to avoid the choice of address form and can be considered a “zero degree of politeness” (Carreira 2005: 313).

Research on address forms has traditionally focused on accounting for variation between competing address forms in different discourse types and between different types of interlocutors, as well as explaining the diachronic development of address forms. The variation is affected by sociolinguistic and sociosituational factors such as age, sex, profession, social rank, personal relation, time of acquaintance, place and type of conversation, among others (cf. Uber 2016: 627 for an overview and references).

The third thematic part of the current volume, *Reference to the addressee*, presents two studies focusing on Catalan and European Portuguese address forms from both synchronic and diachronic perspectives. The paper by Nogué-Serrano & Payrató (2023 [this volume]) discusses interesting data and sheds new light on the use and role of different address forms in Catalan parliamentary debates. The authors examine reference to the participants in comparing two time periods: from 1932 to 1938, and from 1980 to 2020, using qualitative as well as quantitative methods. The authors show that the study of reference to discourse participants in parliamentary debates needs to go beyond the study of first and second-person forms and include several third-person forms as well. Furthermore, they observe a development of address forms, specifically vocatives, from more complex to less complex forms. Likewise, they detect an increase of the use of *vostè* and a loss of *vós* and *vostra senyoria* and other third-person forms. Lastly, on a general level, the study shows a move towards more informality on a continuum. The chapter by Duarte & Marques (2023 [this volume]) examines the address form *o senhor* in a wide variety of data from an interactional perspective. Their analysis also covers contracted forms like *sotor*, deriving from the contraction of the complex address form *senhor doutor* ‘mister doctor’ and other innovations. Similarly to the *peessoa* constructions analyzed by Amaral & Mihatsch (2023 [this volume]), the address forms based on *senhor* form a network of partially variable constructions with different grades of productivity and variability, occupying a position between noun phrases and pronouns in the paradigm of referential devices of European Portuguese.

References

- Aijón Oliva, Miguel Á. & María José Serrano. 2010. El hablante en su discurso: Expresión y omisión del sujeto de creio. *Oralia* 13. 7–38.
- Allen, Ana Sofia Ferreira. 2019. *O sistema de formas de tratamento em português europeu*. Lisbon. (Doctoral dissertation).

- Amaral, Eduardo Tadeu Roque & Wiltrud Mihatsch. 2019. Incipient impersonal pronouns in colloquial Brazilian Portuguese. In Peter Herbeck, Bernhard Pöll & Anne C. Wolfsgruber (eds.), *Semantic and syntactic aspects of impersonality* (Linguistische Berichte, Sonderhefte 26), 149–185. Hamburg: Buske.
- Amaral, Eduardo Tadeu Roque & Wiltrud Mihatsch. 2023. Portuguese *a pessoa* and *uma pessoa*: Emerging inclusive impersonals. In Pekka Posio & Peter Herbeck (eds.), *Referring to discourse participants in Ibero-Romance languages*, 107–146. Berlin: Language Science Press. DOI: 10.5281/zenodo.8124494.
- Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London: Routledge.
- Barbosa, Pilar. 1995. *Null subjects*. Cambridge, MA: MIT. (Doctoral dissertation).
- Barbosa, Pilar. 2009. Two kinds of subject *pro*. *Studia Linguistica* 63(1). 2–58. DOI: 10.1111/j.1467-9582.2008.01153.x.
- Bentivoglio, Paola. 1983. Topic continuity in discourse: A study of spoken Latin-American Spanish. In Talmy Givón (ed.), *Topic continuity in discourse*, 255–312. 10.1075/tsl.3.06ben. Amsterdam: John Benjamins.
- Bentivoglio, Paola. 1987. *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela, Consejo de Desarrollo Científico y Humanístico.
- Bessett, Ryan M. 2023. A cross dialectal comparison of first person singular subject pronoun expression in Southern Arizona and Southeast Texas. In Pekka Posio & Peter Herbeck (eds.), *Referring to discourse participants in Ibero-Romance languages*, 25–39. Berlin: Language Science Press. DOI: 10.5281/zenodo.8124488.
- Brown, Penelope & Stephen C. Levinson. 1972. *Politeness: Some universals in language usage* (Studies in Interactional Sociolinguistics 4). Cambridge: Cambridge University Press.
- Brown, Robert & Albert Gilman. 1960. The pronouns of power and solidarity. In Thomas Albert Sebeok (ed.), *Style in language*, 253–276. Cambridge, MA: MIT Press.
- Cabredo Hofherr, Patricia. 2008. Les pronoms impersonnels humains: Syntaxe et interpretation. *Modèles linguistiques* 57. 35–56. DOI: 10.4000/ml.321.
- Cameron, Richard. 1992. *Pronominal and null subject variation in Spanish: constraints, dialects, and functional compensation*. University of Pennsylvania at Philadelphia. (Doctoral dissertation).
- Cameron, Richard. 1993. Ambiguous agreement, functional compensation and nonspecific *tú* in the Spanish of San Juan, Puerto Rico, and Madrid, Spain. *Language Variation and Change* 5(3). 305–334.

- Carreira, M. H. Araújo. 2005. Politeness in Portugal: How to address others? In Leo Hickey & Miranda Steward (eds.), *Politeness in europe*, 306–316. Clevedon/Buffalo/Toronto: Multilingual Matters.
- Carrilho, Ernestina. 2005. *Expletive ele in European Portuguese dialects*. Lisbon. (Doctoral dissertation).
- Carvalho, Ana M., Rafael Orozco & Naomi Shin (eds.). 2015. *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Washington, DC: Georgetown University Press.
- Carvalho, Janayna. 2019. Incorporated subjects in existential impersonal sentences in Brazilian Portuguese. In Peter Herbeck, Bernhard Pöll & Anne Wolfgruber (eds.), *Semantic and syntactic aspects of impersonality* (Linguistische Berichte, Sonderhefte 26), 47–70. Hamburg: Buske.
- Chomsky, Noam. 1981. *Lectures on government and binding: The Pisa lectures*. Dordrecht: Foris Publications.
- Chomsky, Noam. 1982. *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: MIT Press.
- Cintra, Luís Filipe Lindley. 1972. *Sobre as formas de tratamento*. Lisboa: Livros Horizonte.
- De Jonge, Bob & Dorien Nieuwenhuijsen. 2012. Forms of address. In José Ignacio Hualde, Antxon Olarrea & Erin O'Rourke (eds.), *The handbook of Hispanic linguistics*, 247–262. Malden, MA: Wiley-Blackwell.
- Duarte, Isabel Margarida & Maria Aldina Marques. 2014. As formas pronominais EU/TU: Valor genérico e distância. *Revista Galega de Filoloxía* 15. 69–85.
- Duarte, Isabel Margarida & Maria Aldina Marques. 2023. Referring to discourse participants in European Portuguese: The form of address *o senhor*. In Pekka Posio & Peter Herbeck (eds.), *Referring to discourse participants in Ibero-Romance languages*, 273–306. Berlin: Language Science Press. DOI: 10.5281/zenodo.8124504.
- Enríquez, Emilia V. 1984. *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: C. S. I. C.
- Erteschik-Shir, Nomi. 2007. *Information structure: The syntax-discourse interface*. Oxford: Oxford University Press.
- Fernández Soriano, Olga & Susana Táboas Baylín. 1999. Construcciones impersonales no reflejas. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, vol. 2, 1723–1778. Madrid: Espasa.
- García, Carmen. 1992. Refusing an invitation: A case study of Peruvian style. *Hispanic Linguistics* 5(1-2). 207–243.

- Gast, Volker & Johan van der Auwera. 2013. Towards a distributional typology of human impersonal pronouns, based on data from European languages. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries: Studies in the memory of Anna Siewierska*, 119–158. Berlin: De Gruyter.
- Geeslin, Kimberly L., Thomas Goebel-Mahrle, Jingyi Guo & Bret Linfood. 2023. Variable subject expression in second language acquisition: The role of perseveration. In Pekka Posio & Peter Herbeck (eds.), *Referring to discourse participants in Ibero-Romance languages*, 69–104. Berlin: Language Science Press. DOI: 10.5281/zenodo.8124492.
- Giacalone Ramat, Anna & Andrea Sansò. 2007. The spread and decline of indefinite man-constructions in European languages: An areal perspective. In Paolo Ramat & Elisa Roma (eds.), *Europe and the Mediterranean as linguistic areas: Convergences from a historical and typological perspective*, 95–131. Amsterdam: John Benjamins.
- GIEC. 2022. *Gramàtica de la llengua catalana de l'IEC*. Segona edició (en línia). Barcelona: Institut d'Estudis Catalans. DOI: 10.2436/10.2500.08.1.
- Givón, Talmy. 1983. Topic continuity in discourse: An introduction. In Talmy Givón (ed.), *Topic continuity in discourse*, 1–42. Amsterdam: John Benjamins. DOI: 10.1075/tsl.3.01giv.
- Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.
- Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases*. University of Massachusetts, Amherst. (Doctoral dissertation).
- Hennemann, Anja. 2016. A cognitive-constructionist approach to Spanish *creo* \emptyset and *creo yo* '[I] think'. *Folia Linguistica* 50(2). 449–474. DOI: 10.1515/flin-2016-0017.
- Henriques Pestana, Yoselin. 2023. Personal uses of impersonalizing strategies: Hybrid constructions with *a gente* and *se* in rural Madeiran Portuguese varieties. In Pekka Posio & Peter Herbeck (eds.), *Referring to discourse participants in Ibero-Romance languages*, 147–175. Berlin: Language Science Press. DOI: 10.5281/zenodo.8124496.
- Herbeck, Peter. 2021. Perspectival factors and pro-drop: A corpus study of speaker/addressee pronouns with *creer* 'think/believe' and *saber* 'know' in spoken Spanish. *Glossa* 122(1). 1–34. DOI: 10.16995/glossa.5873.
- Herbeck, Peter. 2022. Variable first person singular subject expression in spoken Valencian Catalan. *Catalan Journal of Linguistics* 21. 129–157. DOI: 10.5565/rev/catjl.380.
- Hochberg, Judith. 1986. Functional compensation for /s/ deletion in Puerto Rican Spanish. *Language* 62(3). 609–621. DOI: 10.1353/lan.1986.0041.

- Holmberg, Anders. 2005. Is there a little pro? Evidence from Finnish. *Linguistic Inquiry* 36(4). 533–564.
- Hualde, José Ignacio, Antxon Olarrea, A. M. Escobar & E. C. Travis. 2010. *Introducción a la lingüística hispánica*. Cambridge: Cambridge University Press.
- Hummel, Martin, Bettina Kluge & María Eugenia Vázquez Laslop (eds.). 2010. *Formas y fórmulas de tratamiento en el mundo hispánico*. México/Graz: El Colegio de México/Karl Franzens Universität.
- Kato, Mary A. 1999. Strong and weak pronominals in the null subject parameter. *Probus* 11(1). 1–38. DOI: 10.1515/prbs.1999.11.1.1.
- Kato, Mary A. & Esmeralda V. Negrão (eds.). 2000. *Brazilian Portuguese and the null subject parameter*. Frankfurt am Main & Madrid: Vervuert. DOI: 10.31819/9783964561497.
- Kibrik, Andrej A. 2011. *Reference in discourse*. Oxford: Oxford University Press.
- Lapesa, Rafael. 1981. *Historia de la lengua española*. Madrid: Editorial Gredos.
- Levinson, Stephen C. 1987. Pragmatics and the grammar of anaphora: A partial pragmatic reduction of Binding and Control phenomena. *Journal of Linguistics* 23(2). 379–434. DOI: 10.1017/S0022226700011324.
- Lipski, John M. 1994. *El español de América*. London: Longman.
- Lopes, Célia Regina dos Santos. 2004. A gramaticalização de a gente em português em tempo real de longa e de curta duração: Retenção e mudança na especificação dos traços intrínsecos. *Fórum Linguístico* 4(1). 47–80.
- Luján, Marta. 1999. Expresión y omisión del pronombre personal. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, 1275–1315. Madrid: Espasa.
- Malchukov, Andrej L. & Akio Ogawa. 2011. Towards a typology of impersonal constructions. In Andrej Malchukov & Anna Siewierska (eds.), *Impersonal constructions: A cross-linguistic perspective*, 19–56. Amsterdam: John Benjamins.
- Malchukov, Andrej L. & Anna Siewierska (eds.). 2011. *Impersonal constructions: A cross-linguistic perspective*. Amsterdam: John Benjamins.
- Masullo, Pascual J. 1993. Two types of quirky subjects: Spanish vs. Icelandic. *North East Linguistics Society* 23(2). 303–317.
- Mayol, Laia. 2010. Contrastive pronouns in null-subject Romance languages. *Lingua* 120(10). 2497–2514. DOI: 10.1016/j.lingua.2010.04.009.
- Morales, Amparo. 1997. La hipótesis funcional y la aparición de sujeto no nominal: El español de Puerto Rico. *Hispania* 80(1). 10.2307/345995, 153–165.
- Moser, Karolin. 2006. La variación entre formas ustedeantes y voseantes a nivel del discurso familiar en la clase media y alta de San José – Costa Rica: ¿una estrategia de cortesía? In Martina Schrader-Kniffki (ed.), *La cortesía en el mundo*

- hispánico: Nuevos contextos, nuevos enfoques metodológicos*, 97–116. Madrid: Iberoamericana.
- Nogué, Neus. 2022. El sistema de tractaments del català. In Neus Nogué, Josep Besa, Lluís Payrató & Montserrat Sendra (eds.), *Els sistemes de tractament: Una visió interlingüística i aplicada*. Barcelona: Edicions UB.
- Nogué-Serrano, Neus & Lluís Payrató. 2023. Variation and change in reference to discourse participants in Catalan parliamentary debate (1932–1938 and 1980–2020). In Pekka Posio & Peter Herbeck (eds.), *Referring to discourse participants in Ibero-Romance languages*, 307–343. Berlin: Language Science Press. DOI: 10.5281/zenodo.8124506.
- Ornelas de Avelar, Juanito. 2023. On pronominal uses of *geral* in Brazilian Portuguese. In Pekka Posio & Peter Herbeck (eds.), *Referring to discourse participants in Ibero-Romance languages*, 177–202. Berlin: Language Science Press. DOI: 10.5281/zenodo.8124498.
- Orozco, Rafael. 2015. Pronominal variation in Colombian Costeño Spanish. In Ana M. Carvalho, Rafael Orozco & Naomi Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective*, 17–58. Washington: Georgetown University Press.
- Orozco, Rafael & Gregory Guy. 2008. El uso variable de los pronombres sujetos: ¿Qué pasa en la costa Caribe colombiana? In Maurice Westmoreland & Juan Antonio Thomas (eds.), *Selected proceedings of the 4th workshop on Spanish sociolinguistics*, 70–80. Somerville, MA: Cascadilla.
- Orozco, Rafael, Luz Marcela Hurtado & Marianne Dieck. 2023. Toward the prevalence of a personal use of impersonal *uno* in Colombian Spanish. In Pekka Posio & Peter Herbeck (eds.), *Referring to discourse participants in Ibero-Romance languages*, 203–236. Berlin: Language Science Press. DOI: 10.5281/zenodo.8124500.
- Otheguy, Ricardo, Ana Celia Zentella & David Livert. 2007. Language and dialect contact in Spanish in New York: Towards the formation of a speech community. *Language* 83(4). 770–802.
- Pierre, Émeline & Barbara De Cock. 2023. Discourse participants in impersonal constructions: The case of first and second person object pronouns with Spanish non-anaphoric third person plural subjects. In Pekka Posio & Peter Herbeck (eds.), *Referring to discourse participants in Ibero-Romance languages*, 237–269. Berlin: Language Science Press. DOI: 10.5281/zenodo.8124502.
- Posio, Pekka. 2011. Spanish subject pronoun usage and verb semantics revisited: First and second person singular subject pronouns and focusing of attention in spoken Peninsular Spanish. *Journal of Pragmatics* 43(3). 777–798. DOI: 10.1016/j.pragma.2010.10.012.

- Posio, Pekka. 2012. Who are ‘we’ in spoken Peninsular Spanish and European Portuguese? Expression and reference of first person plural subject pronouns. *Language Sciences* 34(3). 339–360. DOI: 10.1016/j.langsci.2012.02.001.
- Posio, Pekka. 2013. The expression of first-person-singular subjects in spoken Peninsular Spanish and European Portuguese: Semantic roles and formulaic sequences. *Folia Linguistica* 47(1). 253–291. DOI: 10.1515/flin.2013.010.
- Posio, Pekka. 2014. Subject expression in grammaticalizing constructions: The case of *creo* and *acho* ‘I think’ in Spanish and Portuguese. *Journal of Pragmatics* 63. 5–18. DOI: 10.1016/j.pragma.2013.07.001.
- Posio, Pekka. 2015. Subject pronoun usage in formulaic sequences. In A. Carvalho, R. Orozco & N. Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective*, 59–78. Washington, DC: Georgetown University Press.
- Posio, Pekka. 2017. Entre lo impersonal y lo individual: Estrategias de impersonalización individualizadoras en el español y portugués europeos. *Spanish in Context* 14(2). 209–229. DOI: 10.1075/sic.14.2.03pos.
- Posio, Pekka. 2018. Properties of pronominal subjects. In Kimberly L. Geeslin (ed.), *The Cambridge handbook of Spanish linguistics*, 286–306. Cambridge: Cambridge University Press. DOI: 10.1017/9781316779194.014.
- Posio, Pekka. 2021. *A pessoa* and *uma pessoa*: Grammaticalization and functions of a human impersonal referential device in European Portuguese. *Journal of Portuguese Linguistics* 20(1). DOI: 10.5334/jpl.254.
- Posio, Pekka & Maria Villkuna. 2013. Referential dimensions of human impersonals in dialectal European Portuguese and Finnish. *Linguistics* 51. 177–229. DOI: 10.1515/ling-2013-0006.
- Quesada Pacheco, Miguel Ángel. 2010. Formas de tratamiento en Costa Rica y su evolución (1561–2000). In Martin Hummel, Bettina Kluge & María Eugenia Vázquez Laslop (eds.), *Formas y fórmulas de tratamiento em el mundo hispánico*, 649–669. Ciudad de México/Graz: El Colegio de México/Karl-Franzens-Universität Graz.
- Ranson, Diana. 1991. Person marking in the wake of /s/ deletion in Andalusian Spanish. *Language Variation and Change* 3(2). 133–152.
- Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27. 53–94.
- Rigau, Gemma. 1989. Connexity established by emphatic pronouns. In Maria-Elisabeth Conte, János Sándor Petőfi & Emel Sözer (eds.), *Text and discourse connectedness*, 191–205. Amsterdam: John Benjamins. DOI: 10.1075/slcs.16.17rig.
- Rivas, Javier & Esther L. Brown. 2023. Variable indirect object pronoun expression: A usage-based analysis of Galician and Spanish. In Pekka Posio & Peter

- Herbeck (eds.), *Referring to discourse participants in Ibero-Romance languages*, 41–67. Berlin: Language Science Press. DOI: 10.5281/zenodo.8124490.
- Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris Publications.
- Robinson, Joy L. 1980. Sociolingüística i variació semàntica: Els tractaments em català. In Jordi Burguera & Josep Massot i Muntaner (eds.), *Actes del cinquè col·loqui internacional de llengua i literatura catalanes*, 619–638. Barcelona: Publicacions de l'Abadia de Montserrat.
- Shin, Naomi. 2014. Grammatical complexification in Spanish in New York: 3SG pronoun expression and verbal ambiguity. *Language Variation and Change* 26. 303–330.
- Shin, Naomi & Ricardo Otheguy. 2005. Overt non-specific *ellos* in Spanish in New York. *Spanish in Context* 2(2). 157–174.
- Siewierska, Anna & Maria Papastathi. 2011. Towards a typology of third person plural impersonals. *Linguistics* 49(3). 575–610.
- Silva-Corvalán, Carmen. 2001. *Sociolingüística y pragmática del español*. Washington, DC.: Georgetown University Press.
- Solà, Jaume. 1992. *Agreement and subjects*. Universitat Autònoma de Barcelona. (Doctoral dissertation). <http://hdl.handle.net/10803/4847>.
- Steffen, Joachim. 2010. El tratamiento en el Uruguay. In Martin Hummel, Bettina Kluge & María Eugenia Vázquez Laslop (eds.), *Formas y fórmulas de tratamiento em el mundo hispánico*, 449–464. Ciudad de México/Graz: El Colegio de México/Karl-Franzens-Universität Graz.
- Todolí, Júlia. 2006. Els pronoms. In Joan Solà, Maria-Rosa Lloret, Joan Mascaró & Manuel Pérez Saldanya (eds.), *Gramàtica del català contemporani* (), vol. 2, 1337–1433. Barcelona: Editorial Empúries.
- Toribio, Almeida Jacqueline. 2000. Setting parametric limits on dialectal variation in Spanish. *Lingua* 110. 315–341.
- Travis, Catherine E. & Rena Torres Cacoulios. 2012. What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics* 23(4). 711–748. DOI: 10.1515/cog-2012-0022.
- Travis, Catherine E. & Rena Torres Cacoulios. 2018. Discovering structure. In Naomi Shin & Daniel Erker (eds.), *Questioning theoretical primitives in linguistic inquiry*, 67–90. Amsterdam: John Benjamins.
- Uber, Diane R. 2016. Formas de tratamiento. In Javier Gutiérrez-Rexach (ed.), *Enciclopedia de lingüística hispánica*, 620–629. London & New York: Routledge.
- Zilles, Ana M. S. 2005. The development of a new pronoun: The linguistic and social embedding of a gente in Brazilian Portuguese. *Language Variation and Change* 17. 19–53.

Part I

Variable expression of subjects and objects

Chapter 2

A cross dialectal comparison of first person singular subject pronoun expression in Southern Arizona and Southeast Texas

Ryan M. Bessett

University of Texas Rio Grande Valley

This study provides a cross-dialectic comparison of first person singular subject pronoun expression in the Spanish varieties of two US-Mexico borderland communities, Southern Arizona and Southeast Texas. Using data collected from sociolinguistic interviews of 32 Spanish/English bilingual speakers, this analysis further explores the impact that trans-frontier practices have on the realization of subject pronouns in border communities and demonstrates the similarities in the variable grammar of the Spanish spoken in the US Southwest. The results show that both Arizona and Texas express first person singular pronouns at a similar rate (19.3% and 18.7%, respectively). Additionally, the linguistic factors that condition the variable (switch reference; clause type; tense, mood, and aspect; and whether or not the verb is reflexive) are very similar within each group.

1 Introduction

Subject pronoun expression (SPE) is a variable context in Spanish, in which a speaker can express a pronoun, as in (1) or not express a pronoun, as in (2), both from the Corpus del Español del Valle (CoBiVa) (Christoffersen & Bessett 2019).

- (1) Pues, *yo me fui pensando porque para mí- a mí- yo fui- yo era criado con la intención de que a un padrecito nunca se dice que no. Entonces, yo me fui para la casa.* (CoBiVa005)



‘Well, *I* went about thinking because for me- for me- *I* was- *I* was raised with the intention that one to a parent never says no. So, *I* went home.’ (CoBiva005)

- (2) No, me gusta aquí el Valle. (∅) Estoy a gusto. Ya me (∅) acostumbré aquí, ya (∅) sé cómo – cómo es la gente. (CoBiVa006)
- ‘No, I like it here in the Valley. (∅) am happy. (∅) already got accustomed here, (∅) already know what – what the people are like.’ (CoBiVa006)

An abundance of research has been conducted on this variable and in a wide a range of communities in the United States, including: Arizona (Cerrón-Palomino 2016, Bessett 2018), California (Silva-Corvalán 1994, Bayley & Pease-Alvarez 1997), Florida (Hurtado 2001, Abreu 2009), Georgia (Limerick 2017), New Mexico (Torres Cacoullos & Travis 2010b,a), New Jersey (Flores-Ferrán 2007), and New York (Flores-Ferrán 2004, Otheguy et al. 2007, Otheguy & Zentella 2012, among others). The center focus of SPE studies in the US has been to examine the possible effects of language contact on SPE in US Spanish, as witnessed in the summary of the previous research that follows.

One of the most debated features of SPE in US Spanish is the overall frequency with which bilinguals express pronouns, specifically if they do so at a higher rate than monolinguals, since in English, pronouns are expressed at nearly categorical rates. There are some studies that show a higher rate (Shin & Otheguy 2005, Otheguy et al. 2007, Otheguy & Zentella 2012, among others), and there are others that do not (Silva-Corvalán 1994, Bayley & Pease-Alvarez 1997, Hurtado 2001, Flores-Ferrán 2004, 2007, Torres Cacoullos & Travis 2010a,b, Cerrón-Palomino 2016, Limerick 2017, Bessett 2018, among others).

In order to determine the significance of overall frequency, the extralinguistic factor of community or group has been used as a predictor variable among different groups. For example, in New York, bilinguals who were born in the US produce more expressed pronouns than speakers who were born outside the US (Otheguy et al. 2007, Otheguy & Zentella 2012). However, in Arizona, when comparing Mexican vs Arizona born speakers, community was not a significant factor (Bessett 2018). Another way to group participants to test frequency differences is through proficiency level and in Arizona English-dominant bilinguals are shown to disfavor expressed pronouns, the opposite of what would be expected if the result were due to contact with English (Cerrón-Palomino 2016).

In addition to overall frequency, several linguistic factors have been found to condition SPE. One of the most robust linguistic factors is the tense, mood, and aspect (TMA) of the verb. On the whole, it is noted that in Spanish, in general, TMA forms that are morphologically ambiguous (those whose form does

2 First person singular subject pronoun expression in Arizona and Texas

not expressly provide the subject, like *tuviera* ‘I/(s)he had’ or *hablaba* ‘I/(s)he spoke’) tend to favor expressed pronouns while those forms that do not present morphological ambiguity (like *tengo* ‘I have’ or *hablé* ‘I spoke’), tend to disfavor expressed pronouns (Carvalho et al. 2015). This tendency is also borne out in the context of US Spanish (Otheguy & Zentella 2012, Shin & Montes-Alcalá 2014, Cerrón-Palomino 2016, Limerick 2017, among others), although there do exist a few exceptions (Shin 2014, Limerick 2019).

A second linguistic factor is known as the switch reference constraint. This factor group examines the relationship in subject between a verb and the verb that comes before it. Commonly, three levels are considered: same subject, a change in the subject but same object, and a complete change in subject. In Spanish, when there is a complete change in subject in the present verb as compared to the previous one, pronouns are expressed at a higher rate than when the subject stays the same (Carvalho et al. 2015, among many others). In bilingual communities in the US, some studies note that difference between same and change in subject happens to a lesser degree than among monolingual speakers (Otheguy & Zentella 2012, Limerick 2017). However, in other US bilingual communities, the switch reference constraint shows similar results to monolingual speakers (Silva-Corvalán 1994, Hurtado 2001, Torres Cacoullós & Travis 2018, Cerrón-Palomino 2016, Bessett 2018).

Another factor group considered in previous studies is whether the verb is produced with a reflexive pronoun. When speakers produce a verb with a reflexive pronoun, this context has been found to disfavor the use of an expressed pronoun (Otheguy et al. 2007, Abreu 2009, Otheguy & Zentella 2012, Cerrón-Palomino 2016, among others).

Clause type has also been noted to condition SPE in Spanish. In general, main clauses tend to favor expressed pronouns, while subordinate clauses and coordinate clauses disfavor expressed pronouns (Flores-Ferrán 2009, Otheguy & Zentella 2012, Shin & Montes-Alcalá 2014). However, in some communities, clause type is not a significant predictor variable (Torres Cacoullós & Travis 2010b).

To summarize the overall findings of the realization of SPE in the context of Spanish in the US, the literature is divided into those that show evidence of contact (Silva-Corvalán 1994, Shin & Otheguy 2005, Otheguy et al. 2007, Otheguy & Zentella 2012, Limerick 2017, among others) and those that do not (Hurtado 2001, Flores-Ferrán 2004, Torres Cacoullós & Travis 2010a,b, Cerrón-Palomino 2016, Bessett 2018, among others). Interestingly, in general, a pattern emerges, in which the studies on speech communities in the US Southwest tend to show lack of influence of contact for this variable including in New Mexico (Torres Cacoullós & Travis 2010a,b) and Arizona (Cerrón-Palomino 2016, Bessett 2018).

This brings to the forefront of the discussion the trans-frontier practices from residents on both sides of the US–Mexico border (Jaramillo 1995, Bessett 2015 and Cerrón-Palomino 2016) that help to facilitate connectivity between the Spanish spoken on both sides of the border, especially in terms of subject pronoun expression. These results bring about the focus of the present study, which seeks to document this continuity of linguistic behavior along the border, through the analysis of first person-singular subject pronoun expression in Southern Arizona and Southeastern Texas. Previous SPE research on US Spanish includes different variables and different coding practices, making it difficult to compare clearly between communities. In the present study, by comparing two communities with the same factor groups, within group factors and the same researcher (coder), commonalities and differences noted in the results can be taken to better represent the realization of SPE in the communities in general.

2 Methodology

2.1 Participants

The participants in this study consist of 32 bilinguals (Spanish–English) from two bilingual communities along the US–Mexico border, one in Southern Arizona (Tucson and Nogales) and the other in Southeastern Texas (along the border from the McAllen metropolitan area through Brownsville). The participants are equally divided by community with 16 Arizona participants from the Bessett (2012) corpus and 16 from the Corpus Bilingüe del Valle (Christoffersen & Bessett 2019). They are further divided equally by gender (16 men and 16 women). Table 1 shows the distribution of the participants by community and gender.

Table 1: Participants by community and gender

Gender	Community	
	Arizona	Texas
Male	8	8
Female	8	8
Total	16	16

Both participant groups and corpora represent data from sociolinguistic interviews of about an hour in length. All participants self-evaluated their proficiency.

For the Arizona group, the scale was 0 (low proficiency) to 10 (high proficiency) and all participants rated themselves as 6 or higher in both Spanish and English. For the Texas group, the scale was from 0–6 and all participants rated their Spanish and English as 4 or higher. The speakers also demonstrated their bilingual abilities by participating in a conversation in Spanish for the duration of the approximately one hour-long interview, as well as a post interview conversation in English. The participants all also attend(ed) school in the United States (with English instruction), and they live and work in bilingual communities where they are called upon to speak and interact in both languages.

2.2 Data collection

The first 100 instances of verbs conjugated for the first person singular subject pronoun *yo* were extracted from each of the 32 sociolinguistic interviews. Limiting the study to one grammatical person is a practice based off the knowledge that each grammatical person is conditioned differently in terms of SPE and has been applied in previous studies (e.g., Shin & Otheguy 2005, Torres Cacoullos & Travis 2010a,b, Shin 2014, Bessett 2018). The decision to limit the number of tokens to 100 was to provide as equal a sample as possible among the 32 participants. Some participants use first person singular *yo* quite often, while others do not.

Some specific tokens that refer to structures in which the participant does not make the decision on whether or not to express the subject pronoun *yo* were excluded from the analysis. For example, reported speech (see 3), where the person who produced the original utterance and not the participant made the decision to express, or not, the subject pronoun.

- (3) ...y un día me dijo- mi mamá dijo... dijo “(\emptyset) *Tengo* que ir.” (CoBiVa005)
‘...and one day (she) told me- my mom said... (she) said “(\emptyset) *have* to go.”
(CoBiVa005)

Another such case is with set phrases (as in example 4), where again the participant does not make the decision to express or not a pronoun.

- (4) ¿Cómo te (\emptyset) *diré*? (CoBiVa039)
‘How will (\emptyset) *tell* you?’ (CoBiVa039)

The tokens were coded for the dependent variable of subject pronoun expression (expressed, unexpressed) as well as the four linguistic factors that have

proven to condition the variable, as outlined in the introduction. For the factor group of “tense, mood, and aspect” (TMA), the categories were simplified to two, ambiguous TMA morphology (see example 5) and unambiguous TMA morphology (see example 6), following Cerrón-Palomino (2016), based on the well attested pattern of TMA outlined in the introduction. This decision also makes a comparison between the results of the present study and previous US Spanish communities more straightforward.

- (5) ...para cuando *yo estaba* chiquillo en la escuela... (CoBiVa005)
'...for when *I was* little in school...' (CoBiVa005)
- (6) ...y vino un señor ahí a la casa a c- a curar la vaca y lo (\emptyset) *miré* como le hizo... (CoBiVa005)
'...and a man came there to the house to c- cure the cow and (\emptyset) *saw* how he did it...' (CoBiVa005)

Next, the factor group of “switch reference” documented the relationship of the current verb to the previous verb and consisted of three categories, coreference with subject (no switch, see example 7), switch in subject but coreference with object (see example 8), and a switch with the subject and all objects (complete switch, see example 9).

- (7) ...no puedo comer comida de México cuando (\emptyset) *estoy* acá. (CoBiVa006)
'...I can't eat Mexican food when (\emptyset) *am* here.' (CoBiVa006)
- (8) No, me gusta aquí el Valle. (\emptyset) *estoy* a gusto. (CoBiVa006)
'No, I like it here in the Valley. (\emptyset) *am* content.' (CoBiVa006)
- (9) No es una mala opción, ¿verdad?, pero *yo creo* que... (CoBiVa006)
'It's no a bad option, right?, but *I think* that...' (CoBiVa006)

A third factor group, “reflexive verb”, coded for if the verb was conjugated with a reflexive pronoun (see example 11) or without a reflexive pronoun (see example 10).

- (10) ...ya concordamos más las opciones y de que (\emptyset) *voy a hacer* eso... (T01)
'...we already agreed the options more and that (\emptyset) *am going to do* that...' (T01)
- (11) ...si no era que me (\emptyset) quedaba dormida en una casa de una amiga... (T01)
'...if it wasn't that (\emptyset) *fell* asleep at a friend's house...' (T01)

Lastly, “clause type” coded for the type of clause in which the verb was located. This factor group was divided into the following three categories: main clause (defined as an isolated clause, a clause between pauses, or one that had a subordinate clause) as seen in example (12), coordinate clause (which included a string of main clauses) as seen in example (13), and subordinate clauses (see example 14).

- (12) Pues, (\emptyset) *tengo* un hermano mayor que yo... (T02)
'...Well, (\emptyset) *have* and older brother...' (T02)
- (13) De chiquita no prestaba mucha atención a cómo era Tucson y luego como (\emptyset) *he notado* muchos cambios... (T02)
'...As a little girl I didn't pay much attention to what Tucson was like and later like (\emptyset) *have noticed* a lot of changes...' (T02)
- (14) Yo siempre era muy floja, entonces no es que no (\emptyset) *pudiera hacer* el trabajo... (T02)
'...I was always very lazy, so it wasn't that (\emptyset) *couldn't do* the work...' (T02)

Once the data was coded, it was analyzed through a multivariate model using the statistical program GoldVarb for Mac (Sankoff et al. 2018).

3 Results

This section presents the results of the comparison in the realization of first person singular SPE between the Arizona and Texas communities. To this end, §3.1 provides a summary of the overall frequency of SPE by community and compares these findings to other US Southwest communities. Next, §3.2 presents an overview of the constraint hierarchy for the linguistic factor groups between the Arizona and Texas communities. Then, §3.3 explores a detailed analysis of the factors that condition SPE in the Arizona and Texas and communities and discusses the similarities and differences between the two.

3.1 Overall frequency of SPE in Arizona and Texas

The overall frequency with which the Arizona and Texas speakers produce an expressed pronoun is roughly the same, 19.3% (274 of a total of 1,423 tokens) for Arizona bilinguals and 18.7% (299 of a total of 1,600 tokens) for the Texas bilinguals. Table 2 shows this pattern.

Table 2: Overall frequency of first person singular SPE in Arizona and Texas Spanish

	%	<i>n</i>
Arizona	19.3	274/1423
Texas	18.7	299/1600

These results indicate a first parallel between the two communities as well as to the Phoenix, Arizona community whose speakers produce expressed pronouns at a rate of 23.2% (Cerrón-Palomino 2016). While there are clear parallels between Arizona and Texas, in terms of overall frequency, the speakers in New Mexico (Torres Cacoullós & Travis 2010a) as well as California (Silva-Corvalán 1994) show higher rates of 32% and 34.7%, respectively, diverging from the pattern. However, these rates all fall well within the range of monolingual Mexican communities which run from 16.7% in Sonora (Bessett 2018) to 33% in Xalapa (Orozco 2016), showing similar patterns of subject pronoun expression in the Mexican Spanish spoken on both sides of the US-Mexico border. It is also important to mention, as Travis (2007) warns, overall frequency can be misleading, and so it is crucial to determine whether or not this finding is significant. One way to establish this is to use community as a factor group in the logistic regression and determine if the factor is selected by the model. Table 3 shows the factor groups and their ranges for the factors that were selected by GoldVarb when using community (Arizona vs Texas) as a factor group.

Table 3: Constraint hierarchy of the linguistic factor groups with community (Arizona/Texas) as a factor for the probability of an expressed *yo*

Factor group	Range
Switch reference	38
Clause Type	32
TMA	21
Reflexive	10
Log likelihood = -1242.309	
$p < 0.05$	

2 First person singular subject pronoun expression in Arizona and Texas

Community (Arizona/Texas) is not among the factors that condition SPE expression in this data set. This suggests that the realization of the variable in these two communities is similar. In order to further explore this idea, §3.2 and §3.3 discuss the constraint hierarchies of the factors that condition first person subject pronoun expression in the Arizona and Texas communities in separate regression models.

3.2 Constraint hierarchy of the linguistic factor groups

Having discussed the overall frequency, it is important to compare the linguistic factors that condition SPE in Arizona and Texas. The first measure is to evaluate the hierarchy of the factor groups, in accordance with comparative sociolinguistics which maintains that similarities in the constraint ranking can indicate a common origin for the pattern of the given structure in the communities (Tagliamonte 2003). Table 4 shows the hierarchy of factor groups and range for both communities.

Table 4: Factor groups that condition first person singular SPE in Arizona and Texas

Arizona			Texas	
Factor group	Range		Factor group	Range
Switch/Reference	38	=	Switch Reference	39
Clause Type	35	=	Clause Type	31
TMA	17	=	TMA	24
Reflexive	16	≠	Reflexive	[7]
Log likelihood = -582.725			Log likelihood = -654.674	
$p < 0.05$			$p < 0.05$	

Both the Arizona and Texas groups show switch reference, followed by clause type and then TMA as the significant factor groups. There is a discrepancy between the two data sets in that reflexive is significant for Arizona, but not Texas. However, in both cases, this factor group is the lowest ranking factor. In general, the two groups are strikingly similar in terms of the constraint hierarchy that conditions first person singular SPE. When comparing this to other US Southwest communities, we see more similarities. In Phoenix, Arizona (Cerrón-Palomino 2016), while clause type was not included, switch reference was followed by

TMA and then reflexive. Additionally, in New Mexico (Torres Cacoullós & Travis 2010b), switch reference was followed by TMA (ambiguity of verb morphology). However, for this community clause type was not significant. Again, there seem to be strong similarities between the Arizona and Texas communities of the current study, and possibly with the previously studied Arizona (Cerrón-Palomino 2016) and New Mexico (Torres Cacoullós & Travis 2010b) communities. To better understand this relationship, §3.3 will explore the constraint ranking within each factor group.

3.3 Linguistic factors that condition first person SPE

While §3.2 suggests parallels between the factor groups that condition SPE among the Arizona and Texas speakers of this study, this section now turns to the more detailed analysis of the within-group factors and the directionality of their effects. Table 5 outlines these results.

First, for both groups, switch reference is the highest ranking factor group. A complete switch from the previous subject highly favors an expressed pronoun in both communities and to the same degree (FW = 0.70), while a coreferential subject disfavors an expressed pronoun to a similar degree in both communities with a factor weight of 0.32 for Arizona and 0.31 for Texas. In both communities, a switch in the subject but with a coreferential object is ranked second, but this context disfavors an expressed pronoun in Arizona (FW = 0.35) while it slightly favors an expressed pronoun in Texas (FW = 0.56). These results, in addition to following the well-attested pattern of Spanish in general (Carvalho et al. 2015), are also seen in New Mexico (Torres Cacoullós & Travis 2010b) as well as in Phoenix, Arizona (Cerrón-Palomino 2016).

The second highest ranked factor group, clause type, shows further parallels between the two groups. Main clauses favor expressed pronouns for both Arizona (FW = 0.75) and Texas (FW = 0.71) speakers, while coordinate and subordinate clauses disfavor expressed pronouns. However, in Texas coordinate clauses rank lower (FW = 0.40) than subordinate clauses (FW = 0.46), while in Arizona the pattern is flipped, and subordinate clauses rank lower (FW = 0.40) than coordinate clauses (FW = 0.42). In terms of percent of expressed pronouns in coordinate and subordinate clauses however, the pattern is the same for the two communities. In Arizona, first person singular verbs are produced with an expressed pronoun in subordinate clauses at a rate of 15.2% (35/230) while in coordinate clauses pronouns are expressed at a rate of 12.9% (112/870). In Texas, the rates are 17.2% (52/303) in subordinate clauses and 11.2% (94/836) in coordinate clauses. Overall, in this context there are apparent similarities between the two communities.

2 First person singular subject pronoun expression in Arizona and Texas

Table 5: Comparison of the linguistic factors that condition the expression of first person singular pronouns in Arizona and Texas. FW = factor weight.

Factor	Arizona			Texas		
	FW	%	<i>n</i>	FW	%	<i>n</i>
Switch Reference						
Complete switch	0.70	32.1	207/645	0.70	30.8	208/676
Same object	0.35	8.7	8/92	0.56	20.8	33/159
Same subject	0.32	8.6	59/686	0.31	7.6	58/765
	Range = 38			Range = 39		
Clause type						
Main	0.75	39.3	127/323	0.71	33.2	153/461
Coordinate	0.42	12.9	112/870	0.40	11.2	94/836
Subordinate	0.40	15.2	35/230	0.46	17.2	52/303
	Range = 35			Range = 31		
TMA						
Ambiguous	0.63	23.4	81/346	0.69	27.2	84/309
Not ambiguous	0.46	17.9	193/1077	0.45	16.7	215/1291
	Range = 17			Range = 24		
Reflexive						
Not reflexive	0.52	20.3	253/1244	[0.51]	19.4	266/1372
Reflexive	0.36	11.7	21/179	[0.44]	14.5	33/228
	Range = 16			Range = [7]		
	Log likelihood = -582.725			Log likelihood = -654.674		
	<i>p</i> < 0.05			<i>p</i> < 0.05		

The next factor group is TMA which is separated into ambiguous and not ambiguous verb morphology. Among Arizona speakers, ambiguous verb morphology favors expressed pronouns at a rate of 0.63, similar to the 0.69 rate for Texas speakers, while unambiguous verb morphology disfavors expressed pronouns in both Arizona (FW = 0.46) and Texas (FW = 0.45). These results conform to the general pattern noted for Spanish in general (Carvalho et al. 2015) and, more specifically, to the Phoenix, Arizona (Cerrón-Palomino 2016) and New Mexico (Torres Cacoullós & Travis 2010b) communities.

The last factor group is the presence or absence of a reflexive pronoun. The factor group is only significant for the Arizona community. However, the distribution of the effect is the same for both communities, the absence of a reflexive pronoun favors expressed subject pronouns in Arizona (FW = 0.52) as well as in Texas (FW = [0.51]), while the presence of a reflexive pronoun disfavors expressed subject pronouns in Arizona (FW = 0.36) and Texas (FW = [0.44]). In Phoenix, Arizona, much like the Arizona speakers in the current study, reflexive is a significant factor group (Cerrón-Palomino 2016) and the absence of a reflexive pronoun favors expressed subject pronouns. With this factor group we again see continuities in the realization of SPE within the US Southwest.

4 Conclusion

This study provided a cross-dialectal comparison of first person singular subject pronoun expression in two border communities in the US Southwest, Southern Arizona and Southeastern Texas. Overall, the results indicated clear ties between the Arizona and Texas speech communities under consideration in this study. The overall frequency was nearly identical for the two groups, with the Arizona participants expressing first person singular pronouns at a rate of 19.3%, while the rate for Texas participants is 18.7%. This rate was also similar to the Phoenix, Arizona (Cerrón-Palomino 2016) community as well as to the slightly higher rates of New Mexico (Torres Cacoullós & Travis 2010a) and California (Silva-Corvalán 1994) bilinguals. When taking into account the rates of expressed pronouns in Monolingual Mexican communities from 16.7% in Sonora (Bessett 2018) to 33% in Xalapa (Orozco 2016), US border communities fit well within the range, demonstrating a similarity. Additionally, the constraint hierarchy was identical for both Arizona and Texas speakers, switch reference, clause type, TMA and then reflexive (although reflexive was not significant for Texas bilinguals). The within-group factors were also ordered in the same way, with the exception of the inverse order of coordinate and subordinate clauses (although in both communities the two contexts disfavor expressed pronouns). First person singular subject

pronoun expression appears to be conditioned in the same way in both Arizona and Texas. These results demonstrate continuity in the Spanish spoken in the two communities (Arizona and Texas) and when compared to previous studies, a pattern of continuity emerges among US Southwest bilinguals and within the Spanish spoken on both sides of the US-Mexico border. By examining these patterns with other variables, future studies may be able to demonstrate an overall pattern of cohesion in the Spanish spoken along the US-Mexico border.

References

- Abreu, Laurel. 2009. *Spanish subject personal pronoun use by monolinguals, bilinguals and second language learners*. University of Florida. (Doctoral dissertation).
- Bayley, Robert & Lucinda Pease-Alvarez. 1997. Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change* 9(3). 349–371.
- Bessett, Ryan M. 2012. *La extensión de estar a través del tiempo y el espacio: Un estudio diacrónico de los siglos XIII–XX y sincrónico de Hermosillo, Sonora y el Sur de Arizona*. University of Arizona, Tucson. (MA thesis).
- Bessett, Ryan M. 2015. The extension of *estar* across the Mexico–US border: Evidence against contact induced acceleration. *Sociolinguistic Studies* 9(4). 421–443.
- Bessett, Ryan M. 2018. Testing English influence on first person singular *yo* subject pronoun expression in Sonoran Spanish. In J. E. MacDonald (ed.), *Issues in contemporary trends in Hispanic and Lusophone linguistics*, 355–372. Amsterdam: John Benjamins.
- Carvalho, Ana M., Rafael Orozco & Naomi Shin (eds.). 2015. *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Washington, DC: Georgetown University Press.
- Cerrón-Palomino, Álvaro. 2016. (No) English interference on US Southwest Spanish? A look at variable subject pronoun expression in Phoenix Spanish–English bilinguals. *Sociolinguistic Studies* 10(3). 383–408.
- Christoffersen, Katherine & Ryan M. Bessett. 2019. El corpus bilingüe del valle (CoBiVa). University of Texas Rio Grande Valley. <https://utrgv.edu/cobiva>.
- Flores-Ferrán, Nydia. 2009. Are you referring to me? The variable use of *uno* and *yo* in the oral discourse. *Journal of Pragmatics* 41. 1810–1824.
- Flores-Ferrán, Nydia. 2004. Spanish subject personal pronoun use in New York City Puerto Ricans: Can we rest the case of English contact? *Language Variation and Change* 16. 49–73.

- Flores-Ferrán, Nydia. 2007. Los mexicanos in New Jersey: Pronominal expression and ethnolinguistic aspects. In John Holmquist, Augusto Lorenzino & Lotfi Sayahi (eds.), *Selected proceedings of the third workshop on Spanish sociolinguistics*, 85–91. Somerville, MA: Cascadilla Proceedings Project.
- Hurtado, Luz Marcela. 2001. *La variable expresión del sujeto en el español de los colombianos y Colombo-Americanos residentes en el condado de Miami-Dade*. University of Florida, Gainesville. (Doctoral dissertation).
- Jaramillo, June A. 1995. The passive legitimization of Spanish. A macrosociolinguistic study of a quasi-border: Tucson, Arizona. *International Journal of the Sociology of Language* 114. 67–91.
- Limerick, P. 2017. Language contact in the US Southeast: The case of Spanish subject expression in an emerging bilingual community in Georgia. *Spanish in Context* 14(1). 53–77.
- Limerick, P. 2019. Subject expression in a Southwestern US Mexican community. *Borealis: An International Journal of Hispanic Linguistics* 8(2). 243–273.
- Orozco, Rafael. 2016. Subject pronoun expression in Mexican Spanish: ¿Qué pasa en Xalapa? *Proceedings of the Linguistic Society of America* 1(7). 1–15.
- Otheguy, Ricardo & Ana Celia Zentella. 2012. *Spanish in New York: Language contact, dialect leveling, and structural continuity*. Oxford: Oxford University Press.
- Otheguy, Ricardo, Ana Celia Zentella & David Livert. 2007. Language and dialect contact in Spanish in New York: Towards the formation of a speech community. *Language* 83(4). 770–802.
- Sankoff, David, Sali A. Tagliamonte & Eric Smith. 2018. *Goldvarb Z: A multivariate analysis application for Macintosh*.
- Shin, Naomi. 2014. Grammatical complexification in Spanish in New York: 3SG pronoun expression and verbal ambiguity. *Language Variation and Change* 26. 303–330.
- Shin, Naomi & Cecilia Montes-Alcalá. 2014. El uso contextual del pronombre sujeto como factor predicativo de la influencia del inglés en el español en nueva York. *Sociolinguistic Studies* 8(1). 85–110.
- Shin, Naomi & Ricardo Otheguy. 2005. Overt non-specific *ellos* in Spanish in New York. *Spanish in Context* 2(2). 157–174.
- Silva-Corvalán, Carmen. 1994. *Language contact and change: Spanish in Los Angeles*. Oxford: Oxford University Press.
- Tagliamonte, Sali A. 2003. Comparative sociolinguistics. In J. K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The handbook of language variation and change*, 128–156. Oxford: Blackwell.

2 First person singular subject pronoun expression in Arizona and Texas

- Torres Cacoulllos, Rena & Catherine E. Travis. 2010a. Testing convergence via code-switching: Priming and the structure of variable subject expression. *International Journal of Bilingualism* 15(3). 241–267.
- Torres Cacoulllos, Rena & Catherine E. Travis. 2010b. Variable *yo* expression in New Mexico: English influence? In Susana Rivera-Mills & Daniel Villa Crésap (eds.), *Spanish of the U.S. Southwest: A language in transition*, 189–210. Iberoamericana.
- Torres Cacoulllos, Rena & Catherine E. Travis. 2018. *Bilingualism in the community: Code-switching and grammars in contact*. Cambridge: Cambridge University Press.
- Travis, Catherine E. 2007. Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change* 19(2). 101–135. DOI: 10.1017/S0954394507070081.

Chapter 3

Variable indirect object pronoun expression: A usage-based analysis of Galician and Spanish

Javier Rivas 

University of Colorado, Boulder

Esther L. Brown 

University of Colorado, Boulder

This work analyzes variable first person singular indirect object pronoun expression (*me* vs. *me...a min/a mí*) in Galician and Spanish from a usage-based perspective to determine the linguistic factors that constrain overt (vs. omitted) strong pronominal forms in the two languages. Using Variationist methodology, 1288 tokens were extracted from conversational data and submitted to a generalized mixed effect model using R. Despite significant differences in rates of expression (Galician 25%, Spanish 15%), the same syntactic, discourse, and interactional factors significantly constrain indirect object expression in the two languages. In both Galician and Spanish, expression of *a min/a mí* is favored in utterance initial position, in constructions with *gustar*-type verbs, when primed by a previous *me* or indirect object, and when there is a lack of discourse continuity between the previous subject and *me* in the target clause. These results contribute new empirical findings to the body of literature on indirect object pronoun expression and bring to light marked similarities between factors conditioning indirect objects and those predicting variable subject pronoun expression. The results reveal advantages of studying pronoun expression across different syntactic functions.



1 Introduction

Recent years have brought us fine-grained analyses from a usage-based perspective that contribute to a deeper knowledge of traditional linguistic categories and constructions. For example, researchers have generated a plethora of empirical studies on subject pronoun expression in Spanish. These studies suggest that subjects actually subsume usage-patterns and constructions that are, in fact, quite different. Traditionally, we assign the function of subject to elements such as *yo* 'I' (e.g., *yo creo* 'I think'), *ella* 'she' (e.g., *ella ganó las elecciones* 'she won the election'), and *un café* 'one coffee', (*un café por la mañana te alegra el día* 'one coffee in the morning brightens your day'), because they share a number of commonalities including the same coding devices (e.g., verbal concord and, if pronominal, nominative case). However, these elements differ greatly in terms of animacy (semantics) and givenness, referentiality and definiteness (pragmatics).

In recognition of these differences, some studies of variable subject pronoun expression limit the scope of analysis to first person singular subject pronouns (Morales 1980, Bentivoglio 1987, Travis & Torres Cacoulios 2012, 2021, Posio 2013, Torres Cacoulios & Travis 2014, 2018, 2019, Ramos 2016, Travis et al. 2017). In narrowing the scope of variation, these studies have brought to light a lack of difference in factors constraining variable expression across languages that may have been obscured by considering all subjects simultaneously. In this line, subject expression in English and Spanish, two languages typically opposed as examples of non-pro-drop and pro-drop respectively, share probabilistic constraints on variation despite disparate rates of expression. Such research allows for the discovery of novel emergent patterns and leads us to question the utility of imposing *a priori* labels upon our data.

First person singular pronominal expression is not limited to subject constituents, but it also appears in other syntactic functions such as direct object and indirect object. Do first person pronouns behave similarly across syntactic functions? Do these syntactic functions (subject, object), traditionally described as binary opposites, behave uniformly with regard to conditioning factors present in the target context? In this study, we explore first person singular pronominal expression in indirect object function in order to unveil which linguistic factors constrain overt (vs. omitted) strong pronominal forms as well as identify potential similarities across two syntactic functions (subject and indirect object) that are generally examined independently from one another. This work is informed by cross-linguistic, typological studies (e.g., Givón 2001: 474) that highlight commonalities across these functions (subject, indirect object). Both typically encode highly animate (human) and definite referents that are high in the

agentivity scale and low in their degree of affectedness by the verbal action. Additionally, indirect objects in intransitive constructions have been described as *dative* (Bhaskararao & Subbarao 2004) or *quirky* subjects (Rivero 2004, Gutiérrez-Bravo 2006, Cuervo 2010) in the previous literature because they display behavior and control properties typical of subjects in syntactic processes such as equivalent N(oun) P(hrase) deletion in adverbial infinitival clauses (Campos 1999: 1560).

In order to explore potential commonalities between first person singular pronouns in indirect object and subject functions, we focus upon a case of variation in Galician and Spanish. Grammars from both Galician (Álvarez & Xove 2002: 110, Freixeiro Mato 2006: 126) and Spanish (Alarcos Llorach 1994: 199, Gutiérrez Ordoñez 1999: 1872) make evident that the two languages behave quite similarly by allowing for variable expression of first person singular indirect object pronouns *a min/a mí* (literally ‘to me’) in clauses including *me* ‘me’ as a verbal clitic. Similar constraints of pronominal expression across grammatical relations (i.e., subject, indirect object) would suggest that the boundaries between them are not hard and fast (see Aijón Oliva 2017, 2018, 2019, Serrano 2017, 2018), and that a new avenue of linguistic inquiry would be to focus on first person singular expression in all its functions.

In this study we provide a large-scale quantitative analysis of *a min/a mí* variation in naturally occurring discourse in Galician and Spanish in order to identify the linguistic factors that significantly constrain expression (vs. omission) of strong pronominal forms. This analysis allows us to explore the theoretical questions raised in the previous paragraphs. Additionally, we show that expression of *a min/a mí* is conditioned by multiple factors simultaneously, of syntactic, discourse, and interactional nature. These results, which contribute new empirical findings to the body of literature on indirect object, are interpreted from within a usage-based perspective (Bybee 2010).

2 Background

From a cross-linguistic typological perspective, the indirect object is one of the participants of the ditransitive construction (Givón 2001: 141, Kittilä 2007), together with the subject and the direct object. Prototypical instances of subject, direct object and indirect object respectively encode the semantic roles of agent, patient, and recipient, as in (1) and (2). In addition, in both Galician (Álvarez & Xove 2002: 110) and Spanish grammars (Gutiérrez Ordoñez 1999: 1872), the experiencer of intransitive *gustar* ‘like’ type verbs is also regarded as an indirect object, as shown in (3). In both languages, indirect objects may be expressed by

means of the following grammatical markers: a) a dative clitic, b) an [*a* 'to' + *NP*] construction, and c) an [*a* + *NP*] construction which is coreferential to a dative clitic marked on the verb. Examples in both languages are provided in (1), (2) and (3), respectively.

(1) Indirect objects expressed as clitics

a. Galician (OICO-URB-CBAS-SANTIAGODECOMPOSTELA-01-2013)
 eu quero que me digades a ver que
 1SG.SBJ want.1SG.PRS that 1SG.OBJ say.2PL.PRS.SBJV see.INF that
 vos parece esto ||
 2PL.OBJ seem.3SG.PRS that.DEM

'I want you to tell me what you think about this'

b. Spanish (MADR_H23_O33)
 me decías que tenía una serie de
 1SG.OBJ say.2SG.IPFV that have.3SG.IPFV INDF ART.F.SG series of
 ventajas.
 advantage.PL

'you were telling me it had a number of advantages'

(2) Indirect objects expressed by means of *a* + NP

a. Galician (OIED-SURB-CDUB-SANTIAGODECOMPOSTELA-01-1995)
 e estaba-me dedicando a mandar curriculums
 and be.1SG.IPFV-1SG.OBJ focus.PRS.PTCP to send.INF curriculum.PL
 ás empresas
 ART.F.PL company.PL

'and I was focusing on sending my CV out to companies'

b. Spanish (MADR_H32_O43)
 ¿ustedes saben el mal que están haciendo [...] ?
 2SG.PL know.3PL.PRS ART.M.SG evil that be3SG.PRS do.PRS.PTCP
 a la gente joven?
 to ART.F.SG people young

'do you know how much harm they are doing to young people?'

(3) Indirect objects expressed by clitics (*me*) as well as *a min/a mí*

a. Galician (OIED-SURB-CDUB-SANTIAGODECOMPOSTELA-01-1995)
 A min non me gustaba nada
 to 1SG.OBL NEG 1SG.OBJ like.3SG.IPFV nothing

'I didn't like it at all'

b. Spanish (MADR_H32_043)

¿A mí qué leches me importa?
 to 1SG.OBL what milk.PL 1SG.OBJ matter.3SG.PRS
 ‘What the hell do I care?’

In this paper, we will focus on first person singular indirect objects in both languages. In this context, the occurrence of the clitic (*me*) on the verb is obligatory in both present-day Galician (Expósito Loureiro 2016: 30) and Spanish (Company Company 2006: 536, Pericchi et al. 2020: 52). In this context, then, we have two options: the indirect object may be expressed either by the clitic *me* alone, as in (1), or by both the clitic and the strong pronominal form *a min/a mí* (as shown in (3)). The speech sample represented in (1) could have been accompanied by an expressed strong pronoun and the speaker in (3) could likewise have omitted *a min/a mí*. In this way, the variation we examine is probabilistic.

We restrict our analysis to first person singular indirect objects for several reasons. Results provided by previous studies on Spanish suggest that first person singular indirect objects have a higher token frequency than all the other persons (Dufter & Stark 2008: 119, Vázquez Rozas 2012: 849) and that *a mí* is the most frequent strong pronominal form in the discourse (Aijón Oliva 2018: 587), providing sufficient tokens for analysis. Additionally, the variability with other persons (i.e., third) introduces in the analysis dichotomies such as pronominal/lexical, given/new, definite/indefinite, human/inanimate which makes defining them as “two ways of saying the same thing” (Labov 1994, Tagliamonte 2012) (i.e. a variant) more problematic. First person singular indirect objects are always pronominal, given, definite and human, which makes the number of independent factors contributing to the variation between *me* and *a min/mí...me* more manageable. Focusing exclusively on first person singular, moreover, follows precedent (Morales 1980, Bentivoglio 1987, Posio 2013, Ramos 2016, Torres Cacoullos & Travis 2014, 2018, 2019, Travis & Torres Cacoullos 2012, 2021, Travis et al. 2017) and allows us to situate our work within research on pronominal expression generally.

In both Galician (Freixeiro Mato 2006: 187) and Spanish (Gili Gaya 1980: 231, Luján 1999) grammars, the use of strong pronominal forms such as *a min/a mí* is described as ‘emphatic’, or ‘contrastive’, since the information they provide regarding person and number is already expressed through the clitic form (e.g., *me*). More recent approaches (e.g., Serrano 2017, 2018, Aijón Oliva 2018) account for the occurrence of Spanish strong pronouns in object function (*a mí, a ti, a nosotros/nosotras, a vosotros/vosotras, a usted, a ustedes*) from a cognitive perspective, drawing upon notions such as salience and subjectivity. In this line, Serrano

(2018) assigns the use of *a mí* discourse functions such as informativeness and focus/stress on the referent. This researcher also finds the presence of *a mí* to correlate with argumentative speech styles.

To the best of our knowledge, the study presented here reports the results from the first variationist comparative analysis on this subject. The following sections describe the data and methods we employ in our analysis.

3 Data

In order to identify the usage patterns of first person indirect object expression in Galician and Spanish, we use two corpora of conversational data. We extract all examples of *me* in indirect object function with or without *a min/a mí*. We include verbs that can be interpreted as occurring in both transitive and intransitive constructions (Vázquez Rozas 2006) such as *aburrir* ‘bore’, *afectar* ‘affect’, *alegrar* ‘make happy’, *atraer* ‘attract’, *fastidiar* ‘bother’, *impresionar* ‘impress’, *interesar* ‘interest’, *molestar* ‘bother’ and *preocupar* ‘worry’ ($N = 35$). We use Miglio, Gries, Harris, Willer, and Santana-Paixão’s (2013) contextual factors (e.g., tense of the verb, position of the indirect object, position of the theme) to identify intransitive interpretations in these data.

For Galician, we extract our data from *Corpus Oral Informatizado da Lingua Galega* ‘Computerized Oral Corpus of the Galician Language’ (CORILGA). This corpus (Fernández Rei & Regueira Fernández 2019) consists of approximately 1,400,000 words of spoken Galician from recordings made from the 1960s up to the present day. The corpus includes different registers (informal, formal) as well as conversations, interviews, speeches and lectures. We base our analysis on the data found in 24 randomly selected conversations (circa 142,000 words) of informal spoken Galician recorded between the 1990s and the present day.

Our Spanish data is taken from *Corpus del Proyecto para el Estudio Sociolingüístico del Español de España y de América* ‘Corpus of the Project for the Sociolinguistic Study of Spanish from Spain and America’ (PRESEEA 2014). This corpus contains spoken data from different dialects and social groups across the Spanish-speaking world. In this corpus, we extract data from fifteen randomly selected interviews belonging to the corpus from Madrid. Together, these interviews amount to approximately 191,000 words for analysis.

From all non-truncated and complete examples, we exclude indirect object pronouns that do not fall within the envelope of variation. These include examples of indirect object pronouns used reflexively, as in (4), because they cannot co-occur with a strong first person singular pronoun *a min/a mí*. It is grammatically

possible, though, for reflexive *me* to co-occur with *a mín mesmo/mesma/a mí mismo/misma* ‘myself’ both in Galician (*Pregúntome a mín mesma iso* ‘I am asking myself about that’, Álvarez 2000: 84) and Spanish (*yo me he tomado el pulso a mí mismo* ‘I have taken my pulse’, Otero 1999: 1458). However, we did not find any examples of this type in either language in our corpora:

- (4) Indirect object pronoun *me* used reflexively (excluded)
- a. Galician (OICO-URB-CBAS-SANTIAGODECOMPOSTELA-03-2014)
- non me estou preguntando que tará facendo
 NEG 1SG.OBJ be.1SG.PRS ask.PRS.PTCP that be.3SG.FUT do.PRS.PTCP
 miña nai
 POSS.1SG.F mother
 ‘I am not wondering what my mom is doing’
- b. Spanish (MADR_M11_004)
- no si es que me he puesto un
 NEG if be.3SG.PRS that 1SG.OBJ have.1SG.PRS put.PST.PTCP ART.M.SG
 jersey gordo
 sweater thick
 ‘no, it is just that I am wearing a thick sweater’

We also exclude from the analysis constructions in which the verb is followed by the complementation pattern [NP + INFINITIVAL CLAUSE] if the NP is the ‘logical’ subject of the infinitive. This complementation pattern can be found in Spanish after perception (*ver* ‘see’, *oír* ‘hear’) and causative (*hacer* ‘make’, *mandar* ‘order’) verbs, as well as other verbs of manipulation such as *obligar* ‘force’ and *invitar* ‘invite’. These constructions fall outside the envelope of variation of our study because their syntactic structure is still a matter of controversy (see Enghels 2012 for perception verbs, Rivas 2013 for manipulative verbs and Marchís Moreno & Navarro 2015 for causative verbs): it is not clear what the syntactic function of the infinitive is or whether the intervening NP is a direct or an indirect object.

Additionally, from the Spanish data we exclude enclitic indirect objects in commands (e.g., *háblame* ‘talk to me’) and infinitival and gerundial periphrases such as *empezó a decirme* ‘he began to tell me’ ($N = 118$) because we also find 0% of *a mí* expression (no variation), even though constructions such as *háblame a mí* are grammatically permissible.¹ In the Galician data, we also exclude examples in which *me* is contracted with a third person direct object clitic (*o, a, os,*

¹Since these constructions are grammatically possible, a larger corpus might allow inclusion of these cases in the statistical analysis if enough variation were found.

as). This methodology gives us 760 tokens for Spanish and 528 for Galician (total $N = 1288$) of first person singular indirect object referents on which we base our analysis.

4 Method

In order to determine which factors constrain the occurrence of the strong pronominal form *a min/a mí* in both languages, we code each of these examples for a variety of linguistic predictors. We hypothesize that expression/omission of *a min/a mí* will respond to multiple factors simultaneously. As such, we consider the syntactic-semantic features of the verb accompanying *me*, the specific discourse context preceding the target, as well as the position that the clause occupies within the interactional context. The specific coding we employ is described in more detail in the following paragraphs.

4.1 Presence of *a min/a mí*: yes vs. no

This is our dependent variable. As mentioned above, the clitic *me* is obligatory in the present stages of both languages. We code each instance of indirect object *me* for whether it is accompanied by the strong pronominal form or not. We include both pre-verbal and post-verbal uses. The vast majority of examples of *a min/a mí* are pre-verbal (93%, $N = 228$), with no significant difference in position whether used in a transitive or an intransitive construction (unlike suggested by Dufter & Stark 2008: 117). The propensity to appear preverbally aligns with findings from previous research on Spanish (cf. Vázquez Rozas 2006: 97 and Aijón Oliva 2018: 593).

4.2 Syntactic construction: Transitive vs. intransitive

Previous studies propose that transitive and intransitive constructions condition indirect object usage in different ways. For example, the presence of the dative clitic marker on the verb in Spanish is more common in intransitive than in (di)transitive clauses (Dufter & Stark 2008: 116). Furthermore, as has already been mentioned, indirect objects in intransitive constructions are described as dative or quirky subjects because they share some grammatical characteristics with subjects (e.g., clausal position, behavior and control properties in syntactic processes). This fact might have an impact on pronoun expression. For this reason, we code each token of *me* for whether it occurs in a (di)transitive or

intransitive construction. Following Soto Andión (2008), we regard psychological constructions such as *dar medo/miedo* ‘frighten’ or *dar igual* ‘not care’ as intransitive constructions, because their probabilistic grammars resemble those of *gustar*-type constructions (Rivas 2016). We also include within this category impersonal uses of (di)transitive verbs, such as (5), because these constructions lack a direct object:

- (5) Impersonal (di)transitive verbs with *me*
- a. Galician (OIED-SURB-CDUB-SANTIAGODECOMPOSTELA-01-1995)

a min	empezouseme		a acumular
to 1SG.OBL	begin.3SG.PST-3SG.REFL-1SG.OBJ	to	accumulate.INF
o	traballo		
ART.M.SG	work		
‘The work started to pile up on me’			
 - b. Spanish (MADR_M21_O24)

...para controlar	que no	se	me	acaben	las
to	check.INF	that	NEG 3SG.REFL	1SG.OBJ	finish.3P.SBJV
					ART.F.PL
					pilas
					battery.PL
‘in order to check that the batteries don’t run out’					

4.3 Appearance of *me* in the preceding discourse context: Coreferential vs. non-coreferential vs. not applicable

We code for this factor in order to determine whether previous mention of an oblique first person singular participant in the preceding three finite clauses influences in any way the occurrence of the strong pronominal form. We count all types of *me*: direct object, indirect object and reflexive. We predict presence of a coreferential *me* will disfavor the occurrence of *a min/a mí* in the target clause if speakers are using *a min/a mí* to track reference. Example (6a) illustrates a case in which *a mí* is present in the target sentence (*a mí es que me da un poco...*) when the previous *me* is non-coreferential. In contrast, in example (6b), *a mí* occurs in the target sentence (*a mí siempre me ha gustado mucho ir a Egipto*), even though *me* in the previous sentence is coreferential with it.

(6) Presence of *me* in previous clause

a. Spanish (MADR_M23_O34)

E: está rico <laughter="T"/> a mí también me
 be.3SG.PRS good to 1SG.OBL also 1SG.OBJ
 gusta
 like.3SG.PRS

I: a mí es que me da un poco...
 to 1SG.OBL be.3SG.PRS that 1SG.OBJ give.3SG.PRS ART.M.SG little

‘E: it’s good. I also like it. I: The thing is that to me it is a little...’

b. Spanish (MADR_M12_O10)

I: entonces a lo mejor ir llamando la atención con
 so probably go.INF call.PRS.PTCP ART.F.SG attention with
 un coche tampoco es lo que me llame
 ART.M.SG car neither be.3SG.PRS what 1SG.OBJ call.3SG.SBJV
 / o con joyas o no //
 or with jewelry or NEG

E: y de viaje adónde irías
 and of trip to-where go.2SG.COND

I: pues / siempre a mí me ha gustado
 well always to 1SG.OBL 1SG.OBJ have.3SG.PRS like.PST.PTCP
 mucho ir a Egipto /
 much go.INF to Egypt

‘I: Then going around showing off with a car it is probably not what I like, either with jewels or not. E: and where would you go on a trip? I: Well, I have always liked to go to Egypt’

4.4 Presence of an indirect object in the previous clause

Priming is a psycholinguistic process by which speakers tend to repeat linguistic forms and constructions that are used in the previous context (Cameron & Flores-Ferrán 2004, Travis 2007, Abreu 2012, Travis & Torres Cacoulios 2012). To test whether there is any effect on expression from a preceding similar construction, for each token we determine whether *me* is preceded in the immediate clause by an indirect object expressed by a PP headed by *a*, be it pronominal (e.g., *a ti*) or lexical (e.g., *a Rosa*). We distinguish expressed (as in 7a) vs. other (as in 7b):

(7) Indirect object in clause preceding target clause

- a. Galician (OIED-URB-CDUB-SANTIAGODECOMPOSTELA-07-1995)²
 eu non sei o que lle afecta á
 1SG.SBJ NEG know.1SG.PRS what 3SG.DAT affect.3SG.PRS to.ART.F.SG
 xente | a min non me afectou...
 people to 1SG.OBL NEG 1SG.OBJ affect.3SG.PST
 ‘I don’t know what moves people | I wasn’t moved..’
- b. Spanish (MADR_M11_004)³
 y se te apunta más de uno // a mí
 and 3SG.REFL 2SG.OBJ sign-up.3SG.PRS more of one.M.SG to 1SG.OBL
 me encantaría por ejemplo
 1SG.OBJ love.3SG.COND for example
 ‘And more than one signs up // I would really love that’

In both (7a) and (7b), the target clause has the strong pronoun *a min/a mí*. In (7a), the previous clause has an expressed PP headed by *a* (*á xente* ‘to the people’) in indirect object function, so we code this example as *expressed*. In (7b), the indirect object of the previous clause is expressed exclusively by means of the clitic *te* ‘you’. We therefore code (7b) as *other*. We predict that presence of a similar construction in the preceding clause could favor *a min/a mí* expression if structural priming constrains the variation.

4.5 Subject reference of previous finite verb: Same vs. different

Unlike the linguistic factors just described, which are concerned with the occurrence of an (indirect) object in the preceding discourse, this factor concerns itself with the potential role of preceding subject reference on the appearance of *a min/a mí*. As previously mentioned, among other similarities with subjects, indirect objects tend to be animate and definite (Hopper & Thompson 1980: 259). If *a min/a mí* has subject-like qualities, it might be constrained in ways similar to variable subject pronominal expression in Spanish. Numerous studies concur (e.g., Carvalho et al. 2015) that a factor that strongly conditions expressed vs. unexpressed subject pronouns is reference continuity in discourse. That is, overt vs. null pronominal subjects are significantly conditioned by whether the subject of the preceding finite verb is the same (non-switch) or different (switch) from the

²As is noted by one of the reviewers, the expression of *a mí* in this example could also be conditioned by referential contrast.

³In this example, *te* is a dative of interest with a generic referent.

subject of the target. When referents (target, preceding finite verb) are different, expression is favored. Consequently, we code each token of *me* for whether the subject of the preceding finite verb (spoken by any participant) was the same as *me* (*eu/yo* ‘I’ and/or first person singular verbal agreement) or not:

(8) Previous finite verb subject reference

a. Galician (OIED-SURB-CDUB-SANTIAGODECOMPOSTELA-01-1995)

e ao día seguinte xa empecei a notar que
and to.ART.M.SG day following already start.1SG.PST to notice.INF that
non me doía
NEG 1SG.OBJ hurt.3SG.IPFV

‘and the following day I started noticing it did not hurt’

b. Spanish (MADR_H22_026)

la casa va a estar prácticamente desnuda / y /
ART.F.SG house go.3SG.PRS to be.INF practically empty and
cuando algo me guste me lo iré
when something 1SG.OBJ like.3SG.SBJV 1SG.OBJ 3M.SG.OBJ go.1SG.FUT
comprando
buy.PRS.PTCP

‘The house is gonna be practically empty, and when I like something I will buy it’

Example (8a) illustrates a case in which the subject of the preceding finite verb is expressed by means of a first person singular verbal agreement (*empecei* ‘I started’). We therefore code this example as *same*. In contrast, in (8b) the subject of the previous clause is *la casa* ‘the house’ so this token is coded as *different*. If *a min/a mí* has subject-like qualities, we anticipate preceding subject reference (switch/non-switch) should condition *a min/a mí* expression

4.6 Position of the clause containing the indirect object clitic in the interactional context: Initial vs. other

For all tokens of *me* in both languages, based upon the transcribed data, indicating pauses and turn-taking, we code for whether the clause in which *me* is included is pause-adjacent or not. Tokens were coded for whether they followed a pause in order to test for any potential effects of turn-taking or Intonation Unit-initial effects (cf. results reported for Spanish, Bentivoglio 1987: 40, 62, Travis & Torres Cacoullos 2012: 743, and English, Torres Cacoullos & Travis 2014: 27, regarding subject pronoun expression). Example (9a) is coded as *initial*, whereas (9b) is coded as *other*:

(9) Position of *me* in the clause

- a. Galician (OIED-SURB-CDUB-SANTIAGODECOMPOSTELA-02-1995)
 pero en setembro saliron convocadas | ou xa en
 but in September come.3PL.PST invoke.PST.PTCP or already in
 agosto sería | as da universidá de Vigo || e
 August be.3SG.CON ART.F.PL of.ART.F.SG university of Vigo and
 entós decía-me o tipo da academia...'
 then say.3SG.IPFV-1SG.OBJ ART.M.SG guy of.ART.F.SG school
 'but there was a call in September | it might have been in August
 already | for University of Vigo || and the guy from the school would
 tell me...'
- b. Spanish (MAD_H21_020)
 depende si me toca o no me toca
 depend.3SG.PRS if 1SG.OBJ touch.3SG.PRS or NEG 1SG.OBJ touch.3SG.PRS
 'It depends on whether I win or not'

4.7 Language: Galician vs. Spanish

Although grammars suggest that the use of overt strong pronominal forms in indirect object function in both Galician and Spanish are not at odds, we include *language* in our analysis to test whether this is the case in these oral data.

In the following section, we summarize the results of our quantitative analyses.

5 Results

The percentage of expression for strong first person singular indirect object pronouns in our data is 19%, as is shown in Table 1. If we examine rates across languages separately, we find that the percentage of expressed strong pronouns is higher in Galician (25%) than in Spanish (15%).

Differences in rates of expression across these two languages do not necessarily entail a different probabilistic grammar regarding indirect object usage. In order to test whether the probabilistic grammar constraining usage likewise differs, we submit the data to a generalized mixed effect model using R (R Core Team 2019). We include in the analysis all of the factors described in the previous section: presence of *a min/a mí*, syntactic construction, appearance of *me* in the preceding discourse context, presence of an indirect object in the previous clause, subject reference of previous finite verb, position of the clause containing the indirect object clitic in the interactional context, and language. We also

Table 1: Overt *a min/a mí* expression in oral Galician and Spanish

	expressed	<i>N</i>	% expression
Galician	133	528	25
Spanish	113	760	15
Total	246	1288	19
$p < 0.0000, \chi^2 = 21.4787$			

include speaker ($N = 41$) and verb form ($N = 514$) as random effects. We find no significant interactions between these fixed effects. Table 2 summarizes the results of the best model, determined by the lowest AIC.

The two factor groups that most significantly constrain *a min/a mí* expression are appearance of *me* in the preceding discourse context and syntactic construction. The presence of *a min/a mí* in the target construction is strongly favored in our data when there is either a non-coreferential *me* or no *me* in the previous discourse context. This result suggests that the presence of the strong pronominal form *a min/a mí* is highly disfavored in contexts of continuity of reference across neighboring first person clitic pronouns in object function.

A novel contribution of this work is the finding that previous subject reference also constrains indirect object expression. A preceding subject coreferential to *me* in the target clause disfavors the expression of *a min/a mí*. Similar to the effect of a previous *me* in the discourse, this result suggests that indirect object pronouns are sensitive to discourse continuity, an attribute commonly ascribed to subject referents. When there is a switch in reference between the previous subject and *me* in the target clause, *a min/a mí* expression is more probable than when there is no switch in reference.⁴

The presence of the strong pronominal form is also statistically conditioned by position of the clause containing the indirect object clitic in the interactional context. When the clause containing the clitic *me* occurs in any context other than after a pause, *a min/a mí* expression is disfavored. This result suggests that the presence of the strong pronominal form *a min/a mí* is favored in contexts that highlight the interactional nature of conversational language (cf. the results provided in Travis & Torres Cacoullos 2012: 737 regarding the use of expressed *yo* 'I' in Spanish in combination with cognitive verbs).

⁴As noted by one of the external reviewers, the significant effect of switch reference in these data may relate to the contrastive and emphatic functions attributed to the strong pronominal form by traditional grammars.

3 Variable indirect object pronoun expression

Table 2: Generalized linear mixed model predicting *a min/a mí* expression ($N = 1288$). AIC = 1093.2, Random effects: Speaker ($N = 43$), Verb ($N = 513$). Positive coefficients are associated with *a min/a mí* expression. Significance codes: p : *** 0.001 ** 0.01 * 0.05 . 0.1 ns 1.

Random effect	Var	SD			
Speaker (intercept)	0.2235	0.4728			
Verb form (intercept)	0.4938	0.7027			
Fixed effects	N	% overt	Est. coef.	SE	p
Intercept			-2.47	0.48	<0.0001 ***
Previous <i>me</i>					
Non-coreferential, na - ref	820	25	1.52	0.22	<0.0000 ***
Coreferential	468	9			
Syntactic construction					
Intransitive - ref	791	26	1.36	0.22	<0.0001 ***
(Di)transitive	497	10			
Previous IO					
Null, na - ref	1246	19	-0.70	0.42	<0.1 .
Expressed	42	33			
Previous subject reference					
Same - ref	251	18	-0.46	0.20	<0.05 *
Different	1037	19			
Language					
Spanish - ref	760	15	-0.12	0.27	0.78 ns
Galician	528	25			
Pos. in interactional context					
Non-initial - ref	676	13	-0.70	0.21	<0.01 **
Initial	612	25			

As far as syntactic construction is concerned, Table 2 shows that intransitive constructions strongly favor *a min/a mí* expression. This result suggests that indirect objects in (di)transitive and intransitive constructions behave differently, a pattern that was already pointed out by those studies that suggest that the indirect object of intransitive *gustar*-type constructions is a dative or quirky subject. In addition to the subject-like nature of the strong pronominal form in behavior and control processes at the syntactic level, our results suggest that these forms are also more subject-like than the indirect object of (di)transitive clauses at the discourse level.

When the previous clause has no indirect object or has an indirect object exclusively encoded by means of the dative clitic, the presence of *a min/a mí* is disfavored in the data. This result is only marginally significant. There are relatively few target tokens that are preceded by a clause with an expressed indirect object owing to the low textual frequency of this construction. As we show in Table 2, in these contexts we find a higher rate of expression of *a min/a mí* (33% compared to 19%). This could be suggestive of a priming effect of the preceding strong pronominal form. We will return to this result in the Discussion Section.

The only factor that turned out to be non-significant in our analysis was language. Despite higher rates of strong pronoun expression in Galician than in Spanish, once controlling for the independent factors present in the discourse and interactional context as well as factors related to the verb, language does not independently predict expression of the strong pronominal form. We find no significant interactions between language and the other predictors included in the model, which suggests their effect is similar across the two languages.

6 Discussion

In this paper we analyze variable first person singular indirect object pronoun expression (*me* vs. *me...a min/a mí*) in both Galician and Spanish and our corpus-based analysis demonstrates which linguistic factors significantly constrain expression vs. omission of the strong pronominal form *a min/a mí*. Despite significant differences in rates of expression (Table 1), the statistical analysis determines that expression is conditioned similarly across the two languages considered in this work. In both Galician and Spanish, expression of *a min/a mí* is constrained similarly between interactional (utterance position), syntactic (construction), and discourse (presence of a previous *me*, previous subject reference, and previous indirect object) factors. This result echoes literature on subject pronoun expression. Languages such as English and Spanish differ significantly with

regard to rates of overt subject pronominal expression. Nevertheless, it has now been demonstrated that subject expression in both languages responds to similar linguistic predictors (Torres Cacoullós & Travis 2019). The comparable probabilistic grammar across languages (despite significant difference in rates of variant expression) is also true for dialects of Spanish (Carvalho et al. 2015).

In fact, this generalization holds when applied to variants of multiple linguistic variables. Regarding subject position, Llompart (2016) finds that higher rates of preverbal subjects in the Spanish of Arizona than in the Spanish of Mexico City do not entail a difference in factors constraining the usage patterns. In this same line, Rivas & García Pineda (forthcoming) show that, despite significantly higher rates of expression of Present Perfect vs. Preterit in Costa Rican (20%) than in Mexican (15%) Spanish, the linguistic factors constraining the occurrence of the Present Perfect are consistently the same in both dialects. Similarly, with regard to the extension of *estar* (Silva-Corvalán 1986), although rates of innovative *estar* vary cross-dialectally, again the constraints governing usage coincide (Rivas forthcoming).

In addition to showing a lack of difference between languages, we interpret our results as suggesting a lack of difference in conditioning of the variable expression/omission of first person singular pronouns across syntactic functions (indirect object and subject). For instance, as already mentioned, the overt strong pronoun *a min/a mí* appears preverbally in the vast majority of our examples ($N = 228, 93\%$).⁵ Likewise, Travis & Torres Cacoullós (2012: 713) report that 95% of instances of overt *yo* occur before the verb. Together with this similarity we find regarding clausal position, we identify multiple conditioning factors of indirect object pronoun expression that mirror those employed in previous research on subject pronoun expression. Following Travis & Torres Cacoullós (2012), we classify the significant findings in our study into four different categories: mechanical (priming), cognitive (reference tracking), interactional, and constructional (lexically specific constructions). Each of these categorizations, framed within a usage-based approach, is described in turn.

6.1 Mechanical factors

In the case of indirect objects, our statistical analysis shows only a marginally significant effect of a preceding [*a NP*] construction priming the subsequent target use (*a min/a mí*). Nevertheless, as we note, owing to the low token frequency of indirect objects expressed by *a NP* in discourse, the number of target cases

⁵This result is in line with findings reported by Aijón Oliva (2019: 106) for first-person dative pronouns.

preceded by an indirect object construction is low. To more directly examine a potential role for priming, we study only the target tokens preceded by a clause with an indirect object ($N = 315$). Of these targets, 42 are preceded by an expressed indirect object, that is to say, an $[a NP]$ construction (all the persons: first, second, third) in indirect object function. In contrast, 273 target tokens are preceded by a clause with a null indirect object, i.e., an indirect object exclusively expressed by means of the clitic. When the indirect object is expressed in the previous clause, our targets are overt at a rate of 33%, compared to just 11% when the indirect object in the preceding clause is null (i.e., just a clitic). This tendency is in the direction we would expect if primed by the previous construction. On the basis of this pattern, we can conclude that in our data there is perhaps a weak structural priming effect. However, as can be appreciated in Table 3, if we consider only the 42 targets preceded by an overt indirect object, we find a different picture. Rates of expression for targets preceded by first person singular $a min/a mí$ are lower than rates for targets preceded by other person-number combinations (e.g., $a ti$), (26%, $N = 31$ vs. 55%, $N = 11$). This result does not support a lexical priming interpretation. Again, these interpretations are speculative given the low token counts of these examples. If this result were replicated with a larger data sample, it could be suggestive of a contrastive function of $a min/a mí$.

Table 3: Rates of overt $a min/a mí$ expression in targets preceded by a clause containing an overt indirect object $[a NP]$ construction ($N = 42$)

Preceding clause	N	% expression
First person ($a min/a mí$)	31	26
Second, third persons ($a ti/a Rosa$)	11	55

6.2 Reference-tracking factors

Speakers use pronouns and agreement to track referents in ongoing discourse. For subject pronoun expression, continuity in reference between the target and the subject of the previous finite clause correlates with higher rates of null subjects. Similarly, our results show that the presence of a coreferential me in the previous three clauses significantly disfavors $a min/a mí$ expression. The coreferentiality of me is derived not only from when it functions as an indirect object, but also from when it is a direct object or a reflexive.

The results in Table 2 report the effects of a previous coreferential me in discourse. A more detailed analysis of the rates of expression can be seen in Figure 1.

Figure 1 summarizes target rates of overt *a min/a mí* in three contexts: preceded by overt coreferential forms, preceded by overt non-coreferential forms, and preceded by clauses lacking overt *me*. The number of overt target tokens preceded in discourse in the prior three clauses by an overt non-coreferential *me* is quite low ($N = 41$). Notwithstanding this small number of examples, the results illustrated in Figure 1 do seem to support the possibility of a tracking (cognitive) effect. When the preceding *me* is coreferential ($N = 468$), rates of expression are the lowest (8.5%). When the preceding *me* is non-coreferential (and potentially competing for reference), rates of *a min/a mí* are the highest (48.7%). This result could account for the function of *contrast* attributed to the overt *a min/a mí* expression in traditional grammars.

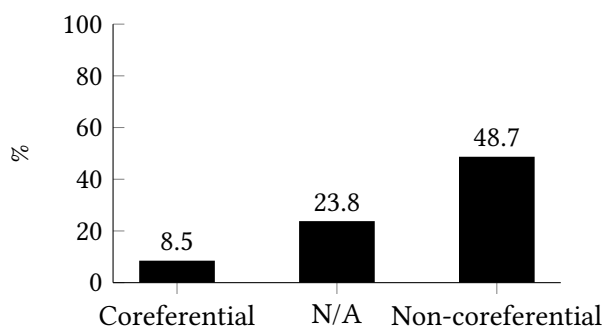


Figure 1: % of *a min/a mí* expression according to appearance of *me* in preceding contexts ($N = 1288$)

Interestingly, our results also reveal that the presence of the overt indirect object strong pronominal form is not only conditioned by the cognitive features of the indirect objects occurring in the preceding discourse but also by the cognitive features of preceding subjects. In this respect, we find that when the referent of the previous subject is the same as the referent of the indirect object, *a min/a mí* is disfavored in our data. This result suggests that first person singular pronouns are interrelated in the discourse regardless of the syntactic function (indirect object, subject) they fulfill in the clause.

6.3 Interactional factors

The quantitative analyses summarized in Table 2 reveal a significant effect of utterance position on expression/omission of *a min/a mí*. When the indirect object clause follows a pause (as indicated in the orthographic transcriptions), the expression of the strong pronoun is favored as compared to when the clause is

embedded elsewhere in the discourse. This result is independent of the role that the subject reference of the previous clause plays in a *min/a mí* expression. In most cases (72%, $N = 440$), when the target clause occurs after a pause, the reference of the indirect object is different from the reference of the subject of the previous clause. However, within switch reference contexts, the percentage of a *min/a mí* expression is significantly higher ($p < 0.0000$, $\chi^2 = 22.48862$) when it occurs in a clause following a pause (28%, $N = 123$) than when it occurs in other interactional positions (15%, $N = 69$). This result suggests that, regardless of reference-tracking factors, the occurrence of expressed *a min/a mí* may also be conditioned by an interactional effect.⁶ As is noted by Travis & Torres Cacoullós (2012: 737) with regard to overt *yo*, this interactional effect may well have led to the traditional interpretations of contrast and emphasis associated with the expression of strong pronominal forms in object function.

This finding also suggests a potential role for Intonation Units (IUs) (Du Bois et al. 1993) in constraining indirect object pronominal expression in line with results reported for subject pronominal expression (Travis & Torres Cacoullós 2012, Torres Cacoullós & Travis 2014). Future research on corpora that are IU-transcribed may be able to ascertain if the same pattern holds true for indirect objects.

6.4 Constructional factors

Our quantitative analysis indicates that intransitive constructions (i.e., grammatical patterns) favor expression of the strong indirect object pronoun over (di)transitive constructions. Similarly, Orozco & Hurtado (2021) have also shown that syntactic construction significantly constrains subject pronoun expression in Spanish. In order to determine whether all verbs within these two categories behave similarly or not, we examine rates of expression for each translation equivalent with 10 or more tokens in our data. Table 4 summarizes our findings. We group together Galician and Spanish forms (for example, *soar* and *sonar* ‘to sound’, are considered jointly). Recall, the average rate of overt *a min/a mí* expression is 19% (see Table 1). We have approximated this in the table with a dotted line. All verbs listed above the dashed line have rates of expression higher than 19%, and those below have lower than average rates. The two bolded types in the list (*gustar*, *dicir/decir*) are the two verbs with the highest token frequency in

⁶An anonymous reviewer suggests that an interpretation of this result is that the speaker is orienting the interpretation of subsequent discourse as specifically regarding his/her perspective (e.g., Fauconnier & Turner 2006).

3 Variable indirect object pronoun expression

Table 4: Rates of overt *a min/a mí* expression in translation equivalents occurring 10+ more times in data

Verb infinitive	N	% expression
<i>pasar</i> ‘happen’	26	61.5
<i>importar</i> ‘matter’	24	54.2
<i>soar/sonar</i> ‘sound’	12	33.3
<i>parecer</i> ‘seem’	108	25.9
<i>encantar</i> ‘love’	50	24.0
<i>gustar</i> ‘like’	289	23.5
<i>dar pena</i> , etc. ‘cause pain’	46	21.7
<i>ocorrer/ocurrir</i> ‘occur’	15	20.0
<i>falar/hablar</i> ‘speak’	11	18.2
<i>saír/salir</i> ‘leave’	11	18.2
<i>dar</i> ‘give’	67	11.9
<i>custar/costar</i> ‘cost’	10	10.0
<i>tocar</i> ‘to be one’s turn’	33	9.1
<i>quedar</i> ‘have left’	11	9.1
<i>dicir / decir</i> ‘say’	178	5.6
<i>contar</i> ‘tell’	25	4.0
<i>facер/hacer</i> ‘do, make’	25	4.0
<i>poñer/poner</i> ‘put’	15	0.0
<i>quitar</i> ‘remove’	10	0.0

each category (intransitive and (di)transitive respectively). The rates of expression for these two particular verbs differ significantly ($\chi^2 = 25.40103$, $p < 0.000$), with *gustar* exceeding the average (23.5%) and *dicir/decir* falling well below the average (5.6%). The high token frequency verb *dicir/decir* has remarkably low rates of overt strong pronoun expression and may work to suppress the rates of expression in the whole category (along with other speech verbs such as *contar* and *hablar*).

The verbs with greater than average rates of *a min/a mí* expression (among our most frequent types), belong to the same semantic category along with *gustar*. Previous studies (Delbecque & Lamiroy 1996: 101, Gutiérrez Ordoñez 1999: 1879) include these verb types into the category of “psych-movement” or “psychological verbs”. Again, here, we can see a parallel with the literature on subject (as opposed to object) pronoun expression. Many studies (e.g., Enríquez 1984, Otheguy

& Zentella 2012, Posio 2013, 2014, 2015, Herbeck 2021, Travis & Torres Cacoullos 2021) report higher rates of overt subject pronouns with cognitive-psych verbs than with other verb types. Usage-based analyses of subject pronoun expression (e.g., Brown & Shin 2022) suggest a verb's history of use conditioning context may help account for the overall higher rate of subject pronoun expression for this verb class. Given the variability apparent in Table 4 within the categories of transitive and intransitive (cf. *pasar* 61.5%, *quedar* 9.1%), the relative contribution of construction as opposed to verb semantics in predicting strong pronoun expression remains to be determined.

7 Conclusion

It has been common practice in linguistics to identify grammatical relations on the basis of coding devices such as case, presence/absence of an adposition, agreement and clausal position. In this line, grammatical relations such as subject, direct object and indirect object have become part of the core metalanguage to describe the structure of (accusative) languages such as Galician and Spanish. Subject, direct object and indirect object can be expressed by means of different grammatical markers including agreement, clitics, strong personal pronouns, and lexical NPs. Undoubtedly, this methodology has contributed enormously to our understanding of linguistic structure. However, by assuming that grammatical relations play a central role in language, we tend to take them as the starting point of our analyses, turning in this way our back to any commonalities that may exist across them. For example, the results of this paper suggest that first person singular behaves similarly across different grammatical relations (indirect object and subject) regarding cognitive, mechanical and interactional factors. Our results also suggest that the expression of first person singular is determined by the occurrence of a first person singular in the preceding discourse, regardless of its grammatical relation. These factors may outweigh syntactic functions in accounting for grammatical variation. Future research might consider giving precedence to grammatical categories (such as first person singular) over grammatical relations in their accounts of processes of language variation and change.

References

Abreu, Laurel. 2012. Subject pronoun expression and priming effects among bilingual speakers of Puerto Rican Spanish. In Kimberly L. Geeslin & Manuel Díaz-

- Campos (eds.), *Selected proceedings of the 14th Hispanic Linguistics Symposium*, 1–8. Cascadilla Proceedings Project.
- Aijón Oliva, Miguel Á. 2017. Speaker self-profiling through discursive indexation and syntactic encoding in Spanish radio talk. *Borealis: An International Journal of Hispanic Linguistics* 6/1. 175–199.
- Aijón Oliva, Miguel Á. 2018. The participants as objects: Variation and meaning of first- and second-person object encoding in Spanish. *Studia Linguistica* 72(3). 571–607.
- Aijón Oliva, Miguel Á. 2019. *Constructing us: The first and second persons in Spanish media discourse*. Berlin: Mouton de Gruyter.
- Alarcos Llorach, Emilio. 1994. *Gramática de la lengua española*. Madrid: Espasa Calpe.
- Álvarez, Rosario. 2000. O redrobo de A + FN en galego moderno. In R. Vieira de Castro & P. Barbosa (eds.), *Actas do XV encontro nacional da asociación portuguesa de lingüística*. Vol. 1, 75–93. Braga: Associação Portuguesa de Lingüística.
- Álvarez, Rosario & Xosé Xove. 2002. *Gramática da lingua galega*. Vigo: Galaxia.
- Bentivoglio, Paola. 1987. *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela, Consejo de Desarrollo Científico y Humanístico.
- Bhaskararao, Peri & Karumuri V. Subbarao (eds.). 2004. *Non-nominative subjects*. Amsterdam/Philadelphia: John Benjamins.
- Brown, Esther L. & Naomi Shin. 2022. Acquisition of cumulative conditioning effects on words: Spanish-speaking children's [subject pronoun + verb] usage. *First Language* 42(3). 361–382.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Cameron, Richard & Nydia Flores-Ferrán. 2004. Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context* 1(1). 41–65. DOI: 10.1075/sic.1.1.05cam.
- Campos, Héctor. 1999. Transitividad e intransitividad. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*. Vol. 2: *Las construcciones sintácticas fundamentales. Relaciones temporales, aspectuales y modales*, 1519–1574. Madrid: Espasa Calpe.
- Carvalho, Ana M., Rafael Orozco & Naomi Shin (eds.). 2015. *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Washington, DC: Georgetown University Press.

- Company Company, Concepción. 2006. El objeto indirecto. In Concepción Company Company (ed.), *Sintaxis histórica de la lengua española. Primera parte: La frase verbal*, 477–572. Ciudad de México: Fondo de Cultura Económica y Universidad Nacional Autónoma de México.
- Cuervo, Cristina. 2010. Two types of (apparently) ditransitive light verb constructions. In Karlos Arregi, Zsuzsanna Fagyal, Silvina Montrul & Annie Tremblay (eds.), *Romance linguistics 2008: Interactions in Romance*, 139–156. Amsterdam/Philadelphia: John Benjamins.
- Delbecque, Nicole & Béatrice Lamiroy. 1996. Towards a typology of the Spanish dative. In William Van Belle & Willy Van Langendonck (eds.), *The dative*, vol. 1: Descriptive studies, 73–117. Amsterdam/Philadelphia: John Benjamins.
- Du Bois, John W., Stephan Schuetze-Coburn, Susanna Cumming & Danae Paolino. 1993. Outline of discourse transcription. In Jane Edwards & Martin Lampert (eds.), *Talking data: Transcription and coding in discourse*, 45–59. Hillsdale, MI: Lawrence Erlbaum Associates.
- Dufter, Andreas & Elizabeth Stark. 2008. Double indirect object marking in Spanish and Italian. In Elena Seoane & María J. López Couso (eds.), *Theoretical and empirical issues in grammaticalization*, 111–129. Amsterdam/Philadelphia: John Benjamins.
- Enghels, Renata. 2012. Transitivity of Spanish perception verbs: A gradual category? *Borealis: An International Journal of Hispanic Linguistics* 2(1). 35–56.
- Enríquez, Emilia V. 1984. *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: C. S. I. C.
- Expósito Loureiro, Andrea. 2016. O redobro do complemento indirecto en galego no verbo pedir. *Cumieira: Cadernos de investigación da nova Filoloxía Galega* 1. 11–36.
- Fauconnier, Giles & Mark Turner. 2006. Mental spaces: Conceptual integration networks. In Dirk Geeraerts (ed.), *Cognitive linguistics: Basic readings*, 303–371. Berlin/New York: Mouton de Gruyter.
- Fernández Rei, Elisa & Xosé L. Regueira Fernández. 2019. CORILGA: Un corpus para o estudo da variación e do cambio lingüístico no galego falado. In Hella Olbertz Marta Blanco & Victoria Vázquez Rozas (eds.), *Corpus y construcciones: Perspectivas hispánicas*, 221–243. Universidade de Santiago de Compostela: Servizo de Publicacións.
- Freixeiro Mato, Xosé Ramón. 2006. *Gramática da lingua galega II*. 2nd edn. Vigo: A Nosa Terra.
- Gili Gaya, Samuel. 1980. *Curso superior de sintaxis española*. 13th edn. Barcelona: Biblograph.

- Givón, Talmy. 2001. *Syntax: An introduction*. Amsterdam/Philadelphia: John Benjamins.
- Gutiérrez Ordoñez, Salvador. 1999. Los dativos. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española. Vol. 2: Las construcciones sintácticas fundamentales. Relaciones temporales, aspectuales y modales, 1855–1930*. Madrid: Espasa Calpe.
- Gutiérrez-Bravo, Rodrigo. 2006. A reinterpretation of quirky subjects and related phenomena in Spanish. In Chiyo Nishida & Jean P. Montreuil (eds.), *New perspectives on Romance linguistics*, vol. 1: Morphology, Syntax, Semantics, and Pragmatics, 127–142. Amsterdam/Philadelphia: John Benjamins.
- Herbeck, Peter. 2021. Perspectival factors and pro-drop: A corpus study of speaker/addressee pronouns with *creer* ‘think/believe’ and *saber* ‘know’ in spoken Spanish. *Glossa* 122(1). 1–34. DOI: 10.16995/glossa.5873.
- Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299.
- Kittilä, Seppo. 2007. On the encoding of transitivity-related features on the indirect object. *Functions of Language* 14(1). 149–164.
- Labov, William. 1994. *Principles of linguistic change*, vol. 1. Oxford: Blackwell.
- Llompарт, Miquel. 2016. Subject-verb order variation with unaccusative verbs of change of location in Mexico and Southern Arizona. *Studies in Hispanic and Lusophone Linguistics* 9(1). 143–173.
- Luján, Marta. 1999. Expresión y omisión del pronombre personal. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española, 1275–1315*. Madrid: Espasa.
- Marchís Moreno, Mihaela & Mario Navarro. 2015. El verbo causativo hacer en el español loísta: Nuevas perspectivas para la nueva gramática de la lengua española. *Revista Internacional de Lingüística Iberoamericana* XIII, No. 1(25). 183–197.
- Morales, Amparo. 1980. La expresión de sujeto pronominal, primera persona, en el español de Puerto Rico. *Boletín de la Academia Puertorriqueña de la Lengua Española* 8(3). 91–102.
- Orozco, Rafael & Luz Marcela Hurtado. 2021. A variationist study of subject pronoun expression in Medellín, Colombia. *Languages* 6(1). Article 5. DOI: 10.3390/languages6010005.
- Otero, Carlos-Peregrín. 1999. Pronombres reflexivos y recíprocos. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, vol. 1: Sintaxis básica de las clases de palabras, 1427–1517. Madrid: Espasa Calpe.

- Otheguy, Ricardo & Ana Celia Zentella. 2012. *Spanish in New York: Language contact, dialect leveling, and structural continuity*. Oxford: Oxford University Press.
- Pericchi, Natalia, Kristin Davidse, Bert Cornillie & Freek Van de Velde. 2020. A diachronic study of indirect object doubling in Argentinian Spanish. *Acta Linguistica Hafniensia* 52(1). 45–80.
- Posio, Pekka. 2013. The expression of first-person-singular subjects in spoken Peninsular Spanish and European Portuguese: Semantic roles and formulaic sequences. *Folia Linguistica* 47(1). 253–291. DOI: 10.1515/flin.2013.010.
- Posio, Pekka. 2014. Subject expression in grammaticalizing constructions: The case of *creo* and *acho* ‘I think’ in Spanish and Portuguese. *Journal of Pragmatics* 63. 5–18. DOI: 10.1016/j.pragma.2013.07.001.
- Posio, Pekka. 2015. Subject pronoun usage in formulaic sequences. In A. Carvalho, R. Orozco & N. Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective*, 59–78. Washington, DC: Georgetown University Press.
- R Core Team. 2019. *R: A language and environment for statistical computing*.
- Ramos, Miguel. 2016. Continuity and change: First person singular subject pronoun expression in earlier Spanish. *Spanish in Context* 13(1). 103–127.
- Rivas, Javier. 2013. A usage-based analysis of morphosyntactic variation in the complementation of manipulative verbs. *Revista Internacional de Lingüística Iberoamericana* XI(1). 155–171.
- Rivas, Javier. 2016. Verb-object compounds with Spanish *dar* ‘give’: An emergent *gustar* ‘like’-type construction. *WORD: A Journal of the International Linguistics Association* 62(1). 1–21.
- Rivas, Javier. Forthcoming. Ser, estar y los verbos semicopulativos. In Guillermo Rojo, Victoria Vázquez Rozas & Rena Torres Cacoullos (eds.), *Sintaxis del español: The Routledge handbook of Spanish syntax*. Oxford: Taylor & Francis.
- Rivas, Javier & Érick García Pineda. Forthcoming. Rates and constraints of present perfect and preterit in Costa Rican Spanish: A variationist approach. In Brandon Baird, Osmer Balam & María del Carmen Parafita Couto (eds.), *Linguistic advances in Central American Spanish*. Leiden/Boston: Brill.
- Rivero, María L. 2004. Spanish quirky subjects: Person restrictions and the person-case constraint. *Linguistic Inquiry* 35. 494–502.
- Serrano, María José. 2017. A variable cognitive and communicative resource in Spanish: The first-person plural subject and object. *Journal of Pragmatics* 108. 131–147.
- Serrano, María José. 2018. Managing subjectivity: Omission and expression of first-person singular object *a mí* in Spanish media discourse. *The Canadian Journal of Linguistics* 63(3). 423–443.

- Silva-Corvalán, Carmen. 1986. Bilingualism and language change: The extension of *estar* in Los Angeles Spanish. *Language* 62(3). 587–608.
- Soto Andión, Xosé. 2008. Esquemas intransitivos con *Dar* en gallego: Variación semántica y combinatoria sintáctica. *Revue Roumaine de Linguistique* LIII(4). 405–425.
- Tagliamonte, Sali A. 2012. *Variationist sociolinguistics: Change, observation, interpretation*. Malden, MA/Oxford: Wiley-Blackwell.
- Torres Cacoullós, Rena & Catherine E. Travis. 2014. Prosody, priming and particular constructions: The patterning of English first-person singular subject expression in conversation. *Journal of Pragmatics* 63. 19–34.
- Torres Cacoullós, Rena & Catherine E. Travis. 2018. *Bilingualism in the community: Code-switching and grammars in contact*. Cambridge: Cambridge University Press.
- Torres Cacoullós, Rena & Catherine E. Travis. 2019. Variationist typology: Shared probabilistic constraints across (non-)null subject languages. *Linguistics* 57(3). 653–692.
- Travis, Catherine E. 2007. Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change* 19(2). 101–135. DOI: 10.1017/S0954394507070081.
- Travis, Catherine E. & Rena Torres Cacoullós. 2012. What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics* 23(4). 711–748. DOI: 10.1515/cog-2012-0022.
- Travis, Catherine E. & Rena Torres Cacoullós. 2021. Categories and frequency: Cognition verbs in Spanish subject expression. *Languages* 6(3). DOI: 10.3390/languages6030126.
- Travis, Catherine E., Rena Torres Cacoullós & Evan Kidd. 2017. Cross-language priming: A view from bilingual speech. *Bilingualism: Language and Cognition* 20(2). 283–298.
- Vázquez Rozas, Victoria. 2006. Gustar-type verbs. In J. Clancy Clements & Jiyoung Yoon (eds.), *Functional approaches to Spanish syntax: Lexical semantics, discourse and transitivity*, 80–114. New York: Palgrave Macmillan.
- Vázquez Rozas, Victoria. 2012. Construyendo emociones: Sintaxis, frecuencia y función comunicativa. In Tomás Jiménez Juliá, Belén López Meirama, Victoria Vázquez Rozas & Alexandre Veiga Rodríguez (eds.), *Cum corde et in nova grammatica: Estudios ofrecidos a Guillermo Rojo*, 841–854. Universidade de Santiago de Compostela: Servizo de Publicacións.

Chapter 4

Variable subject expression in second language acquisition: The role of perseveration

Kimberly L. Geeslin^a, Thomas Goebel-Mahrle^a, Jingyi Guo^a & Bret Linford^b

^aIndiana University ^bGrand Valley State University

The current study examines the development of variable subject expression in Spanish across multiple proficiency levels of second language learners, and compares their patterns with a group of native speakers from the same speech community. A cross-sectional design and a written contextualized preference task are employed to explore the differences in rates of subject form selection, the degree to which the linguistic and psychological construct known as perseveration, constrains the acquisition of subject expression, as well as the potential interaction between perseveration and other linguistic factors. Our analysis examines null and overt pronominal subjects as well as full lexical noun phrase verbal subjects. The results show that as proficiency level increases, learners' selection rates of subject forms and rates of perseveration become gradually more native-like, and an increasing number of linguistic factors (prime form, gender continuity, tense mood aspect continuity) predict the occurrence of perseveration. In addition, for learners from the two most advanced levels and native speakers, feminine primes are more likely to perseverate than masculine primes, suggesting the effects of psychological processes (i.e., surprisal) on perseveration.

1 Introduction

The current study examines the patterns of the perseveration of variable subject forms in Spanish across multiple proficiency levels of second-language (L2)



learners, and for a group of native speakers (NS) from the same speech community. In a general sense, perseveration refers to the tendency for a particular form to appear again (i.e., to persevere) in subsequent discourse. We note that this is sometimes referred to as linguistic priming, but priming may also refer to the cognitive explanation for the effect known as perseveration (see Otheguy 2015). Thus, we will employ the term perseveration throughout, and we will take this to describe a distribution of subject forms attested in our dataset, such that a given form (i.e., a prime) is followed by a second subject of the same type (e.g., an overt subject pronoun is followed by a subsequent overt subject pronoun). What is particularly interesting about perseveration is that it may occur even in the absence of a discourse-based or functional explanation. For example, Poplack (1980), studying Puerto Rican Spanish which tends to delete word-final -s, found that plural marking was more likely to occur if a previous element in an NP was already -s marked for plural (e.g., *la[∅] mujere[∅] bonita[∅]* vs. *las mujeres bonitas* ‘the beautiful women’). This marking cannot be accounted for by appealing to discourse constraints or a disambiguation strategy; rather, overt plural marking with -s is favored when a previous element is already marked, even though a functional hypothesis would consider this marking redundant. In short, the phenomenon can be described as the distribution that results when one form begets another subsequent like form, even when this is unnecessary for reasons such as disambiguation. We will return to a more detailed discussion of this distribution later.

Our selection of subject forms as the object of study is based in large part on the wealth of research available on their variable use. Collectively, sociolinguistic studies show that variation between null subjects and overt subject pronouns (SPs) is constrained by factors such as person/number, verbal tense, mood, and aspect (TMA), reflexivity of the verb, lexical content of the verb, and specificity of the referent as well as by discourse-related factors beyond the verb phrase, such as perseveration, discourse genre, referent cohesiveness and clause type (e.g., Carvalho et al. 2015, Otheguy et al. 2007, Shin & Otheguy 2009, Silva-Corvalán 1994, Travis & Torres Cacoullos 2012). Although there is a tendency to limit the study of subject forms to null and overt SPs, there has recently been expansion to study the patterns that influence the realization of subject forms as full lexical noun phrases (NPs), even in contexts where they have been mentioned previously. These studies have demonstrated that many of the same independent linguistic constraints influence use of these forms relative to null and overt pronouns (see Bentivoglio 1993, Dumont 2006, Gudmestad & Geeslin 2022, Gudmestad et al. 2013, Silva-Corvalán 2015). In general, we see that rates of use of null forms vary dialectally, but the constraints on these rates of use are often

steady across studies and speech communities (Carvalho et al. 2015, Gudmestad & Geeslin 2022). Similarly, there are ample studies of subject form expression in L2 Spanish, examining variability in learner-directed input as compared to patterns of second language development (Gurzynski-Weiss et al. 2018), the subtle differences attested between highly-advanced non-native speakers (NNSs) and NSs of Spanish (Geeslin & Gudmestad 2011, 2016), stages of development across multiple levels of proficiency (Geeslin et al. 2015), and differences between group and individual patterns for variable structures (Geeslin et al. 2013), to name only a few key issues. As with the studies of NS patterns of use, we find that the constraints on the expression of subject forms are relatively stable. This makes subject form expression a particularly good test case for a variety of theoretical questions.

The current study adopts a variationist framework, which is characterized by its attention to the many factors that simultaneously influence patterns of use. An advantage of this approach is that it allows researchers to determine the constraints that influence the realization of specific (socio)linguistic variables in the interlanguage, how these constraints develop over time, and if learners approximate native-like usage. For NSs of Spanish, the wealth of research available has shown the multiple independent linguistic factors, as well as social factors such as regional location of a speech community come into play in studying the patterns of subject form use. Likewise, there now exists a significant body of L2 research showing that sensitivity to the factors that constrain NS variation can be acquired by NNSs of the language (for overviews see Geeslin & Long 2014, Kanwit 2018). The native-like variation between two or more grammatical forms that perform the same function (i.e., variable structures) is guided probabilistically by the semantic, morphosyntactic, and discourse-level features of the linguistic context as well as the social features of the extra-linguistic context and is often referred to as Type II variation (Bayley & Preston 1996, Mougeon & Dewaele 2004, Young 1991). In Spanish, research has examined the SLA of the variation between forms used to express copulas, mood contrasts, the progressive aspect, future and past-time marking, as well as grammatical subjects and objects (see Geeslin 2018 for a review). Together these studies demonstrate that the variationist perspective can be applied profitably to the study of second language development.¹

¹For a generative approach to the study of L2 subject expression, see Lozano (2002), who utilizes experimental methodology to investigate the acquisition of universal properties of subject expression versus Spanish language-specific parameters. Additionally, Lozano (2016) uses corpus data to argue that advanced learners are pragmatically redundant in their subject expression and may struggle with the syntax-discourse interface.

In this brief introduction, we have established that subject form expression is a well-studied structure in both first and second language contexts and, that it lends itself to the study of perseveration. We close this introduction by highlighting the contributions of the current investigation to the larger whole. While there are studies that analyze the relative influence of many morphosyntactic factors on subject form expression, there is still a need for careful examinations of the more complex patterns that are exhibited in extended discourse. For example, we know that the form of the previous mention of the referent predicts rates of production of subject forms for highly advanced NNSs (e.g., Geeslin & Gudmestad 2011). Moreover, this particular factor tends to be one that is controlled in elicitation tasks designed to study L2 development, rather than the focus of the analysis (e.g., Geeslin et al. 2015). To date, less is known about the developmental path that L2 learners follow in their acquisition of sensitivity to these more complex factors and whether this sensitivity might lead to similar patterns of perseveration in learner language. Finally, within the variationist framework, studies of subject expression by L2 learners have been based primarily on oral production data. Thus, the current study contributes to the growing body of research on the acquisition of variable structures by using a cross-sectional design to examine the L2 development of subject expression and sensitivity to the form of the previous mention of the referent. We accomplish this through an analysis of subject form selection on a controlled preference task, which allows us to ensure that each participant responds to the same confluence of independent variables. In so doing, we also continue the cross-disciplinary dialogue between research on language variation and second language acquisition and provide common ground to move both fields forward.

2 Spanish subject expression

The syntax of Spanish allows for grammatical subjects to be expressed overtly as a personal pronoun (overt SP; example 1), a lexical noun phrase (lexical NP; example 2), demonstrative pronoun (example 3), indefinite pronoun (example 4), interrogative pronoun (example 5), as well as allowing phonetically unexpressed, or null subjects (example 6).

- (1) Él habla español.
'He speaks Spanish.'
- (2) Juan habla español.
'Juan speaks Spanish.'

- (3) Ese habla español.
'That one speaks Spanish.'
- (4) Alguien habla español.
'Someone speaks Spanish.'
- (5) ¿Quién habla español?
'Who speaks Spanish?'
- (6) ∅ Habla español.
'(He/she-null) speaks Spanish.'

With few exceptions, variationist research on subject expression in Spanish has focused on the variation between null and overt SPs and the analyses are limited to contexts that are determined to permit variation between the two forms (Otheguy & Zentella 2007). Nonetheless, there are studies that suggest that these two forms are in variation with others, most notably full lexical NPs which have been shown to occur even following an adjacent previous mention of the same subject (Bentivoglio 1993, Dumont 2006, Gudmestad & Geeslin 2022, Gudmestad et al. 2013, Silva-Corvalán 2015).² Previous variationist research on subject expression in Spanish has found that variation between null and overt SPs is constrained by morphosyntactic factors such as person/number, tense, mood, and aspect (TMA), lexical frequency and reflexivity of the verb, semantic factors such as lexical content of the verb and specificity of the referent, discourse-level factors such as switch reference³, referent cohesiveness, the form of the previous mention of the subject (i.e., perseveration), discourse genre, clause type, as well as some extra-linguistic factors (Ávila-Jiménez 1995, Bayley & Pease-Alvarez 1996, 1997, Bentivoglio 1987, Cameron 1994, 1995, Cameron & Flores-Ferrán 2004, Enríquez 1984, Erker & Guy 2012, Flores-Ferrán 2005, Hochberg 1986, Morales 1986, Otheguy et al. 2007, Otheguy & Zentella 2012, Shin 2006, 2012, Shin & Cairns 2009, Shin & Otheguy 2009, Silva-Corvalán 1994, Travis 2007, Torres Cacoullos

²While it may seem counterintuitive given the general rules of use for full lexical NPs, recent research has begun to provide two strong arguments for their consideration within the same envelope of variation as other subject forms. Firstly, their use can be constrained by similar factors to other subject forms (e.g., Gudmestad & Geeslin 2022, Dumont 2006) and secondly, the forms have been shown to appear in interview speech with some regularity in contexts where they would not be expected, such as those where they are mentioned previously and where there is no need to disambiguate from other referents.

³Also known as “continuity of reference” (Shin & Otheguy 2009), “coreferentiality” (Silva-Corvalán 1994), “subject continuity” (Torres Cacoullos & Travis 2010) and “discourse connection” (Carvalho & Child 2011).

& Travis 2010; *inter alia*). The discourse-level factors previously mentioned have been found to be crucial in explaining subject expression in Spanish and are the focus of the current study. Hence, the following review will be limited specifically to the impact of discourse-level factors on subject expression and how these findings have influenced the goals of the present investigation.

Perhaps the most widely studied discourse-level factor is switch reference (Bayley & Pease-Alvarez 1997, Bentivoglio 1987, Cameron 1994, 1995, Cameron & Flores-Ferrán 2004, Erker & Guy 2012, Otheguy et al. 2007, Silva-Corvalán 1994, Shin & Cairns 2009, Shin & Otheguy 2009, Torres Cacoullós & Travis 2010, Travis 2007). Contexts where the subject referent is different from the referent of the previous tensed verb are known as “switch reference” while situations where these two referents are the same are known as “same reference”. Results across studies based on oral speech of speakers from a variety of linguistic backgrounds demonstrate that overt SPs are more frequent in contexts of switch reference than same reference. Moreover, several studies find that this factor is one of the most important factors influencing subject form variation (see Cameron 1994: 28).

Some studies have expanded the analysis of switch reference (Bayley & Pease-Alvarez 1997, Cameron 1995, Travis 2007). For example, Bayley & Pease-Alvarez (1997) analyzed degrees of “discourse connectedness” in Mexican-descent children’s oral and written narratives. The categories of the discourse connectedness variable accounted for the continuity of TMA between tensed verbs in the discourse, the clause distance to the previous mention of the referent, the previous mentions of the referent in different syntactic functions, and changes in narratives. By combining these factors, Bayley & Pease-Alvarez (1997) identified five degrees of discourse connection: from the most connected discourse (where the referent and TMA of the verb were the same as the referent and TMA in the preceding tensed verb) to the least connected discourse (where the narrative section or discourse topic changed). Their findings showed that there was a steady increase in the probability of using overt SPs as the discourse became less connected. They also found that the effect of the discourse connectedness variable was a more robust predictor of subject use than switch reference alone. As we turn our attention to the focus of the current study, the role of perseveration in subject form acquisition and use, we will see that discourse connectedness must remain in view as well.

3 Research on perseveration

Interest in perseveration⁴, the focus of the current study, stems from research findings that showed that patterns of deletion could not be explained using discourse constraints alone. For example, Poplack (1980) analyzed the factors that constrain overt plural -s marking (in variation with -s aspiration and deletion) in Puerto Rican Spanish. From a corpus of naturalistic productions, 6439 tokens of words in plural NP strings were extracted (e.g., *las nenas bonitas* ‘the pretty girls’) and were coded for grammatical category, following phonological segment, following stress, presence of disambiguating plural information, position of the word in the NP string, and presence of preceding plural marking. Apart from following stress, the functional and discourse-related factors did not account for the realization of final -s. In fact, Poplack found that “[p]resence of a plural marker before the token favors marker retention on that token, whereas absence of a preceding marker favors deletion (...) (Additionally,) [t]he most favorable context for marker deletion is precisely when the two preceding markers have already been deleted” (pp. 63–64). Thus, her findings do not fit a functional explanation and, instead, are consistent with Torres Cacoullós & Travis (2010: 4), who state that “the use of a certain structure in one utterance functions as a prime on a subsequent utterance, such that that same structure is repeated”.

With respect to subject form perseveration, most studies distinguish between what Travis & Torres Cacoullós (2018) dub *co-referential subject priming*, which focuses on the previous expression of the same referent regardless of clausal distance, and *adjacent clause subject priming*, which analyzes the subject forms in adjacent clauses regardless of co-referentiality. Variations of these analyses were carried out by Torres Cacoullós & Travis (2010), Carvalho & Child (2011), Abreu (2012), Flores-Ferrán (2002), Geeslin & Gudmestad (2011) and Travis (2007).⁵ Additional studies focus solely on first- and third-person singular forms (de Prada Pérez 2020) and still others have examined perseveration’s role in the expression of second person singular pronouns (Callaghan & Travis 2021). The diverse coding schemes employed in these studies may well stem from the fact that persever-

⁴Also known as “parallelism” (e.g., Carvalho & Child 2011), “linguistic priming” and “structural priming” (Pickering & Ferreira 2008) among other terms.

⁵We recognize that there are additional nuances within the various coding schemes employed in these studies. For example, some studies examine only the form of the preceding co-referential subject (Carvalho & Child 2011, Torres Cacoullós & Travis 2010), others examine the form of the preceding subject, even when it is not co-referential (Cameron 1994, Cameron & Flores-Ferrán 2004), while others examine the form of the preceding mention of the referent in subject position regardless of the distance between mentions (Abreu 2012, Flores-Ferrán 2002, Geeslin & Gudmestad 2011, Travis 2007).

ation has been shown to happen on various levels of linguistic representation including syntactic, semantic, structural, and lexical (see Pickering & Ferreira 2008 for a review, Travis 2007). For example, while lexical repetition appears to boost the strength of perseveration, perseveration also appears to happen between linguistic structures in the absence of lexical repetition (Pickering & Ferreira 2008, Travis 2007) and even when speakers switch between languages (de Prada Pérez 2018, Gries & Kootstra 2017, Sodaci 2018, Torres Cacoullos & Travis 2010).

In general, this research has shown perseveration to occur through findings such as higher rates of overt SPs in contexts where they are preceded by overt SPs than when they are preceded by nulls and vice versa. In fact, the form of the previous mention of a referent has been found to be one of the strongest predictors of subject form variation in some studies (e.g., Carvalho & Child 2011, Torres Cacoullos & Travis 2010). A functional hypothesis, on the other hand, would predict that repeated marking of forms (such as overt subjects) would be unnecessary once the information was clearly established in accordance with Grice's (1975) maxim of quantity, which states that speakers should "not make a contribution more informative than is required" (p. 45). In contrast to the use of subject forms for functional reasons, perseveration is unique because it appears to happen involuntarily without any pragmatic or functional motivation (Cameron & Flores-Ferrán 2004). Thus, as Cameron & Flores-Ferrán (2004) describe, it can be considered the part of language that is expressive, that is, where the message may be more spontaneous and less carefully planned. There are accounts that address the relative importance of perseveration vis-a-vis function, shifting the predictive importance toward one or the other (e.g., see Otheguy 2015 for discussion of the importance of function). It is our view that allowing for a role for functional factors does not diminish the apparent importance of perseveration. Instead, the current study recognizes a role for each.

4 Subject expression in L2 Spanish

Research on subject expression in L2 Spanish began with a focus on the L2 acquisition of the null subject parameter and its associated properties (Al-Kasey & Pérez-Leroux 1998, Bini 1993, Emberson 1987, Galvan 1999, Isabelli 2004, Licerias 1989, Licerias et al. 1997, Phinney 1987, White 1985). Some work, informed by generative and optimality theory, explored discourse-pragmatic features as well (LaFond 2002, LaFond et al. 2000, Montrul & Rodríguez Louro 2006, Rothman 2007), although studies that investigated the role of pragmatics on subject expression generally focused on the acquisition of subject expression in *obligatory*

rather than *variable* contexts (Blackwell & Quesada 2012, Quesada & Blackwell 2009, Rothman 2009). Both generative and discourse-pragmatic approaches to this issue continue to be of interest today (e.g., Lozano 2002, 2016).

In contrast to the aforementioned approaches, the variationist method allows us to measure and track patterns of use over the course of development without relying on an assessment of accuracy of a single form in a given context. This is especially helpful for charting acquisition in contexts where more than one subject form is allowable. Existing variationist research on subject expression in L2 Spanish has sought to determine the various linguistic and extra-linguistic factors that guide the use of subject forms. Through a series of studies, Geeslin & Gudmestad (2008, 2010, 2011) and Gudmestad & Geeslin (2010) showed that advanced L2 learners appear to reach a native-like sensitivity to the predictors of subject form variation in sociolinguistic interviews. They analyzed all forms produced in the subject position, including null, overt SPs and lexical NPs. Like Bayley & Pease-Alvarez 1997, Geeslin & Gudmestad (2011) examined referent cohesiveness and found that NSs and NNSs used more overt subjects as distance between mentions of the referent increased or functions of the referent changed. Relevant to the current study, they found that perseveration occurred for these speakers given that “null subjects were followed by a higher frequency of null subjects and overt forms were followed by a higher frequency of overt forms” (Geeslin & Gudmestad 2011: 10).

Continuing this line of work, Linford & Geeslin (2022) studied the L2 acquisition of sensitivity to referent cohesiveness on variable subject expression in Spanish. For their study, 125 NNSs across five levels of proficiency (beginner to highly-advanced) and a group of 25 NSs completed a written contextualized preference task (WCPT) in which aspects of referent cohesiveness were manipulated. Specifically, the distance to the previous mention of the referent, the syntactic function of the previous mention of the referent and the TMA of the verb with the previous mention of the referent were manipulated. For the WCPT, participants selected either a null subject, an SP or a lexical NP to complete phrases that were embedded into a written dialogue. Each item was categorized into one of eight referent cohesiveness categories, from most to least cohesive based on the manipulated factors. The results showed that native-like rates of selection of the three forms did not occur until the highest level of proficiency, suggesting that acquiring the rates of variation across subject forms occurs rather late in the acquisition process. In addition, unlike previous research on oral production data that found a consistent decrease in the use of null subjects as the discourse became less cohesive, they did not find this to be a case between all categories for any group – even the native speakers – including after they reorganized the

categories based on the varying degrees of importance of the sub-factors. Hence, even in this highly controlled and structured task, we see that referent cohesiveness and its associated properties, despite their clear importance, cannot be taken as the lone explanatory factor.

In connecting these findings across studies, we hypothesize that better understanding the role of the form of the previous mention of subject referents and the resulting perseveration is a key step in understanding the limitations of previous findings. In addition, other associated properties of referent cohesiveness not explored yet, such as referent gender, might be related to perseveration as well. In fact, there is complementary research on the acquisition of gender as well as research on the psychological notion of *surprisal*, which suggests that the gender of the referent and the cohesiveness between referents in terms of gender may further play a role in understanding perseveration. For example, Malovrh (2014) found that even the most advanced learners performed less accurately on a written and oral short film retell when producing feminine clitics (i.e., *la[s]*) versus masculine clitics (i.e., *lo[s]*). He further posited that “masculine forms are used as defaults under conditions in which access to working memory is more restricted” (p. 66), such as experimental tasks. Earlier studies such as Klee (1989) also found that, with respect to object clitics, learners tend to acquire feminine clitics last, and default to *lo* as an archmorpheme in all accusative contexts. Relatedly, studies on noun and adjective agreement such as Alarcón (2010) have found that learners are typically more accurate at producing gender agreement between masculine nouns and adjectives and are often guided by semantic notions such as animacy. Taken together, these disparities in learner performance between masculine and feminine referents suggest that masculine and feminine gender are activated differently in psychological representation. It is possible that, if masculine gender acts as a default, perseveration may obtain less in cases where the prime gender is masculine. Conversely, overt feminine referents may prime a preference for overt feminine referents. This relates to the notion of *surprisal*, as described by Jaeger & Snider (2007). These authors analyzed the English ditransitive construction which varies between the more frequent double NP construction (e.g., *I gave him the book*) and the less frequent NP PP construction (e.g., *I gave the book to him*). The authors found that the less frequent construction was more likely to result in perseveration on the following ditransitive structure. They interpret this result as a product of the *surprisal* caused by the less frequent construction. According to this surprisal-sensitive persistence hypothesis (Jaeger & Snider 2007: 3), “less expected prime structures are predicted to prime more (i.e., to lead to a bigger increase in the probability of repetition) than more expected prime structures”. In other words, less frequent variants exert a stronger priming effect due

to their salience in the discourse. If we assume that masculine is the unmarked gender (and indeed, it has been argued that feminine is the marked gender in Spanish, see Beatty-Martínez & Dussias 2019, Harris 1991), then it is possible for a feminine form to more strongly activate an underlying feminine representation, leading to higher rates of perseveration.

5 The current study

The current project was designed to examine the development of subject expression in L2 Spanish, looking specifically at perseveration in referential third person singular contexts⁶ in order to deepen our understanding of the relationship between perseveration of referential pronouns and other discourse-related factors, such as referent cohesion. To meet these goals, the current study answers the following questions:

1. What is the overall frequency of subject form selection by native speakers and L2 learners across different levels of proficiency on a written contextualized preference task?
2. Do L2 learners across different levels of proficiency and native speakers perseverate subject forms on a written contextualized preference task?
3. If perseveration is attested for a speaker group, is it constrained by independent factors such as the prime form (null, overt or lexical NP), and factors related to referent cohesiveness, such as, switch reference, TMA continuity, gender continuity, and/or gender of the referent?

5.1 Participants

The participants were 125 L2 learners of Spanish and 25 NSs. All L2 learners were native speakers of English and ranged in age from 18 to 47 years (average = 21.9 yrs.). There were 70 female and 55 male L2 learners. The L2 learners were divided into five groups of Spanish proficiency (split into five percentile ranges) based on their scores on a 24-item grammar proficiency test (see §5.2 for additional information about the grammar test). Table 1 summarizes these participant characteristics.

⁶Here, we describe third person pronouns as “referential” since they refer to persons who are not actively participating in the discourse between interlocutors, as opposed to first and second person pronouns which are deictic in nature (see de Prada Pérez 2020 for further discussion of the referential nature of third person pronouns).

Table 1: Participant characteristics by year of university Spanish course enrollment and mean score in a grammar proficiency test ("G": Graduate).

Level	N	Year					Test mean (%)
		1	2	3	4	G	
1	25	19	6				30.8
2	18	6	11	1			41.7
3	22		8	13	1		51.3
4	31			11	20		68.0
5	29				4	25	89.9
NS ^a	25			n/a			97.3

^aTwo native speaker participants did not complete the grammar test.

The results of a One-Way ANOVA revealed that differences in grammar test scores across participant groups were significant [$F(5, 142) = 280.6, p < 0.0001$], and Games-Howell post hoc tests⁷ revealed significant differences between all participant groups.

The NS group included 19 females and 6 males. Their countries of origin were Argentina (1), Bolivia (1), Colombia (3), Costa Rica (1), Mexico (6), Nicaragua (1), Peru (3), Puerto Rico (1) and Spain (8). Their ages ranged from 22 to 44 (mean = 31.24 years). All were university-educated instructors of Spanish residing in the U.S. at the time of data collection. In addition, all were bilingual in at least Spanish and English. Similar to Geeslin & Gudmestad (2008), this group of native speakers was chosen precisely because it is these speakers with whom the L2 learners in our study interact and as such constitute a reasonable target for acquisition.

5.2 Elicitation tasks

All participants completed three tasks in the following order: a written contextualized preference task (WCPT), a grammar proficiency test, and a background questionnaire. The tasks were administered to the first four groups of L2 participants either on paper or online by means of *Quia Web*⁸ during the participants' regularly scheduled class time. The graduate-level L2 learners and the native speakers completed the tasks online at a location of their choice.

⁷Games-Howell post hoc tests were employed since the test of homogeneity of variances was significant.

⁸<http://www.quia.com>

Participants first completed the WCPT, which consisted of 20 items embedded within a fictional dialogue in Spanish. Each item had three response choices, which were identical except for the subject forms: null, overt SP, or lexical NP.⁹ Participants were instructed to read the dialogue and select the phrase with the form that sounded most natural in each context. The preceding mention of the referent was varied throughout the WCT in order to provide a means for examining perseveration. Additionally, the instrument presented contexts with same and switch reference, same and switch TMA, and also varied the gender of the referent, making it possible to study the way these factors conspire to influence perseveration.

Various linguistic features of the context were controlled in the instrument to avoid potential confounding factors. All referents in each item were animate third person singular referents, and all clauses were independent clauses. Furthermore, the verb forms in each item were divided evenly between the simple present indicative and the imperfect indicative, allowing for a balance between clearly defined and potentially ambiguous verb forms. Although at first glance this task may resemble those used under other theoretical approaches (e.g., generative), it falls within variationist framing because of (1) its attention to the many independent linguistic factors that come to bear on form selection and (2) the response format that allows participants to indicate a preference without reference to accuracy or acceptability. While not always the case, another defining feature is that it creates extended narrative context rather than eliciting sentence-level judgements. Example 7 is an excerpt taken from the written contextualized task followed by a translation into English:

- (7) **Jorge:** *¿De verdad? Pues ya nunca llego tarde porque cada vez que yo llegaba tarde, ella siempre se ponía muy seria.*
- a. *Decía que le daba igual...*
 - b. *Ella decía que le daba igual...*
 - c. *Juanita decía que le daba igual...*

Jorge: Really? Well, now I never arrive late because every time I arrived late, she always got really serious.

- a. \emptyset Said that it didn't matter...
- b. She said that it didn't matter...
- c. Juanita said that it didn't matter...

⁹The reason for limiting the options to these three forms was that Geeslin & Gudmestad (2008) found that the other forms (e.g., demonstrative pronouns, indefinite pronouns, and interrogative pronouns) represented a very small portion of the data (only 4.6% for highly-advanced L2 learners and 8.4% for native speakers).

The second task, a grammar test, consisted of a fictional narration in Spanish that contained 25 contextualized items in which the participants were instructed to choose between three possible options to complete the sentences grammatically.¹⁰ In the current study, we excluded from the analysis one item due to high levels of variability among native speakers,¹¹ specifically, the item that included Spanish copulas and is marked with preterit or imperfect aspect. As both aspectual marking and copula contrast in Spanish tend to vary, the variability on this item is not unusual, leaving a total of 24 items.

The background questionnaire for the L2 learners consisted of 33 questions in English that gathered demographic information (e.g., age, gender, etc.) as well as determined the participants' current and previous experience with Spanish and other languages. Another questionnaire was created for the NSs which contained 10 items to gather information regarding demographics, time spent in the U.S., and experience with other languages. The data collected through these tasks yielded the description of the participants provided in §5.1.

5.3 Coding and analysis

In this study, we examined contexts in which the previous mention of the referent was in subject position regardless of the distance to the previous mention, thereby following the operationalizations of Abreu (2012), Flores-Ferrán (2002), Geeslin & Gudmestad (2011), and Travis (2007). This operationalization allowed us to examine the potential relationship to perseveration of continuity of reference and/or gender of the referent. However, three of the total twenty contexts were excluded from this analysis because there is no previous mention of the referent in subject position or this mention falls in a previous item, where the subject depends on the participant response to the previous item. In addition, one context was excluded from the analysis because the potential previous mention of the same referent has an ambiguous reference. These exclusions left us with 16 items for analysis and a total of 2,396 tokens.

Our dependent variable in the current analysis is whether perseveration occurred, that is, whether the form selected by the participant was the same or different from the previous form of the referent. Additionally, we coded our data for

¹⁰As mentioned earlier, only 8 of the native speaker participants scored 100% on the proficiency test. Even so, it is important to note that the results on the same grammatical proficiency test of over 500 native and L2 learners of Spanish were submitted to a reliability test using SPSS. The Cronbach's Alpha for the proficiency test was 0.868, well above the minimum requirement of 0.70 for a test to be reliable (George & Mallery 2012).

¹¹The item that was removed asked test-takers to select between options that contained the Spanish copulas.

4 Variable subject expression in second language acquisition

several independent variables. It will be recalled that previous research indicates that referent cohesiveness, which has been operationalized with varying degrees of distinction, sometimes based on TMA continuity, position of the referent, or other factors, plays an important role in understanding patterns of subject form use. To this end, we explored multiple factors related to referent cohesiveness. From studies such as Geeslin & Gudmestad (2011) and Linford & Geeslin (2022), we know that TMA continuity (which distinguishes adjacent tensed verbs that contrast in TMA from those that do not) adds dimension to our understanding of the relationship between referent cohesiveness and perseveration and, thus, we included this factor in our coding. We also examined the variable *gender continuity*, which captures whether a referent has the same gender as the referent of the subject of the previous third person singular verb. As noted earlier, this factor is particularly relevant for third person subjects because third person pronouns are referential. For this factor, we disregarded intervening referents that were not third person singular because we assume that gender continuity is most likely a relevant influence for perseveration of third person singular subject forms. In cases of same reference, there is by definition, also gender continuity and, thus, the key contrast occurs in contexts of switch reference. This interaction is reflected in our coding scheme (details below).¹² This variable sheds light on the relationship between perseveration and the function that forms, such as overt SPs or lexical NPs, play in distinguishing the current referent from a previous one.

In addition to gender continuity, we also coded for the gender (masculine vs. feminine) of the prime.¹³ Our reasons for including this variable stem from the research reviewed previously on the acquisition of various other L2 Spanish structures that reflect learners' differing patterns of acquisition with feminine referents (e.g., Alarcón 2010, Klee 1989, Malovrh 2014) as well as the possible role that a non-default form may play in processes such as perseveration (Jaeger & Snider 2007). We summarize our coding of these independent variables in Tables 2–6.

¹²There is only one item that has an intervening referent that is not third person singular. This item was coded as switch reference for the “same/switch reference” factor, since the intervening verbal subject is first person singular. The same item was coded as same reference, same gender for the “gender continuity” factor, since the closest preceding third person singular verbal subject has the same reference as the subject of this item.

¹³There is only one item where the prime gender and the current referent gender do not correspond. In this item, the current referent is singular feminine [*∅/Ella/Olivia no siente nada por él* ‘null/She/Olivia does not feel anything for him’], while the prime is a plural masculine referent null subject *ellos* ‘they’ which includes the current referent.

Table 2: Analysis of independent variable: Prime form (The form of the preceding mention of the referent in subject position, regardless of the distance)

Categories	Example (prompt with response options)
Null	<p><i>El único problema es que hace cinco meses que ∅ rompió con su exnovio tras una relación de dos años y todavía está un poco triste.</i></p> <p>‘The only problem is that it has been five months since she (∅) broke up with her ex-boyfriend after a two-year relationship and she is still a little sad.’</p> <p>∅/Ella/Juanita salía con Paco García.</p> <p>‘∅/She/Juanita was dating Paco García.’</p>
Overt pronoun	<p><i>Sé que ELLA no siente nada por él ahora...Antes...</i></p> <p>‘I know that SHE doesn’t feel anything for him now...Before...’</p> <p>∅/Ella/Olivia sentía algo por él, ¿no?</p> <p>‘∅/She/Olivia felt something for him, right?’</p>
Lexical NP	<p><i>Sí, sí. Me voy. JUANITA ya me está esperando.</i></p> <p>‘Yes, yes. I’m leaving. JUANITA is already waiting for me.’</p> <p>∅/Ella/Juanita se irrita un poco cuando llego tarde</p> <p>‘∅/She/Juanita becomes a little irritated when I arrive late.’</p>

The initial step in our analysis was to examine the distribution of the subject forms selected by each participant group. Although this is not the dependent variable for the remaining analyses, it is important to provide this distribution as a backdrop for comparison to other studies. Following the reporting of the distribution of the forms selected on our WCPT, we provide a similar report of the distribution of the perseveration attested by each participant group. Although we do provide the overall rates of perseveration by group, we note that it is the rate of perseveration within the prime form that is more meaningful in answering our research questions and contextualizing our findings within the existing research on this subject. Our analysis then turns to a statistical examination of the degree to which the independent factors in our analysis are related to perseveration for each group. We present the findings of a binary logistic regression analysis¹⁴ for each participant group as a means for answering our third research

¹⁴We used the Generalized Estimating Equations tool in SPSS 27 with participant as an exchangeable subject variable for the regressions.

4 Variable subject expression in second language acquisition

Table 3: Analysis of independent variable: Same vs. switch reference (Whether the subject referent of the immediately preceding tensed verb is the same)

Categories	Example (prompt with response options)
Same reference	<p>Pablo: <i>Pues, en primer lugar, Olivia no tiene exnovios.</i> ‘Pablo: Well, first, Olivia doesn’t have ex-boyfriends.’ Jorge: \emptyset/Ella/Olivia <i>salía con Enrique el año pasado, ¿no?</i> ‘Jorge: \emptyset/She/Olivia dated Enrique last year, right?’</p>
Switch reference	<p><i>Ok, ok...la verdad es que tengo una cita con mi novia.</i> ‘Ok, ok...the truth is that I have a date with my girlfriend.’ \emptyset/Ella/Mi novia <i>quiere comer en un restaurante elegante así que...</i> ‘\emptyset/She/My girlfriend wants to eat at an elegant restaurant so ...’</p>

Table 4: Analysis of independent variable: TMA continuity (Is TMA of previous mention of the referent in subject position the same?)

Categories	Example (prompt with response options)
Same TMA	<p><i>Tal vez, pero también él le dice a Juanita que no pasa nada si ella sale con otro chico. Si...</i> ‘Possibly, but he also tells Juanita that it’s ok if she dates another guy. If...’ \emptyset/él/Paco <i>tiene problema conmigo...</i> ‘\emptyset/he/Paco has a problem with me...’</p>
Different TMA	<p><i>Antes Paco me trataba como a un amigo pero ahora que Juanita sale conmigo,...</i> ‘Before Paco treated me as a friend but now that Juanita dates me,...’ \emptyset/él/Paco <i>ni me mira.</i> ‘\emptyset/he/Paco doesn’t even look at me.’</p>

Table 5: Analysis of independent variable: Gender continuity (Is the gender of the 3rd person subject referent of the immediately preceding tensed verb the same?)

Categories	Example (prompt with response options)
Same gender, same referents	Pablo: <i>Pues, en primer lugar, OLIVIA no tiene exnovios.</i> ‘Pablo: Well, first, OLIVIA doesn’t have ex-boyfriends.’ Jorge: <i>∅/Ella/Olivia salía con Enrique el año pasado, ¿no?</i> ‘Jorge: ∅/She/Olivia dated Enrique last year, right?’
Same gender, different referents	Jorge: <i>¿Seguro que no la conoces? Pues, es la chica con quien hablaba Ana Ramírez después de la clase de biología todos los días el semestre pasado. De hecho, ANA es su mejor amiga.</i> ‘Jorge: Are you sure you don’t know her? Well, she is the girl with whom Ana Ramírez chatted after the biology class everyday las semester. In fact, ANA is her best friend.’ Pablo: <i>Ah, ok...sí...</i> ‘Pablo: Ah, ok...yes...’ <i>∅/Ella/Juanita es muy guapa entonces.</i> ‘∅/She/Juanita is very beautiful then.’
Different gender, different referents	<i>Antes Paco me trataba como a un amigo pero ahora que JUANITA sale conmigo,...</i> ‘Before Paco treated me as a friend but now that JUANITA dates me,...’ <i>∅/él/Paco ni me mira.</i> ‘∅/he/Paco doesn’t even look at me.’

4 Variable subject expression in second language acquisition

Table 6: Analysis of independent variable: Prime gender

Masculine	<p><i>Tal vez, pero también él le dice a Juanita que no pasa nada si ella sale con otro chico. Si...</i></p> <p>‘Possibly, but he also tells Juanita that it’s ok if she dates another guy. If...’</p> <p>Ø/él/Paco tiene problema conmigo...</p> <p>‘Ø/he/Paco has problem with me...’</p>
Feminine	<p><i>¿De verdad? Pues ya nunca llego tarde porque cada vez que yo llegaba tarde, ELLA siempre se ponía muy seria.</i></p> <p>‘Really? Well now I never arrive late because every time I arrived late, SHE always got very serious.’</p> <p>Ø/Ella/Juanita decía que le daba igual.</p> <p>‘Ø/She/Juanita said that she didn’t care.’</p>

question. The independent variables included in the regression are prime form, TMA continuity, and gender continuity. We included prime form in order to examine how this variable affects perseveration when other factors are considered in the same statistical model. Gender continuity rather than same vs. switch reference was included since the former further specifies the degrees of discourse cohesion captured in the switch reference variable and it allows us to tease apart the effects of selecting each form for functional reasons (e.g., contrast/clarity) and psychological or expressive reasons (e.g., priming). Specifically, one would expect based on functional use of subject forms that null subjects would be perseverated most often in same reference contexts, overt SPs would be perseverated most often in switch reference contexts where the gender is contrastive with the previous subject referent, and lexical NPs would be perseverated most often in switch reference contexts, especially without contrastive gender of the referents. The degree to which these functional predictions (do not) account for the patterns attested, indicates a role for psychological effects, such as perseveration. We did not include the prime gender variable in the regression because there were unintended correlations between the prime gender and switch reference in the instrument design. The final step of our analysis is to focus more directly on the prime gender and its relationship to perseveration. By looking at the prime gender only in contexts of switch reference, we are able to test the hypotheses put forth earlier regarding the degree to which perseveration is differential for default vs. non-default forms.

6 Results

As described previously, we begin our presentation of the results with an overview of the distribution of the forms selected according to proficiency level. Table 7 shows the overall distribution of the subject forms selected by each participant group and the percentage each form selected constitutes within each participant group. The note below the table provides details of ANOVA tests of differences between groups for each subject form.

Table 7: Distribution of forms selected by group. Note: ANOVA tests between groups *** = $p < 0.001$, Nulls [$F(5, 144) = 36.13, p < 0.001$], Overt SPs [$F(5, 144) = 8.817, p < 0.001$], Lexical NPs [$F(5, 144) = 18.433, p < 0.001$].

Level	***		***		***		Total <i>N</i>
	Null subjects		Overt SPs		Lexical NPs		
	#	%	#	%	#	%	
1	133	33.5	151	38.0	113	28.5	397
2	99	34.5	106	36.9	82	28.6	287
3	127	36.1	155	44.0	70	19.9	352
4	249	50.2	179	36.1	68	13.7	496
5	334	72.0	90	19.4	40	8.6	464
NS	307	76.8	63	15.8	30	7.5	400

Table 7 demonstrates that the L2 learners select a relatively large proportion of overt SPs and lexical NPs at lower proficiency levels and gradually select a higher percentage of null subjects as L2 proficiency increases. This trend is especially noticeable at level 4 where the selection rates of overt SPs and lexical NPs decrease, accompanied by a sharp increase in the selection of null subjects. Results from three one-way ANOVAs comparing the selection rates of each of the forms between groups were significant (see note below Table 7). Post-hoc Tukey tests showed that for null subjects, the selection rates for learners in levels 1-3 were not significantly different from each other, level 4 learners' selection rate of null subjects was significantly different from all other levels, and level 5 and NSs were not significantly different from each other. For overt SPs, post hoc Tukey tests showed that the selection rates from level 1 to 4 were not significantly different from one another and level 5 and NSs were not significantly different from each other. Finally, post hoc Games-Howell tests showed that for lexical NPs, levels

4 Variable subject expression in second language acquisition

1–3 were not significantly different from each other, and level 4, 5 and the NSs were not significantly different from one another. In sum, we see that these apparent shifts, at level 4 for null subjects and lexical NPs, and at level 5 for overt SPs represent significant shifts in rates of selection.

We now turn our analysis toward the overarching question of whether we find perseveration, for all groups and for all forms, and what other independent variables help us understand the patterns of perseveration attested in our dataset. Table 8 shows the rates of perseveration within participant groups for each form and the overall rates of perseveration. For example, for level 3 learners, when the prime form was null, the form selected was also null in 34 tokens, which represent 30.9 percent of the cases; and in total, they selected the same form as the prime with 90 tokens (25.6 percent of all the cases). The use of the asterisks (*) indicates the results of chi-square tests that measured the degree to which these patterns were significantly different by form, within the group. Building on the previous example, this means that for level 3, the *p*-value for a test measuring the degree to which rates of perseveration differed by form was smaller than 0.001.

Table 8: Number and percentage of perseveration by prime form and overall perseveration (OP). Note: χ^2 tests, *** = $p < 0.001$.

Level	Null		Overt SP		Lexical NP		OP	
	#	%	#	%	#	%	#	%
1	27	21.8	39	31.5	36	24.2	102	25.7
2	27	30.0	28	31.5	22	20.4	77	26.8
3***	34	30.9	41	37.3	15	11.4	90	25.6
4***	78	50.3	50	32.3	13	7.0	141	28.4
5***	102	70.3	22	15.2	6	3.4	130	28.0
NS***	93	74.4	22	17.6	2	1.3	117	29.3

Table 8 demonstrates that starting at level 3 and continuing for each of the more advanced groups, and the NSs, the perseveration rates differ significantly across the three forms examined. This result is highly anticipated as we would not expect similar rates of perseveration for null subjects and lexical NPs in natural discourse. On the contrary, we might expect that because the perseveration of lexical NPs is functionally unnecessary for content recovery, it would occur at lower rates. A more interesting question is whether the rates of perseveration for a given form differ by participant group, as this would demonstrate the path of acquisition, and whether our learners arrive at native-like pat-

terns of perseveration. To assess this, we compared perseveration rates between groups by means of One-Way ANOVAs for each prime form (i.e., three separate ANOVAs). The results of the ANOVAs show that there are significant differences in perseveration rates between groups for null primes [$F(5, 144) = 25.003$, $p < 0.001$], overt SP primes [$F(5, 144) = 5.187$, $p < 0.001$] and for lexical NP primes [$F(5, 144) = 10.519$, $p < 0.001$]. Post hoc Tukey tests revealed that for perseveration of null primes, levels 1–3 were not significantly different from each other, level 4 was significantly different from all other groups and level 5 and NSs were not significantly different from each other. For overt SPs, post hoc Games-Howell tests showed no significant differences in perseveration between levels 1–4; level 5 was significantly different from levels 1, 3 and 4, and NSs were only significantly different from level 3. For lexical NP primes, levels 1–3 were not significantly different from each other, levels 3–5 were not significantly different from each other, and levels 4–NSs were not significantly different from one another. As a whole, this indicates that level 5 was the only group that demonstrated consistent native-like patterns of perseveration across primes. Adding detail, for null primes it appears that there is a shift at level 4 that leads to more native-like patterns by level 5, whereas this happens somewhat sooner for lexical NPs, showing the transition between levels 3 and 4. Patterns for overt SPs are not as linear as for null subjects and lexical NPs and, thus, the patterns attested by the ANOVA are more complex, but show a general trend toward reduction in perseveration over time, with a dramatic shift between levels 4 and 5.

Thus far, our analysis demonstrates that perseveration does in fact vary by form and proficiency level. To understand these patterns, we conducted separate binary logistic regressions for each level to examine other factors that may influence perseveration (see §5.3 for independent variable details). Table 9 summarizes the results of each of these six statistical models.

As is demonstrated in Table 9, prime form is a significant predictor of perseveration for level 3 and above. In contrast, the factors related to referent cohesiveness, such as gender continuity (which includes a measure of switch reference) and TMA continuity do not seem to play a significant role in patterns of perseveration until level 5. Additionally, the role of gender continuity is apparent for highly advanced learners and for NSs. We did find one anomalous result in that TMA continuity is significant at level 2, but then not for other levels until level 5. With regard to the direction of these effects, for prime form, we find that for nearly all groups for which this variable was significant, perseveration was most common with null primes, followed by overt SP primes, then lexical NP primes. The only exception to this trend was Level 3 in which perseveration occurred more with overt SP than null primes. For gender continuity, for level 5

4 Variable subject expression in second language acquisition

Table 9: Results of binary logistic regressions for perseveration by level (significant results bolded)

Level	Prime form		Gender continuity		TMA continuity	
	Wald χ^2	<i>p</i>	Wald χ^2	<i>p</i>	Wald χ^2	<i>p</i>
1	2.399	0.301	0.738	0.691	0.942	0.332
2	3.907	0.142	0.733	0.693	7.335	0.007
3	32.731	<0.001	7.107	0.029	0.012	0.912
4	52.368	<0.001	2.221	0.329	1.642	0.200
5	59.282	<0.001	22.687	<0.001	8.655	0.003
NS	51.358	<0.001	11.163	0.004	1.900	0.168

and the NSs, perseveration occurred least often in same reference contexts and more often in switch reference contexts, with little apparent difference with and without switches in gender. Finally, for continuity of TMA, level 3 demonstrated more perseveration when there was a switch in TMA whereas level 5 showed the opposite trend. We will discuss this result further in the section that follows. To summarize the overall patterns in terms of development, we see that learners first perseverate at different rates by prime form and then, at much higher levels of development, begin to demonstrate patterns of perseveration that are sensitive to independent variables related to referent cohesiveness. Such factors are indicators of functional explanations for (lack of) perseveration and it is to be expected that the interplay between functional patterns and psychological ones, such as perseveration, requires advanced ability in a language. We will return to these results in the discussion section.

The final step in our analysis was to examine whether the prime gender plays a role in perseveration. This analysis focused on switch reference contexts for which the nearest previous mention of the referent was in subject position. This narrower scope was selected because we identified interactions with switch reference as well as the distance of the mention of the reference, on the one hand, and the prime gender, on the other. These interactions were artifacts of the instrument rather than indications of how these might operate in naturally-occurring language. For example, on the WCPT we found that items with male primes were found in significantly more contexts of switch reference (57.2%) than those with female primes (33.4%) [Pearson $\chi^2 = 135.391$, $df = 1$, $p < 0.001$]. We further noted that although our original analysis included previous referents in subject position, regardless of intervening mentions of the referent, that this might in-

fluence the role that gender of the prime played. Consequently, we further narrowed our analysis to those contexts where the previous mention of the referent was in subject position (e.g., there were no intervening mentions of the referent as verbal objects). Table 10 presents the results of Chi-square tests to determine if the prime gender significantly correlated with perseveration for each participant group in contexts of switch reference where the previous mention of the referent was in subject position.

Table 10: Perseveration by prime gender in switch reference contexts with previous mention of referent in subject position (significant results bolded)

Level	Pearson χ^2	<i>p</i>	M prime		F prime	
			<i>N</i>	%	<i>N</i>	%
1	0.350	0.554	20	26.7	11	22.0
2	0.854	0.355	9	17.0	9	25.0
3	0.287	0.582	18	27.3	10	22.7
4	6.596	0.010	21	22.6	26	41.9
5	13.647	<0.001	18	20.7	29	50.0
NS	13.539	<0.001	13	17.3	24	48.0

As shown in Table 10, prime gender was found to significantly correlate to perseveration for levels 4, 5 and NSs. In every case, when this factor was found to be significant, there was significantly more perseveration for feminine primes than for masculine ones. As with the results from the regression analysis, we will further explore this finding in the discussion that follows.

7 Discussion

The current study was designed to expand our understanding of how perseveration operates with third person referential subjects for second language learners of Spanish and how these patterns change as proficiency increases. Our first research question examined the distribution of subject forms selected on the WCPT and how this differed by group. Our analysis showed that in the early stages of acquisition (levels 1–3), L2 learners select each form at similar rates, suggesting that their selection is not (heavily) influenced by contextual factors. At levels 3–4, however, learners begin showing differences in selection rates for each form.

As L2 proficiency level increased, patterns indicated a shift toward greater selection rates of null forms and lowered selection rates of overt SPs and lexical NPs. For null and lexical NPs, the shift toward native-like rates occurred at level 4, whereas this occurred later (only at level 5) for overt SPs. These results confirm previous research employing both experimental and spontaneous oral data (e.g., Geeslin & Gudmestad 2010, 2011) and suggest that acquiring native-like patterns of subject form variation occurs relatively late in the acquisition process. As previous research has suggested, native-like variation of this structure is guided by a myriad of semantic and discursive attributes of the linguistic context, as well as psychological processes such as priming, and this complexity likely contributes to acquisition rates of these patterns.

Our second research question examined the rates of perseveration. Our analysis showed that from level 3 onward the rates of perseveration differed by prime form, and from level 4 the perseveration rate was highest for null primes and lowest for lexical NP primes. We note that this difference across forms, and the direction of those differences, is not surprising, but it is important from a developmental standpoint to understand when L2 learners begin to reflect these patterns in their own use. Adding further depth to these developmental patterns, our analysis of patterns within prime forms showed that for null subjects, only level 5 learners reached native-like patterns (i.e., did not differ significantly from the native speaker group). For overt SPs the trend over levels was not linear, but level 5 did reach native-like rates of perseveration. Finally, for lexical NPs this occurred slightly earlier in the process and level 4 learners were shown not to differ significantly from level 5 or from the NSs. The gradual differentiation of patterns by form as proficiency increases is consistent with previous research, regardless of the additional factors under examination (e.g., Geeslin et al. 2015). However, the current study is the first to our knowledge to look at changes in perseveration rates by level.

The remainder of our analysis (research question 3 and its sub-questions) sought to explore the role of additional factors in understanding perseveration among our L2 learner groups. Henceforth our dependent variable is whether perseveration occurred, and the goal of the analysis is to determine which independent factors contribute to the occurrence of perseveration. The regression analyses conducted for each level indicate stable patterns of change across proficiency levels, with the exception of the effect of TMA continuity, which was a significant predictor of perseveration at levels 2 and 5, but in opposite directions. We hypothesize that the results for the role of TMA continuity for level 5 are in line with the general trends and the level 2 results are anomalous. This may reflect the pattern of acquisition of the two morphosyntactic forms in alternation

on our instrument, rather than a fact related to perseveration. Specifically, the acquisition of the two forms that mark past tense in Spanish is likely in progress for level 2 learners (Salaberry 2011) and the forms are likely to draw additional attention until they are incorporated into the learners' grammar (VanPatten 1990). Lower rates of perseveration with switches in TMA for level 2 may indicate the attention required to process these forms at this particular level, whereas by level 5 we see the direction of effect that explanations based on referent cohesiveness would predict. Returning then to overall patterns of development attested by the regression analyses, we see that as proficiency increases, learner grammars move toward patterns that are predicted by an increasing number of factors. The importance of the prime form is attested for level 3 and above and the role of gender continuity becomes apparent at level 5, as does TMA continuity. The reader will recall that because of the overlap between gender continuity and a dichotomous switch reference distinction, whereby same reference contexts are also, by definition, contexts where gender is also continuous, we combined these factors into a single, three-part distinction. Consequently, this variable represents a level of complexity that a simpler switch vs. same reference distinction would not. Our hypothesis is supported by earlier studies that show a relatively earlier effect for switch reference when not combined with gender continuity (Geeslin et al. 2015). It is likely that this complexity explains why the variable is significant only for the highest proficiency level and for NSs, whereas Linford & Geeslin (2022) found that referent cohesiveness alone constrained subject expression as early as level 3 among their learners.

The final step in our analysis was to look not only at whether there were shifts in gender of the referent, a reflection of referent cohesion that offers a functional explanation of perseveration, but also whether the prime gender is related to perseveration. This final variable speaks to hypotheses related to surprisal or default forms and is based on psychological processes rather than functional ones (Jaeger & Snider 2007). Our results do, in fact, demonstrate an effect for prime gender beginning at level 4 and continuing for level 5 and the NS group. Specifically, for these groups we see that feminine primes are more likely to perseverate than masculine ones. This finding is consistent with our predictions as the feminine form is described in the literature as less frequent, and the less likely form to serve as the default, both for learners and NSs (Alarcón 2010, Klee 1989, Malovrh 2014). In terms of L2 development, it is reasonable to expect that learners must reach a fairly advanced level of proficiency in order to demonstrate these sophisticated patterns of language processing. Although our findings are as expected, our study is the first of its kind to demonstrate a role for both functional and psychological processes as they relate to L2 perseveration.

4 Variable subject expression in second language acquisition

Throughout the discussion, we have provided an account of the differences between levels in an effort to use our cross-sectional design to illustrate the path of L2 acquisition. In order to provide a snapshot of development that captures all of our findings, we summarize the results in Table 11 in terms of developmental trends by level of proficiency.

Table 11: Developmental stages for subject form perseveration

Level	Summary of patterns
1	Subject form selection rates are near chance; Perseveration rates are not influenced by prime form, gender or TMA continuity or prime gender
2	Similar to level 1, except perseveration rates for null SPs increase, and TMA continuity has a level-specific relationship to perseveration
3	Subject form selection rates are similar to levels 1 and 2; Perseveration rates begin to show differentiation by prime form (null, overt SP and lexical NP); Perseveration rates for null primes and lexical NP primes begin to shift toward native-like patterns
4	Rates of selection for null and lexical NP subjects are significantly different from lower levels (overt SPs are not); Perseveration rates for null primes are significantly different from lower levels and those for overt SP and lexical NP primes are like NSs; Perseveration rates are significantly influenced by prime form, but not by other discourse factors; prime forms denoting feminine referents start to be more likely perseverated and this continues through higher levels
5	Subject selection rates are like NSs for all forms and uniquely so, for null and overt SPs; Perseveration rates for all prime forms are like NSs; perseveration rates are significantly related to prime form, gender continuity and TMA continuity; prime forms denoting feminine referents continue to be perseverated at higher rates
NS	Subject form selection rates are significantly different by form; perseveration rates are significantly related to prime form and gender continuity; prime forms denoting feminine referents are perseverated at higher rates

As seen in Table 11, there is little difference in form selection and rates of perseveration between levels 1 and 2 and patterns at this level do not appear to be influenced by the independent factors in the current study. By level 3, however, rates of selection remain similar to earlier levels but rates of perseveration for null subjects and lexical NPs are shifting towards native-like tendencies. Nevertheless, we do not see a marked influence on rates of perseveration by the referent cohesiveness variables, nor by the prime gender. The learners at level 4 show the sharpest differentiation from earlier levels. They exhibit significantly different rates of selection of null and lexical NP subjects, and they have reached native-like rates of perseveration for lexical NPs and also differ significantly from lower levels in their rates of perseveration of null subjects. However, not all independent variables in the current study have begun to demonstrate a significant relationship to perseveration given that gender continuity nor TMA continuity was not significant for this group. The patterns documented for learners at level 5 are similar to the native speaker group in several ways. First, level 5's rates of selection of all forms are comparable to that of native speakers. Additionally, their rates of perseveration for null, overt SP and lexical NP primes do not significantly differ from native speakers. It is also at level 5 that we begin to see the native-like influence of gender continuity on patterns of perseveration. However, TMA continuity was significant for level 5 but not for the NSs.

In sum, the current study adds to the body of literature on the L2 acquisition of subject expression by exploring factors related to perseveration. Furthermore, our study is the first to demonstrate the role that the prime gender plays in perseveration. Finally, and perhaps most importantly, the results provide evidence for a developmental path in which learners move towards nativelike subject expression as they gain proficiency in Spanish. It is noteworthy that differences remain even between highly advanced learners and native speakers as patterns that reflect this complex interplay of factors are likely to develop only with extensive exposure to the language. Taken together, our findings suggest a role for functional factors that proceed from the discourse at hand, as well as the psychological factors, such as priming, that lead to perseveration. Indeed, the study stresses the importance of considering functional accounts (see Otheguy 2015) as well as the effect of perseveration in order to account for subject expression in both native and L2 Spanish. Our findings further suggest that controlled instruments such as the WCPT in the present study can be effective means for teasing out these subtle differences (see also Geeslin et al. 2015).

8 Conclusion and future directions

The present study has shown that perseveration occurs among L2 learners of Spanish and its study provides important information about the acquisition of subject forms. Specifically, we see that learners first come to differentiate rates of selection of subject forms and then, at higher levels of proficiency, demonstrate varying rates of perseveration by form. Additionally, we see that learners do come to make use of other related factors, such as gender continuity and TMA continuity. We also showed that for learners as well as for native speakers, the examination of prime gender adds dimension to our knowledge of the perseveration of subject forms. These findings are consistent with existing literature but also provide new insights related to L2 acquisition and the role of prime gender in particular.

Benefits of our findings notwithstanding, there are limitations to the current project that provide impetus for future investigations. As is often the case, the benefits of using a highly controlled elicitation instrument were appropriate given the goals of our investigation and, at the same time, it is important to take what we have learned and explore these same patterns in more freely produced samples of language. Specifically, it will be important to expand these findings in contexts where there are a greater number of referents in play and the narrative structure is more complex. We further recognize that learner populations differ, and these results should be expanded to include learners in other contexts and with other first languages. Similarly, adding additional native speaker groups to the study of these factors is essential to corroborate and build on our findings. Clearly our native speaker group serves as an example of the input our learners receive, but they do not represent the diverse speech communities throughout the Spanish-speaking world.

References

- Abreu, Laurel. 2012. Subject pronoun expression and priming effects among bilingual speakers of Puerto Rican Spanish. In Kimberly L. Geeslin & Manuel Díaz-Campos (eds.), *Selected proceedings of the 14th Hispanic Linguistics Symposium*, 1–8. Cascadilla Proceedings Project.
- Alarcón, Irma. 2010. Gender assignment and agreement in L2 Spanish: The effects of morphological marking, animacy, and gender. *Studies in Hispanic and Lusophone Linguistics* 3(2). 267–300.

- Al-Kasey, Tamara & Ana T. Pérez-Leroux. 1998. Second language acquisition of Spanish null subjects. In Suzanne Flynn, Gita Martohardjono & William O'Neil (eds.), *The generative study of second language acquisition*, 161–185. Routledge.
- Ávila-Jiménez, Bárbara. 1995. A sociolinguistic analysis of a change in progress: Pronominal overtiness in Puerto Rican Spanish. *Cornell Working Papers in Linguistics* 13. 25–47.
- Bayley, Robert & Lucinda Pease-Alvarez. 1996. Null and expressed pronoun variation in Mexican-descent children's Spanish. In Jennifer Arnold, Renee Blake & Brad Davidson (eds.), *Sociolinguistic variation: Data, theory, and analysis*, 85–99. Center for the Study of Language & Information.
- Bayley, Robert & Lucinda Pease-Alvarez. 1997. Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change* 9(3). 349–371.
- Bayley, Robert & Dennis Preston (eds.). 1996. *Second language acquisition and linguistic variation*, vol. 10 (Studies in Bilingualism). John Benjamins Publishing.
- Beatty-Martínez, Anne L. & Paola Dussias. 2019. Revisiting masculine and feminine grammatical gender in Spanish: Linguistic, psycholinguistic, and neurolinguistic evidence. *Frontiers in Psychology* 10. Article 751. DOI: 10.3389/fpsyg.2019.00751.
- Bentivoglio, Paola. 1987. *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela, Consejo de Desarrollo Científico y Humanístico.
- Bentivoglio, Paola. 1993. Full NPs in spoken Spanish: A discourse profile. In William J. Ashby, Marianne Mithun & Giorgio Perissinotto (eds.), *Linguistic perspectives on Romance languages: Selected papers from the XXI linguistic symposium on Romance languages* (Current Issues in Linguistic Theory 103), 211–224. John Benjamins Publishing.
- Bini, Milena. 1993. La adquisición del italiano: Más allá de las propiedades sintácticas del parámetro pro-drop. In Juana M. Liceras (ed.), *La lingüística y el análisis de los sistemas no nativos*, 126–193. Dovehouse.
- Blackwell, Sarah & Margaret L. Quesada. 2012. Third-person subjects in native speakers' and L2 learners' narratives: Testing (and revising) the Givenness Hierarchy for Spanish. In *Selected proceedings of the 14th Hispanic linguistics symposium*, 142–164. Cascadilla Proceedings Project.
- Callaghan, Matthew & Catherine E. Travis. 2021. *Languages* 6(1). DOI: 10.3390/languages6010001.
- Cameron, Richard. 1994. Switch reference, verb class, and priming in a variable syntax. In Katharine Beals et al. (eds.), *CLS 30: Papers from the 30th regional*

4 Variable subject expression in second language acquisition

- meeting of the Chicago Linguistic Society, vol. 2: The parasession on variation and linguistic theory, 27–45. Chicago Linguistic Society.
- Cameron, Richard. 1995. The scope and limits of switch reference as a constraint on pronominal subject expression. *Hispanic Linguistics* 6(7). 1–27.
- Cameron, Richard & Nydia Flores-Ferrán. 2004. Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context* 1(1). 41–65. DOI: 10.1075/sic.1.1.05cam.
- Carvalho, Ana M. & Michael Child. 2011. Subject pronoun expression in a variety of Spanish in contact with Portuguese. In James Michnowicz & Robin Dodsworth (eds.), *Selected proceedings of the 5th workshop on Spanish sociolinguistics*, 14–25. Cascadilla Press.
- Carvalho, Ana M., Rafael Orozco & Naomi Shin (eds.). 2015. *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Washington, DC: Georgetown University Press.
- de Prada Pérez, Ana. 2018. Subject pronoun expression and language mode in bilingual Spanish. *Studies in Hispanic and Lusophone Linguistics* 11(2). 303–336.
- de Prada Pérez, Ana. 2020. The interaction of functional predictors and the mechanical predictor perseveration in a variationist analysis of Caribbean Spanish heritage speaker subject pronoun expression. *Languages* 5(4). Article 36. DOI: 10.3390/languages5040036.
- Dumont, Jenny. 2006. Full NPs as subjects. In Nuria Sagarra & Almeida J. Toribio (eds.), *Selected proceedings of the 9th Hispanic linguistics symposium*, 286–296. Cascadilla Proceedings Project.
- Emberson, Jan E. 1987. *The pro drop parameter in the acquisition of Spanish as a second language*. University of Texas at Austin. (Doctoral dissertation).
- Enríquez, Emilia V. 1984. *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: C. S. I. C.
- Erker, Daniel & Gregory Guy. 2012. The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language* 88(3). 526–557.
- Flores-Ferrán, Nydia. 2002. *A sociolinguistic perspective on the use of subject personal pronouns in Spanish narratives of Puerto Ricans in New York City*. Munich: Lincom-Europa.
- Flores-Ferrán, Nydia. 2005. La expresión del pronombre personal sujeto en narrativas orales de puertorriqueños de Nueva York. In Luis A. Ortiz López & Manel Lacorte (eds.), *Contactos y contextos lingüísticos: El español en los Estados Unidos y en contacto con otras lenguas*, 119–130. Iberoamericana Editorial Vervuert.

- Galvan, Javier A. 1999. *The distribution of lexical and null subjects in the acquisition of L2 Spanish by adult speakers of English*. University of Southern California. (Doctoral dissertation).
- Geeslin, Kimberly L. 2018. Variable structures and sociolinguistic variation. In Paul Malovrh & Alessandro Benati (eds.), *The handbook of advanced proficiency in second language acquisition*. Wiley-Blackwell.
- Geeslin, Kimberly L. & Aarnes Gudmestad. 2008. Variable subject expression in second-language Spanish: A comparison of native and non-native speakers. In Melissa Bowles, Rebecca Foote, Silvia Perpiñán & Rakesh Bhatt (eds.), *Selected proceedings of the 2007 second language research forum*, 69–85. Cascadilla Proceedings Project.
- Geeslin, Kimberly L. & Aarnes Gudmestad. 2010. An exploration of the range and frequency of occurrence of forms in potentially variable structures in second-language Spanish. *Studies in Second Language Acquisition* 32(3). 433–463. DOI: 10.1017/S10000033.
- Geeslin, Kimberly L. & Aarnes Gudmestad. 2011. Using sociolinguistic analyses of discourse-level features to expand research on L2 variation in forms of Spanish subject expression. In Luke Plonsky & Maren Schierloh (eds.), *Selected proceedings of the 2009 second language research forum*, 16–30. Cascadilla Proceedings Project.
- Geeslin, Kimberly L. & Aarnes Gudmestad. 2016. Subject expression in Spanish: Contrasts between native and non-native speakers for first and second-person singular referents. *Spanish in Context* 13(1). 53–79. DOI: 10.1075/sic.13.1.03gee.
- Geeslin, Kimberly L., Bret Linford & Stephen Fafulas. 2015. Variable subject expression in second language Spanish. In Ana M. Carvalho, Rafael Orozco & Naomi Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective*, 191–209. Georgetown University Press.
- Geeslin, Kimberly L., Bret Linford, Stephen Fafulas, Avizia Long & Manuel Díaz-Campos. 2013. The group vs. the individual: Subject expression in L2 Spanish. In Jesse Aaron, Jennifer Cabrelli Amaro, Gillian Lord & Ana de Prada Pérez (eds.), *Selected proceedings of the 16th hispanic linguistics symposium*, 156–174. Cascadilla Proceedings Project.
- Geeslin, Kimberly L. & Avizia Y. Long. 2014. *Sociolinguistics & second language acquisition: Learning to use language in context*. Routledge.
- George, Darren & Paul Mallery. 2012. *IBM SPSS statistics 19 step by step: a simple guide and reference*. 12th edn. Pearson Higher Education.
- Grice, H. P. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Speech acts, syntax and semantics*, vol. 3, 41–58. Academic Press.

- Gries, Stephen T. & Gerrit Jan Kootstra. 2017. Structural priming within and across languages: A corpus-based perspective. *Bilingualism: Language and Cognition* 20(2). 235–250. DOI: 10.1017/S6001085.
- Gudmestad, Aarnes & Kimberly L. Geeslin. 2010. Exploring the roles of redundancy and ambiguity in variable subject expression: A comparison of native and non-native speakers. In Claudia Borgonovo, Manuel Español-Echeverría & Philippe Prévost (eds.), *Selected proceedings of the 12th Hispanic linguistics symposium*, 270–283. Somerville, MA: Cascadilla Proceedings Project.
- Gudmestad, Aarnes & Kimberly L. Geeslin. 2022. Overlapping envelopes of variation: The case of lexical noun phrases and subject expression in Spanish. In Manuel Díaz-Campos (ed.), *The Routledge handbook of variationist approaches to Spanish*, 437–449. Routledge.
- Gudmestad, Aarnes, Leanna House & Kimberly L. Geeslin. 2013. What a Bayesian analysis can do for SLA: New tools for the sociolinguistic study of subject expression in L2 Spanish. *Language Learning* 63(3). 371–399.
- Gurzynski-Weiss, Laura, Kimberly L. Geeslin, Danielle Daidone, Bret Linford, Avizia Long, Ian Michalski & Megan Solon. 2018. Examining multifaceted sources of input: Variationist and usage-based approaches to understanding the L2 classroom. In Andrea E. Tyler, Lourdes Ortega, Mariko Uno & Hae In Park (eds.), *Usage-inspired L2 instruction: Researched pedagogy*, 291–311. John Benjamins.
- Harris, James W. 1991. The exponence of gender in Spanish. *Linguistic Inquiry* 22(1). 27–62.
- Hochberg, Judith. 1986. Functional compensation for /s/ deletion in Puerto Rican Spanish. *Language* 62(3). 609–621. DOI: 10.1353/lan.1986.0041.
- Isabelli, Casilde. 2004. The acquisition of the null subject parameter properties in SLA: Some effects of positive evidence in a naturalistic learning context. *Hispania* 87(1). 150–162.
- Jaeger, T. Florian & Neal Snider. 2007. Implicit learning and syntactic persistence: Surprisal and cumulativity. In Lynsey Wolter & Jill Thorson (eds.), vol. 3, 26–44.
- Kanwit, Matthew. 2018. Variation in second language Spanish. In Kimberly L. Geeslin (ed.), *The Cambridge handbook of Spanish linguistics*, 716–732. Cambridge University Press.
- Klee, Carol A. 1989. The acquisition of clitic pronouns in the Spanish interlanguage of Peruvian Quechua speakers. *Hispania* 72(2). 402–408.
- LaFond, Larry. 2002. *The pro-drop parameter in second language acquisition revisited: A developmental account*. University of South Carolina. (Doctoral dissertation).

- LaFond, Larry, Rachel Hayes & Rakesh Bhat. 2000. Constraint demotion and null-subjects in Spanish L2 acquisition. In Joaquim Camps & Caroline Wiltshire (eds.), *Romance syntax, semantics and L2 acquisition: Selected papers from the 30th linguistic symposium on Romance languages*, 121–135. John Benjamins.
- Liceras, Juana M. 1989. On some properties of the “pro-drop” parameter: Looking for missing subjects in nonnative Spanish. In Susan Gass & Jacqueline Schachter (eds.), *Linguistic perspectives on second language acquisition*, 109–133. Cambridge University Press.
- Liceras, Juana M., Denyse Maxwell, Biana Laguardia, Zara Fernández, Raquel Fernández & Lourdes Díaz. 1997. A longitudinal study of Spanish non-native grammars: Beyond parameters. In Ana Pérez-Leroux & William Glass (eds.), *Contemporary perspectives on the acquisition of Spanish*, vol. 1: Developing grammars, 99–132. Cascadilla Press.
- Linford, Bret & Kimberly L. Geeslin. 2022. The role of referent cohesiveness in variable subject expression in L2 Spanish. *Spanish in Context* 19(3). 508–536. DOI: [10.1075/sic.19029.lin](https://doi.org/10.1075/sic.19029.lin).
- Lozano, Cristobal. 2002. The interpretation of overt and null pronouns in non-native Spanish. *Durham Working Papers in Linguistics* 8. 53–66.
- Lozano, Cristobal. 2016. Pragmatic principles in anaphora resolution at the syntax-discourse interface. In Margarita Alonso-Ramos (ed.), *Spanish learner corpus research: Current trends and future perspectives*, 235–266. John Benjamins.
- Malovrh, Paul. 2014. Variability and systematicity in interlanguage development: An analysis of mode and its effect on L2 Spanish morphology. *Studies in Hispanic and Lusophone Linguistics* 7(1). 43–78.
- Montrul, Silvina & Celeste Rodríguez Louro. 2006. Beyond the syntax of the null subject parameter. In Vincent Torrens & Linda Escobar (eds.), *The acquisition of syntax in Romance languages*, 401–418. John Benjamins.
- Morales, Amparo. 1986. *Gramáticas en contacto: Análisis sintácticos sobre el español de Puerto Rico*. Guaynabo, Puerto Rico: Editorial Playor.
- Mougeon, Raymond & Jean-Marc Dewaele. 2004. Preface for special issue of *The International Review of Applied Linguistics* on variation in the interlanguage of advanced second language learners. *International Review of Applied Linguistics* 42. 295–301.
- Otheguy, Ricardo. 2015. Variationist sociolinguistics and linguistic theory in the context of pronominal perseveration. In Rena Torres Cacoullous, Nathalie Dion & André Lapierre (eds.), *Linguistic variation: Confronting fact and theory*, 319–334. Routledge.

- Otheguy, Ricardo & Ana Celia Zentella. 2007. Apuntes preliminares sobre el contacto lingüístico y dialectal en el uso pronominal del español en Nueva York. In Kim Potowski & Richard Cameron (eds.), *Spanish in contact*, 275–295. John Benjamins.
- Otheguy, Ricardo & Ana Celia Zentella. 2012. *Spanish in New York: Language contact, dialect leveling, and structural continuity*. Oxford: Oxford University Press.
- Otheguy, Ricardo, Ana Celia Zentella & David Livert. 2007. Language and dialect contact in Spanish in New York: Towards the formation of a speech community. *Language* 83(4). 770–802.
- Phinney, Marianne. 1987. The pro-drop parameter in second language acquisition. In Thomas Roeper & Edwin Williams (eds.), *Parameter setting*, 221–238. Reidel.
- Pickering, Martin J. & Victor S. Ferreira. 2008. Structural priming: A critical review. *Psychological Bulletin* 134(3). 427–459. DOI: 10.1037/.134.3.427.
- Poplack, Shana. 1980. The notion of the plural in Puerto Rican Spanish: Competing constraints on (s) deletion. *Locating language in time and space* 1. 55–67.
- Quesada, Margaret L. & Sarah Blackwell. 2009. The L2 acquisition of null and overt Spanish subject pronouns: a pragmatic approach. In *Selected proceedings of the 11th Hispanic linguistics symposium*, 117–130. Cascadilla Proceedings Project.
- Rothman, Jason. 2007. Pragmatic solutions for syntactic problems: Understanding some L2 syntactic errors in terms of discourse-pragmatic deficits. In Sergio Baauw, Frank Drijkoningen & Manuela Pinto (eds.), *Romance languages and linguistic theory 2005: Selected papers from Going Romance* (Current Issues in Linguistic Theory 29), 299–320. John Benjamins.
- Rothman, Jason. 2009. Pragmatic deficits with syntactic consequences?: L2 pronominal subjects and the syntax–pragmatics interface. *Journal of Pragmatics* 41(5). 951–973. DOI: 10.1016/j.pragma.2008.07.007.
- Salaberry, M. Rafael. 2011. Assessing the effect of lexical aspect and grounding on the acquisition of L2 Spanish past tense morphology among L1 English speakers. *Bilingualism: Language and Cognition* 14(2). 184–202.
- Shin, Naomi. 2006. *The development of null vs. overt subject pronoun expression in monolingual Spanish-speaking children: The influence of continuity of reference*. City University of New York. (Doctoral dissertation).
- Shin, Naomi. 2012. Variable use of Spanish subject pronouns by monolingual children in Mexico. In Kimberly L. Geeslin & Manuel Díaz-Campos (eds.), *Selected proceedings of the 14th Hispanic linguistics symposium*, 130–141. Cascadilla Proceedings Project.

- Shin, Naomi & Helen Smith Cairns. 2009. Subject pronouns in child Spanish and continuity of reference. In Joseph Collentine, Maryellen García, Barbara Laford & Francisco Marcos Marín (eds.), *Selected proceedings of the 11th Hispanic linguistics symposium*, 155–164. Cascadilla Proceedings Project.
- Shin, Naomi & Ricardo Otheguy. 2009. Shifting sensitivity to continuity of reference: Subject pronoun use in Spanish in New York City. In Manel Lacorte & Jennifer Leeman (eds.), *Español en Estados Unidos y en otros contextos: Cuestiones sociolingüísticas, políticas y pedagógicas*, 111–136. Iberoamericana.
- Silva-Corvalán, Carmen. 1994. *Language contact and change: Spanish in Los Angeles*. Oxford: Oxford University Press.
- Silva-Corvalán, Carmen. 2015. The acquisition of grammatical subjects by Spanish-English bilinguals. In Ana M. Carvalho, Rafael Orozco & Naomi Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective*, 211–229. Georgetown University Press.
- Sodaci, Hande. 2018. *Role of structural priming in contact-induced change: Subject pronoun expression in NL-Turkish*. Radboud University Nijmegen. (MA thesis).
- Torres Cacoullous, Rena & Catherine E. Travis. 2010. Variable *yo* expression in New Mexico: English influence? In Susana Rivera-Mills & Daniel Villa Crésap (eds.), *Spanish of the U.S. Southwest: A language in transition*, 189–210. Iberoamericana.
- Travis, Catherine E. 2007. Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change* 19(2). 101–135. DOI: 10.1017/S0954394507070081.
- Travis, Catherine E. & Rena Torres Cacoullous. 2012. What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics* 23(4). 711–748. DOI: 10.1515/cog-2012-0022.
- Travis, Catherine E. & Rena Torres Cacoullous. 2018. Discovering structure. In Naomi Shin & Daniel Erker (eds.), *Questioning theoretical primitives in linguistic inquiry*, 67–90. Amsterdam: John Benjamins.
- VanPatten, Bill. 1990. Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition* 12(3). 287–301.
- White, Lydia. 1985. The “pro-drop” parameter in adult second language acquisition. *Language Learning* 35(1). 47–62. DOI: 10.1111/j.1985.tb01014.x.
- Young, Richard. 1991. *Variation in interlanguage morphology*, vol. 1. Peter Lang.

Part II

Between personal and impersonal

Chapter 5

Portuguese *a pessoa* and *uma pessoa*: Emerging inclusive impersonals

Eduardo Tadeu Roque Amaral

Universidade Federal de Minas Gerais

Wiltrud Mihatsch

Universität Tübingen

In both Brazilian and European Portuguese, *a pessoa* ('the person'), *uma pessoa* ('a person'), *as pessoas* ('the persons'), *o povo* ('the people'), *o pessoal* ('the people') and some other, more colloquial expressions such as *geral* 'general' (Ornelas de Avelar 2023 [this volume]) are currently developing new impersonal uses (see Afonso 2008, Amaral & Mihatsch 2019, Posio 2021). In this contribution we will analyse the functional changes of *a pessoa* and *uma pessoa*, with a focus on Brazilian Portuguese. Interestingly, all these expressions are originally third person noun phrases excluding reference to speakers and addressees. In impersonal contexts, however, *a pessoa* and *uma pessoa* are predominantly used in non-referential contexts where speaker and addressee may be included. We will try to shed light on the evolution of the functions of the emerging impersonal pronouns *a pessoa* and *uma pessoa* in Brazilian Portuguese, starting with a macro-diachronic analysis tracing the earliest impersonal uses on the basis of the Corpus do português (CDF, Genre/Historical) by Mark Davies and by comparing Brazilian oral colloquial data from the 20th century based on the comparative subcorpus of NURC RJ with contemporary corpus data from Rio de Janeiro (CORPORAPORT) and Minas Gerais (MOC). The corpus analysis will be complemented by acceptability judgments. The different data types will be combined in order to trace the diachronic development of the restrictions determining the impersonal uses and the differences and parallels between the two expressions. We will close by comparing our results with existing studies by Posio (2017, 2021) and Martins (2022) on parallel developments in European Portuguese.



1 Introduction

A considerable number of studies have shown the existence of impersonal and personal pronouns that originate from noun phrases with highly general human nouns such as French *on* (impersonal ‘one’ or ‘you’) or Portuguese *a gente* (personal pronoun ‘we’ with an earlier impersonal meaning, from *a gente* ‘the people’). In Brazilian Portuguese (BP) and European Portuguese (EP) there are several expressions such as *a pessoa*, *uma pessoa*, *as pessoas*, *o povo* and *o pessoal* which take over impersonal functions (Amaral & Mihatsch 2019, Posio 2021). However, although all these expressions are originally third person noun phrases excluding speakers and addressees, *a pessoa* (literally ‘the person’) and *uma pessoa* (literally ‘a person’) are particularly used in contexts where speaker and addressee are included. This study aims to analyse these expressions in order to shed light on the different restrictions determining the exclusive and inclusive impersonal uses and their referential functions, starting out with the classification by Gast & van der Auwera (2013). As typically observed in other processes of grammaticalisation, the functional changes are accompanied by other, rather formal changes not focused on in this study (but see Posio 2021 for European Portuguese and Amaral & Mihatsch 2019 for Brazilian Portuguese), such as their increasingly common occurrence in the syntactic subject position, decategorialisation, which leads to the loss of gender and number inflection and the possibility of adjectival modification, a certain degree of prosodic weakening (Posio 2021) and more specific developments leading to well-established impersonal pronouns such as the impossibility of referring anaphorically to an impersonal antecedent (Cabredo Hofherr 2008: 39–42, 45–48).

Posio (2021) analyses these constructions in EP and discusses their status as potentially grammaticalised referential devices. Posio (2021: 3) assumes that human impersonal referential devices, in opposition to prototypical personal pronouns (such as *I* or *you*), receive their interpretation through inference rather than reference. In example (1), according to Posio (2021: 11), the NP *a pessoa* ‘the person’ gains its speaker-oriented interpretation by inference from the immediate discourse context. The statement about the speaker’s life and work, formulated in the first person plural, is followed by the generalising but still inclusive expression *a pessoa*, corresponding to English *one*. This is another typical use of the *pessoa* constructions, as will be confirmed by our analysis.

- (1) penso que o trabalho absorve-nos muito, (1.3) e acho que nos ocupa mesmo muito atualmente

‘I think that the work absorbs us a lot, (1.3) and I think that it occupies us really a lot these days’

5 Portuguese *a pessoa* and *uma pessoa*: Emerging inclusive impersonals

acho que *a pessoa* vive muito para o trabalho. (.) e já sai do trabalho muito cansada. (0.6) eh:(0.4)

‘I think that *the person* lives a lot for the work. (.) and already leaves the job really tired. (0.6) eh: (0.4) (Posio 2021: 11)

Based on previous studies, Posio (2017, 2021) points out some differences between EP, BP and Peninsular Spanish varieties in the domain of human impersonal constructions. Although Spanish is a language syntactically close to Portuguese, it does not present evidence of a grammaticalisation process resembling the *pessoa* constructions, but both EP and BP do, and recent research has shown some interesting similarities and differences between them. Posio (2021) highlights that it is not surprising that a construction resembling the MAN-impersonals has emerged in the varieties of Portuguese but not in Spanish, since Brazilian Portuguese and to some extent European Portuguese (Posio 2012: 346) do now fill the subject position more frequently and in more contexts than Spanish (see the chapters in Kato & Negrão 2000 and Lamoglia Duarte 2000). The impersonal uses of *a pessoa* and *uma pessoa* are thus a particular feature of Portuguese within the Ibero-Romance language family. When translating the above example into Spanish, the use of the non-grammaticalised Spanish NPs *la persona* and *una persona* leads to the loss of the impersonal reading and in many cases the use of this etymological equivalent sounds awkward. This is clear evidence of functional changes occurring beyond contextual effects and mere inference, by which these expressions have changed.

This contribution focuses on the diachronic and, to some extent, diatopic tendencies in the referential functions of these emerging impersonals. Our analysis is structured as follows: the next section (§2) will give an overview of the main functions and properties of impersonal pronouns. In §3 we will sketch the known paths of pronominalisation of impersonal pronouns and formulate a hypothesis as to how *a pessoa* and *uma pessoa* might have emerged and the degree to which their paths might be related. The following section §4 will explain the methodology. In §5, we present our diachronic corpus analysis, starting with the earliest uses in the *Corpus do Português* (CDP) Genre/Historical and ending with 21st-century colloquial data from Brazil. We will bring together the results of the corpus analysis in §6, where we will outline the most important referential changes of *a pessoa* and *uma pessoa* and point out parallels and differences between BP and EP. These results will be complemented by some of our results from a large-scale acceptability study before we close with a brief conclusion.

2 Impersonality and impersonal pronouns

There are various discourse contexts or communicative settings where speakers need or want to avoid specifying arguments that must be expressed syntactically, for instance when speakers make general statements, when they want to avoid naming known referents or when they ignore the identity of referents. This is particularly important for the subject position, which is typically occupied by an argument referring to a human agent. This position needs to be filled in non-pro-drop languages such as French and is increasingly filled in Brazilian Portuguese (see, for instance, Kato 2000). In European Portuguese it may also be filled even when subject expression is not syntactically, semantically or pragmatically obligatory (Posio 2012: 346). This is also one reason, at least for non-pro-drop languages, why many impersonal pronouns occur in the subject position, as in the case of French *on*, or at least prefer the subject position as in the case of English *one*. Similarly, in the pro-drop language Spanish the impersonal pronoun *uno* ‘one’ occurs most often although not exclusively in the subject position (Cabredo Hofherr 2008: 43–44, 2017: 263). Other aspects play a role in explaining the preference for the subject position, for instance the correlation between the subject position and agency and the frequency of subjects referring to humans.

Apart from the above-mentioned speaker motivations, impersonal pronouns thus allow the subject position to be filled without specifying the referents, a function captured by the definition offered by Gast & van der Auwera (2013: 124):

Impersonalization is the process of filling an argument position of a predicate with a variable ranging over sets of human participants without establishing a referential link to any entity from the universe of discourse.

(Gast & van der Auwera 2013: 124)

The crucial difference between impersonals and other pronouns is indeed their lack of referential anchoring, i.e. they do not introduce new discourse referents that can be taken up anaphorically (see Siewierska 2011: 67), which is a consequence of their primary function, agent defocusing (Achard 2015: 52–55).

Impersonal pronouns, i.e. different types of fillers of argument positions, go back to several distinct types of diachronic sources. They may go back to lexical sources, in most cases general human nouns meaning ‘human being’ as in the case of the so-called MAN-impersonals such as French *on* or German *man*, but also collective or plural general human nouns with the meaning ‘people’, as in the case of the earlier impersonal uses of Portuguese *a gente* and the incipient impersonal uses of Portuguese *as pessoas*, *geral*, *(a) galera*, *(o) povo*, *(o) pessoal*

(Amaral & Mihatsch 2019, Silva & Coelho 2020, Ornelas de Avelar 2023 [this volume]); a similar tendency has been observed for French *les gens* (Cappeau & Schnedecker 2015).

Another important source for impersonal pronouns are personal pronouns as in the case of the impersonal uses of second person pronouns (see Kluge 2016) and third person plural pronouns (Siewierska 2011), but also indefinite pronouns as in the case of English *one* (Moltmann 2010) or Spanish *uno* (Company Company & Pozas Loyo 2009). Notably, we thus find both definite and indefinite sources.

The evolution of the impersonal pair *a pessoa/uma pessoa* with both a definite and an indefinite source is still unclear and we do not know for sure whether they have a common source or two separate sources. Related to their diachrony, we need to find out whether they are functionally equivalent, and whether there are any differences between BP and EP.

3 The evolution of impersonal pronouns and the case of *a pessoa/uma pessoa*

A pessoa and *uma pessoa* share their function of agent defocusing and their emerging status as pronouns, as suggested by Amaral & Mihatsch (2019) and Mihatsch (2017). The more established impersonal pronouns (e.g. French *on*, Portuguese *a gente*) are known to go back to different sources: a generic NP in the case of *on* (see Giacalone Ramat & Sansò 2007, 2011), and a definite referring NP with plural reference as in the case of *a gente* ‘the people’ (see Lopes 2003, 2004) or in the emerging impersonals *as pessoas* ‘the persons’ and *o povo* ‘the people’. The sources are linked to particular discourse strategies, which also explains some of their contemporary functional differences (Amaral & Mihatsch 2019). We will start out with Gast & van der Auwera’s semantic map¹ of impersonal pronouns, refining some aspects. The circular map established by Gast & van der Auwera (2013) for impersonal pronouns synthesises the functions distinguished for the well-studied MAN-impersonals and impersonals based on third person plural pronouns (see Siewierska 2011 and Cabredo Hofherr 2006, 2008, and others). Gast & van der Auwera (2013) distinguish seven main functional clusters, which are each related with the preceding and following cluster, with a link between functions 1 and 7 closing the circle. In what follows we will adopt their terminology:

¹Semantic maps visualise synchronic relations between functions typically expressed by polysemous items as well as diachronic relations (see Haspelmath 2003). Neighbouring functions on a semantic map typically show minor differences.

S refers to the sentence type and HP to the properties of the human participants (Gast & van der Auwera 2013: 27). Sentence contexts can be either veridical or non-veridical, i.e. not showing truth values, as is the case with conditionals or questions and generally modal contexts such as hypothetical sentences. Episodic sentences are anchored in time and space and refer to particular (but not necessarily identifiable) referents, whereas generic sentences are not anchored in any specific point in time or specify any particular referents. Furthermore, impersonals can include or exclude the speaker or, if the speaker is not included, can be oriented toward the speaker and adopt the perspective of the speaker.

Table 1 summarises and illustrates this classification.

Table 1: Clusters of properties of impersonal pronouns (Gast & van der Auwera 2013: 27)

1.	S: veridical/episodic, HP: existential/indefinite/vague <i>They're knocking on the door.</i>
2.	S: veridical/episodic, HP: existential/indefinite/plural <i>They've surrounded us.</i>
3.	S: veridical/episodic, HP: existential/definite <i>They've raised the taxes again.</i>
4.	S: veridical/generic, HP: universal, external <i>They eat dragonflies in Bali.</i>
5.	S: veridical/generic, HP: universal, internal <i>One only lives once.</i>
6.	S: non-veridical/modal, HP: universal, internal <i>One should never give up.</i>
7.	S: non-veridical/non-modal, HP: universal, internal <i>What happens if one drinks sour milk?</i>

The distinctions between inclusive and exclusive, episodic and generic and veridical or non-veridical uses seem to be binary ones. However, there are degrees of inclusiveness, ranging from a clear reference including the speaker to differing degrees of speaker orientation, or degrees of episodicity with more or less explicit and specific temporal anchors, or degrees of universality, between the reference to a totality (humankind in 5) to vague spatio-temporal restrictions as in function 4 in Table 1.

The link between function 1 and function 7, leading to a circular map, might not be obvious at first sight. When we take into account the closeness and referential equivalence of the indefinite pronouns *someone* or *somebody* and the impersonals in these contexts, this link becomes plausible. The contrast between functions 1 and 7 regarding inclusiveness is a pragmatic effect which Gast & van der Auwera (2013: 154) relate to informativity in discourse, with a highly marked and therefore unlikely inclusive interpretation for episodic contexts as in function 1. Nevertheless, impersonal pronouns may also develop inclusive episodic readings, and they may leave the domain of impersonals and become first person personal pronouns as in the cases of BP *a gente* (Lopes 2004) or French *on*. In what follows we will situate known grammaticalisation paths of impersonal pronouns with respect to these functions, point out the main differences between the grammaticalised uses and the non-grammaticalised source constructions and try to position the evolution of *a pessoa* and *uma pessoa* with respect to the semantic map and the known paths.

The most detailed diachronic analyses exist for MAN-impersonals, which are considered an SAE feature² well-attested in many European languages, including Ibero-Romance languages, at least in the medieval period (Giacalone Ramat & Sansò 2007). They only survive in some languages such as French *on* and German *man* and Mainland Scandinavian (Egerland 2003), while they disappeared from Portuguese and Spanish in the 16th century (Company Company & Pozas Loyo 2009, Lopes 2003: 54). The source construction is a generic noun phrase with a lexical noun 'man, human being'. Generic NPs are, of course, not restricted to humans although, in the case of generic NPs with general human nouns, we automatically arrive at a speaker-inclusive interpretation. It is crucial for the study of these pronominalisation processes to consider NPs and not just lexical items. This is easily overlooked since MAN-impersonals do not show any determiners. The starting point of the process of pronominalisation is the medieval or even earlier singular generic bare NP, at a time when the generic function of the definite article is gradually developing (a process fossilised in the French impersonal variant *l'on*). It is notable that the indefinite article cross-linguistically develops its generic readings rather late, in the 16th century in Romance languages, and is generally the last step in the evolution of indefinite articles as in the case of Spanish *un(a)* 'a' (see Givón 1978, 1981, Elvira 1994: 48).

In the analysis of MAN-impersonals, the generic reading as illustrated in the following example does not require grammaticalisation:

²The linguistic area of Standard Average European (SAE) covers Romance, Germanic, Balto-Slavic, Balkan and to some extent Finno-Ugrian, which share a number of grammatical features due to contacts in late antiquity (Haspelmath 2001).

- (2) Non in solo pane vivit *homo* (Matthew 4:4)

‘Man does not live by bread alone’ (Giacalone Ramat & Sansò 2007: 100)

Generic uses as a source of impersonals are associated with particular discourse traditions. Company Company & Pozas Loyo (2009: 1171–1174) identify the role played by religious and moral genres for this development in Old Spanish. In these texts the positing of general truths about human beings is highly relevant. The kind-reference is lost as the impersonal use arises. This is evident in the Old Italian example in (3), a hypothetical context corresponding to function 7 in Table 1, the point where MAN-impersonals enter the domain of impersonals:

- (3) ...quando *uomo* truova la donnola nella via... (Novellino, 32, rr. 7–8)

‘When one finds a weasel on his [sic] way’ (Giacalone Ramat & Sansò 2007: 101)

Here, *uomo* refers to any member of humanity but not to humankind, and it is referentially equivalent to the indefinite pronoun *one* (see also Giacalone Ramat & Sansò 2011: 94). In the course of grammaticalisation these impersonals may subsequently adopt the other functions from 6 to 1 in Table 1. An indicator for the generic source are the first uses in non-episodic non-referring contexts and the still prevailing generic or non-episodic inclusive use of the less entrenched MAN-impersonals in peripheral areas of the SAE languages (Giacalone Ramat & Sansò 2007).

Well-entrenched impersonals (such as English *one*, going back to a different source, however) can be distinguished from generic uses of general human nouns. Our English example (4) features a fossilised bare generic use of *man*, which can also be capitalised (OED, s.v. *man*), while the pronoun *one* cannot refer to the kind and therefore cannot replace *man* in this use:

- (4) With the agricultural revolution, *man* started to settle down, but still many followed a wandering path throughout history.
(<https://eu.coloradoan.com/story/opinion/2017/07/07/editorial-fort-collins-should-stay-course-homeless-services/449108001/>, page last consulted on 28/04/2022)

The referentially equivalent lexical item as well the definite article of *a pessoa* might suggest an analogous path of pronominalisation. In order to clarify this question we need to look at different types of generic NPs.

In English, as well as in other modern European languages with a strongly grammaticalised definite and indefinite article, there are two basic ways of establishing generic reference (see Krifka & Gerstner-Link 1993, Krifka et al. 1995).

The definite article (singular and plural) may directly refer to kinds. This is why they are common with kind predicates, shown in the Portuguese examples taken from Ferreira & Correia (2016):

- (5) O urso polar está quase extinto. (BP, EP)
'The polar bear is almost extinct.'
- (6) Os ursos polares estão quase extintos. (BP, EP)
'The polar bears are almost extinct.'

Kind-generic uses also allow plural (and singular) interpretations:

- (7) The antelope gathers near waterholes. (Krifka & Gerstner-Link 1993)

We believe that this flexibility eases the subsequent step of grammaticalisation of *MAN*-impersonals. Relevant for the grammaticalisation processes of impersonal pronouns is the distinction between the previously described kind-generic interpretation (D-genericity according to Krifka & Gerstner-Link 1993) and generic readings with the indefinite singular article. According to Krifka & Gerstner-Link (1993) this is a case of I-genericity arising at the sentence level. We suggest that diachronically, indefinite articles must have undergone changes to allow for these uses which, as mentioned above, arise relatively late. Krifka & Gerstner-Link (1993) argue that indefinite generic uses are tied to modal uses and habitual uses. We think that diachronically the generic uses of indefinite NPs arise from the communicative strategy of exemplification, i.e. generalisations based on selecting one exemplar that represents the kind. This also explains why indefinite generics cannot refer to accidental properties such as *popular* as in (8) and (9), since this property does not necessarily apply to each exemplar but, rather, is a typical feature characterising many instances of the whole category. Although the property *popular* does not apply to each instance, it is a characteristic feature of the whole kind, therefore allowing the definite generic NP, which does not show this restriction:

- (8) The madrigal is popular. / The madrigal is polyphonic.
- (9) * A madrigal is popular. / A madrigal is polyphonic. (Krifka & Gerstner-Link 1993)

We think that the exemplar-based generalisation not requiring an established category (but an applicability of a predicate to each instance) also explains the flexible noun selection of sentences with indefinite generics as opposed to the

restriction of definite generics to well-established kinds (Krifka & Gerstner-Link 1993). There is no well-established category ‘green bottle’, therefore the definite generic NP is not possible (example 10). However, the generalisation starting out from one exemplar is a cognitive strategy for allowing new categories to be created (see Barsalou 1983 and Mauri 2017 on the linguistic means of creating ad hoc categories) and since the indefinite generic is based on a generalisation strategy starting out from a representative instance, its use is not restricted to well-established categories, but can also refer to ad hoc categories such as *green bottle*:

(10) * The green bottle has a narrow neck.

(11) A green bottle has a narrow neck.

As for the evolution of MAN-impersonals, the early contexts of use in moral and religious texts (Giacalone Ramat & Sansò 2007) go far beyond focusing on just the essential properties of the human species applying to each individual. At the same time the human species is a well-established kind. Therefore definite generics must be the source of MAN-impersonals, although the medieval bare NPs do not give us a direct formal clue.

Could a *pessoa* also arise in definite generic contexts? Our own previous lexical analyses of French, Spanish and Portuguese cognates of PERSŌNA rather exclude this path. While nowadays the Romance cognates of HOMŌ in the gender-neutral reading still almost exclusively occur in generic contexts and were common both in bare NPs and with the singular definite article in the middle ages, the Romance cognates of PERSŌNA (e.g. French *personne*; Spanish *persona*; Portuguese *pessoa*) in turn are only marginally acceptable in generic uses referring to the human species today (Mihatsch 2017: 77f.), although Amaral & Mihatsch (2016) show that, unlike in the other languages investigated, in BP generic uses are slightly more acceptable for *pessoa*. The lexical uses of the cognates of PERSŌNA are rather used in indefinite non-specific contexts with the indefinite article in several Romance languages. The most acceptable definite uses tend to be anaphoric uses with indefinite non-specific antecedents (Amaral & Mihatsch 2016, Mihatsch 2017: 89, Amaral & Mihatsch 2019). We counted 100 occurrences of *homem* in the 15th century subcorpus of CDP (in their order of appearance up to five attestations per text) and found, alongside bare impersonal readings, about 25% definite NPs with a definite generic interpretation. A search for generic uses of *pessoa*, including the bare generic uses of the 16th-century data clearly show that generic uses are isolated cases and that the context is indefinite generic rather than definite generic, even with the definite determiner, appearing mostly in anaphoric uses with an indefinite-generic antecedent. In the

typical example (12) below *a pessoa* does not refer to humankind, but to a hypothetical individual who might happen to taste the water of the described territory. There is no explicit indefinite antecedent that might explain the determiner, but perhaps the preceding impersonal expression *se bebem* ‘are drunk/that people drink’ introduces an indeterminate agent and, just as importantly, establishes the hypothetical situation of a person drinking of that particular water:

- (12) Ha por baixo destes aruoredos grande matto e mui basto, e de tal maneira esta escuro e serrado em partes que nunca participa o chão da quetura rie da claridade do sol & assy esta sempre humido e manando agoa de sy. As agoas que na terra se bebem são mui sadias e sabrosas, por muita que se beba não preiudica a saude *da pessoa*, a mais della se torna logo a suar e fica o corpo desalliuado e saõ. (CDP, Pêro de Magalhães de Gândavo (1570?): *Tractado da prouinçia do Brasil*)

‘Underneath these groves there is a large and vast forest and it is so dark and sawn up in parts that the ground never participates in the heat or brightness of the sun, and so it is always humid and flowing with water. The waters that are drunk on that land are very healthy and tasty, however much water you drink it does not harm a *person’s* (*one’s*) health, most of them (lit. ‘her’) soon sweat and the body is relieved and healthy.’³

This use is related to similar and quite widespread uses of *a pessoa* which can be glossed as ‘the respective person/the person in question’ with a possible indefinite antecedent, but also with a possible indirect anaphoric association with an implicit antecedent, as in (12), and as can be seen in example (13) where *a pessoa* does not have an explicit antecedent, though this can be inferred as people who receive the payment:

- (13) E pera que os vassalos se animem a servir seu rei, principalmente aqueles que servem na guerra, são seus serviços escritos em livro e em modo de crónica. Estes actos dos homens são lidos ante el-Rei, assi pera com a lembrança averem igual premio de seu serviço, como pera gloria de seu nome aos que dele descenderem, e todos são pagos nestes rendimentos da terra; dela se dá per anos, e algũa em vida *da pessoa*, e nenhũa de juro. (CDP, João de Barros (1552): *Décadas da Asia (Década Terceira, Livros I-X)*)

‘And in order that the vassals are encouraged to serve their king, especially those who serve in war, their services are written down in a book and in mode of a chronicle. These acts of the men are read in front of the King,

³The italics in all the examples have been added by the authors.

as a reminder to have thus equal reward of their service, as for the glory of their name to their descendants, and all are paid by the revenue of the land; it is given for years, and some in *the person's* life, and no interest.'

The indirect or associative-anaphoric contexts typical of *a pessoa* might point to another grammaticalisation path, namely the path leading to the evolution of third person impersonals (either with or without an explicit pronoun, depending on the language; see Carvalho 2020 on BP). This is another source with a definite expression which, however, does not have a generic interpretation, but one that is anaphorically linked to an antecedent that remains implicit, similarly to third person plural impersonals.

Third person plural impersonals are more widespread than MAN-impersonals worldwide according to Siewierska (2011: 69). They differ from MAN-impersonals since they are constructions that exclude the speaker and the addressee. Siewierska & Papastathi (2011: 604) suggest an anaphoric source extending to a partially known explicit universal and a partially known source deduced from corporate uses. Third person plural impersonals may plausibly enter the semantic map established by Gast & van der Auwera (2013) as impersonals corresponding to the corporate function 3 in Table 1, referring to indeterminate groups of people associated with established and prominent institutions, the starting point of their impersonal uses. Their reference to (anaphoric) third person referents also explains their prevailing plural interpretation and their external perspective, clearly excluding the speaker.

The emerging Portuguese impersonals *as pessoas, o povo, o pessoal* (see Afonso 2008: 147 for their use in EP, Amaral & Mihatsch 2019 in BP)⁴ share two important features with third person impersonals – their plural reference and their definiteness. We have previously argued (Amaral & Mihatsch 2019) that these undergo a process based on mechanisms comparable to the evolution of third person plural pronouns. However, these lexically-based expressions are impossible or awkward in the corporate function 3 (English *They've raised the taxes again*; Portuguese **As pessoas/o povo/o pessoal aumentou/-aram os impostos*), so there must be some difference in the source and the path of impersonal uses of the former personal pronoun and the lexically filled NPs. We believe that the greater referential vagueness of the lexically based impersonals of this group and the possibility to develop inclusive readings (impossible for third person plural impersonals), at least for the most entrenched expression *as pessoas* (Amaral

⁴Also see Ornelas de Avelar (2023 [this volume]), Silva & Coelho (2020) for further candidates in colloquial Portuguese, and Cappeau & Schnedecker (2015) on the impersonal uses of French *les gens*.

& Mihatsch 2019), is due to the greater ease of bridging inferences of lexically filled NPs as opposed to pronouns (Koenig & Mauner 1999: 230f.), so that no narrow institutional context is required or even possible as a starting point. Apart from the greater flexibility of lexically filled NPs, the pronominalisation of such a large number of expressions in Portuguese might be related to a tendency in Portuguese to use lexically filled NPs for discourse participants.

Notably, the plural or collective Portuguese impersonals *as pessoas*, *geral*, (*a galera*), (*o povo*), (*o pessoal*) (with the approximate meaning of ‘people’), have a functional profile clearly differing from *a pessoa* and *uma pessoa* (Amaral & Mihatsch 2019). *A pessoa* and *uma pessoa* are commonly used in inclusive non-episodic functions while *o povo* and *o pessoal* are quite complementary and tend to appear in exclusive episodic contexts (*as pessoas* shows more flexibility, possibly because of its greater degree of grammaticalisation). Due to these differences between the collective and plural expressions and *uma pessoa* and *a pessoa* which pattern together, we assume that *a pessoa* must arise in a context related to *uma pessoa*.

There is in fact a type of impersonal pronoun that shares important semantic properties, notably the preference for inclusive and non-episodic interpretation, with the lexically filled indefinite NP *uma pessoa*, namely English *one* and Spanish *uno*. In Amaral & Mihatsch (2019) we propose related paths for *a pessoa* and *uma pessoa* and we propose a mechanism closer to indefinite generics (in the sense of Krifka & Gerstner-Link 1993) in discourse strategies of generalisation. These are expected to start in the context of function 7, which neutralises the differences between impersonals and indefinites because in these cases indefinites tend to lead to indefinite-generic readings. This explains why *uno* and *one* and equivalent impersonals found in Spanish, Catalan, Italian, but not in French, Portuguese or Romanian (Cabredo Hofherr 2017: 262), are still restricted to non-episodic uses. According to Company Company & Pozas Loyo (2009: 1197–1207) *uno* starts developing in the 16th century, possibly replacing the Old Spanish MAN-impersonal. Spanish *uno* and English *one* show a strong tendency towards speaker-inclusive or speaker-oriented uses (Moltmann 2010) and we will argue here that this is also true for both *uma pessoa* and *a pessoa*. Speaker-oriented exemplar- or case-related generalisation strategies also explain why their use in function 4 referring to a large, vaguely delimited group might require a more advanced stage of grammaticalisation (also see Moltmann 2010 on relevant restrictions for English *one*). Example (14) shows a possible starting point for the pronominalization of *uma pessoa*, i.e., a use in a hypothetical context:

- (14) *Uma pessoa é bipolar quando apresenta um comportamento no qual ocorrem, com certa frequência, variações entre períodos de bom humor, irritabilidade e tristeza. Essas mudanças podem ocorrer em duas fases: a maníaca, onde a pessoa estará muito feliz e com os ânimos elevados e a hipomaniaca (...) (https://www.significados.com.br/como-identificaruma-pessoa-bipolar/, page last consulted on 28/02/2018)*

‘A person is bipolar when they present a behaviour which, with some frequency, varies between times of good humour, irritability and sadness. These changes can take place in two phases. The manic one, where *the person* will be very happy and in high spirits, and the hypomaniac one’

We have argued (in Amaral & Mihatsch 2019) that the variant *a pessoa* goes back to the same type of context, but refers anaphorically to the first mention featuring the indefinite non-referential general NP *uma pessoa*.

Posio (2021) points out two early uses of *a pessoa* and *uma pessoa* going in the direction of impersonality:

- (15) *a fruta é de maravilhoso gosto, tão leve e sadia que, por mais que uma pessoa coma, não há faltar-se*

‘the fruit has a wonderful taste, so light and healthy that, no matter how much *a person* eats, they will not get tired’ (CDP, Fernão Cardim (1590): *Carta de relação da viagem e missão a Província do Brasil*)

- (16) *e tenho observado que o chocolate é alimento dominante que, em se habituando a ele, não se toma quando a pessoa quer, senão quando quer ele (CDP, Manuel Bernardes (1688): Nova Floresta)*

‘and I have noticed that chocolate is an addictive foodstuff which, once being used to, is not eaten when *the person* wants [to eat it], but when it wants [to be eaten]’

We suggest that both examples perfectly illustrate the generalising move of the speakers (or writers): in both cases particular situations are described, in (15) a hypothetical person who tastes the described fruit, and in (16) a hypothetical person accustomed to eating chocolate (a minority at the time), again following an impersonal expression with *se*, which establishes an impersonal antecedent that might be picked up by *a pessoa*. Possibly *a pessoa* does not require an explicit antecedent, but may establish a rather wide anaphoric bridge from contextually given information. In (16), a translation using *the person* sounds awkward, and *one* would be a more accurate formulation. Unlike Posio, we think that in (16) there is a clear speaker orientation, and this becomes clear if we look at the preceding passage:

- (17) Não usando de chocolate este venerável prelado, formaram disto alguns matéria de reparo, por haver no seu bispado (que era então La Puebla de los Angeles) os melhores ingredientes daquela solene bebida. Respondeu-lhes: Não o faço por mortificar-me, senão porque não haja em minha casa quem mande mais que eu (...) (CDP, Manuel Bernardes (1688): *Nova Floresta*)

‘Since this venerable prelate did not use chocolate, some people took issue with him for having in his episcopate (which was then La Puebla de los Angeles) the best ingredients of that solemn drink. He answered them, I do not do this to mortify myself, but so that there is no one in my house who commands more than I do (...)’

There is also independent evidence for a plausible anaphoric source of the definite variant *a pessoa*, namely the high incidence of anaphoric uses of *a pessoa* and its equivalents in other languages with indefinite unspecific antecedents (cf. Amaral & Mihatsch 2019), note also the probability of topical indefinite Old Spanish *un(o)* to be taken up by an anaphor in discourse in Old Spanish and possibly other languages (see Elvira 1994 who analyzed anaphoric chains in the *Primera Crónica General* and detected this marked tendency for topical uses).

As for the diatopic variation, Martins (2022) suggests that in EP *uma pessoa* is always inclusive and the definite variant *a pessoa* can be both inclusive and exclusive. Posio (2021), on the other hand, expects an inclusive preference for the definite variant due to the definiteness of the speaker role, but detects a general overwhelming tendency toward inclusive or speaker-oriented uses and does not observe any functional differences between the definite and the indefinite variant in his EP data (Posio 2017: 220). According to our hypothesis, the common pronominalisation path (only distinguished by the anaphoric use of the definite variant from the above-mentioned hypothetical use of the indefinite variant) should not lead to differences between the two expressions, although varying degrees of grammaticalisation of the two variants might lead to a subsequent differentiation.

In most uses – leaving aside the entry point of the semantic map in hypothetical uses (function 7) – the translations of the Portuguese examples into other languages make it evident that *a pessoa* and *uma pessoa* have undergone a clear functional change, because replacing either one of them by a lexical NP sounds awkward or is impossible and we need to choose an impersonal pronoun.

4 Methodology

Our contribution has a clear empirical focus on corpus data, complemented by acceptability judgments. The analysed data include both spoken and written language and diachronic as well as synchronic data of Brazilian and European Portuguese varieties. The diachronic corpus is composed of data collected in *Corpus do Português* (Genre/Historical), which features texts from the 13th to the 20th centuries distributed between the genres of spoken, fiction, newspaper, and academic texts. We examine all texts in EP and BP until 1899 and take a selective qualitative look at written Brazilian texts in the 20th century until 1969, when the first oral corpora appear. The contemporary Portuguese data we analysed comprise interviews carried out in different locations in Brazil, especially in Rio de Janeiro and Minas Gerais states. In order to obtain a micro-diachronic data set, a sub-sample of the NURC RJ interviews of two synchronies (the comparative subcorpus from the 1970s and '90s, with speakers from Rio de Janeiro) was analysed and compared to data from the 2010s of the Corporaport project, also from Rio de Janeiro. We selected the subcorpora Copacabana and Nova Iguaçu, all speakers between 18 and 35 years old at the time of recording. In the case of the Minas Gerais data, the sample is composed of sociolinguistic interviews carried out in Montes Claros (Minas Gerais) in this decade (Santos 2021 and Table 2). We have not included occurrences in utterances made by the interviewers. We are, of course, aware of gaps in the corpus sample. This is due to the difficulty of finding minimally comparable spoken language corpora (in this case sociolinguistic interviews) from different periods.

Table 2: Sources of the oral data

Period	1970s	1990s	2010s	2020s
Corpus	NURC RJ	NURC RJ	Corporaport	MOC – Minas Gerais
Size (tokens)	80,759	120,332	139,450	181,256

Since ambiguous uses that allow both a pronominal and a lexical interpretation are necessary for grammaticalisation to take place, we also expect many unclear cases in our data. For our analysis we have isolated the clear cases that ought not to be translated by *a person* and *the person* but rather by an impersonal pronoun, and which deviate from the lexical uses, notably in terms of number indeterminacy and inclusiveness. Although we both checked all occurrences we identified as candidates for impersonal uses and discussed potentially ambiguous

attestations, especially in the case of possibly anaphoric uses with an implicit antecedent, there remains a certain subjectivity in the data classification.

Although the size of the analysed data base is large, the absolute frequencies of the impersonal uses are not very high. We have therefore decided to work with the relative frequencies, but have not undertaken an analysis checking for statistical significance. Due to the infrequency of occurrences in some periods and to the diverse nature of the data, the focus will thus be more qualitative than quantitative.

It is well-known that some constructions do not occur in a corpus with the same regularity as others, and this applies to some of the constructions analysed in this study. It is difficult to find each of the different functional types in sociolinguistic interviews. For this reason we opted to complete the corpus analysis with an acceptability study, on the basis of eight online questionnaires of acceptability judgments which altogether featured 237 different sentences that included *a pessoa* and *uma pessoa*, but also other expressions with impersonal uses not discussed in this contribution, such as *as pessoas*, *o pessoal*, *o povo*, *a galera* and third person plural impersonals. The participants had to evaluate the acceptability on a Likert scale of 1 to 5, with 1 corresponding to a “totally acceptable use” and 5 to a “not acceptable use”. The questionnaires were uploaded to Google Forms and the participants were asked to complete them on two occasions, in 2019 and 2021. The 240 participants are all native speakers of Brazilian Portuguese and, at the time the data were collected, were university students of the Faculty of Letters of UFMG in Minas Gerais, and were between 18 and 55 years of age. We obtained 30 answers per sentence. The results of the acceptability judgments were submitted to an SPSS analysis (a Games-Howell post hoc test) to check the statistical significance of the responses.

5 Corpus analyses

5.1 A macro-diacronic overview: From the earliest occurrences to the 20th century

The source constructions as well as the first reanalysed uses of *a pessoa* and *uma pessoa* go back several centuries (see also Martins 2022 and Posio 2021 with observations on diachrony). For this study we manually sorted all uses of *pessoa* and its graphic variants with a full range of determiners and all text types and varieties of CDP until 1899. We also looked at the time-span between 1900 and 1970 and the beginning of the first oral subcorpus analysed by us (NURC from the '70s), with data from between 1971 and 1978.

Since attestations are existent but scarce before the 19th century, we took into consideration all syntactic positions. Similarly to Martins (2022), we found the first attestations in the 16th century data.

As pointed out in §3, generic uses are not a plausible source for emerging impersonal uses of *a pessoa* and definite generic attestations are in fact quite rare; the following examples are two candidates from the 16th century and one from the 17th century that might offer an isolated generic reading. (18) is perhaps a difficult case as it is used in a proverb, thus in a fixed expression, while in (19) and (20) we might also think of a case-based use corresponding to ‘the person in question’, referring in (19) to the hypothetical person offended and in (20) to the person likely to appreciate the paintings:

- (18) Fome, & frio, mette *a pessoa* com seu inimigo. (CDP, Antonio Delicado (1651): Adagios)⁵
‘Hunger and cold unite *the person/the human being* with one’s enemy’
- (19) Uma escândola com’esta / enche de birra *a pessoa* / nem tal chufa nam é boa pera béspera de festa.
(CDP, Gil Vicente (C16th): Obra completa (A-M) - Clerigo da Beira)
‘An offence like this fills *the person/the human* with tantrums, such mockery is not good for the eve of a party’
- (20) E d’isto fará o pintor, para ser visto com mór gosto e que muito commova *a pessoa*. (CDP, Francisco de Holanda (1561): Da Pintura Antiga)
‘And the painter will do this, in order to be regarded with greater appreciation and in order to strongly move *the person/the human*.’

Other examples are less clear; in (21) *a pessoa* seems rather to correspond to ‘outer appearance’:

- (21) lhe merquei eu em Lixboa / dum que chamam solivão / que faz luzir *a pessoa* / e merquei-lhe dum judeu (CDP, Gil Vicente (C16th): Obra completa (N-Z)
‘I bought it in Lisbon / of one they call *solivão* / that makes *the person* shine / And I bought it from a Jew.’

Closest to the first attested uses are plausibly anaphoric contexts as in (22). Note the generic use of *o homem* ‘the human being/Man’ referring to all humans and the following impersonal *se*, then the transition to the more particular case (*acontece algúuas vezes* ‘it happens sometimes’) and the indirect anaphor *a pessoa*

⁵The same proverb appears in an identical form in a later collection.

(which explains the definite article) referring to one type of case described here, a use clearly close to indefinite genericity:

- (22) Capítulo LVI – Que *o homem* nom deve presumir de si, posto que virtuoso seja, porque muitas vezes acontece que soo per hûu defeito se perde: Ora está o coração em seu castello alto aseentado, bem fundado e cercado e bem guarnido de vitalhas e d'água. Ora acontece algûuas vezes que *a pessoa* esguarda o castello de seu coração, e vee-o tam forte que *se* segura mais que rrazom, per que caae em algûua niglligência (CDP, 1400–1500: *Castelo Perigoso*)

‘Chapter 56th – That *man* should not boast if he is virtuous, because it often happens that by a single defect he is lost. Now the heart is set high in its castle, well-founded, and surrounded, and well-furnished with food and water. Now it happens sometimes that *the person* guards the castle of his heart and sees it to be so strong that he thinks himself safe more than is reasonable, so he falls into some neglect.’

Up to the 19th century there were only four possibly impersonal use of *a pessoa* found, for instance example (16), repeated in (23)

- (23) e tenho observado que o chocolate é alimento dominante que, em se habituando a ele, não se toma quando *a pessoa* quer, senão quando quer ele
‘and I have noticed that chocolate is an addictive foodstuff which, once being used to, is not eaten when *the person* wants [to eat it], but when it wants [to be eaten]’ (CDP, Manuel Bernardes (1688): *Nova Floresta*)

First impersonal uses of *uma pessoa* are slightly more frequent (seven occurrences) than *a pessoa*, possibly because the indefinite variant may occur in hypothetical uses, while the use of the definite variant *a pessoa* needs an additional anaphoric bridge. The inclusive interpretation in (24) in a typical combination with another impersonal expression (*se*) points to a functional change:

- (24) aliem dos caminhos serem ingrimes, estreitos e perigosos para a banda do mar, e *se* passava por humas balsas ou bruassaes, dos quaes desviando hum pé hia *huma pessoa* cahir nos abismos por aquella rocha abaixo junto do mar em penedia solta. (CDP, Frois (1560–1580): *Historia do Japam* 3)
‘because, in addition to the paths being steep, narrow and dangerous towards the sea, and one passed by rafts or heather, from which *a person/one*, deviating with one foot, would fall into the abysses by that rock below by the sea in loose boulders.’

Example (25) is less straightforward because this is most likely an exclusive use by a European writer commenting on Japanese practices, so a lexical interpretation ‘a person’ might also be possible:

- (25) ...tenho achado hum meio que me parece não pouco conveniente e acomodado para poder evitar todos os males que daqui se podião seguir. O qual hé repar-me eu (que hé manifesto sinal em Japão de *huma pessoa* deixar e renunciar o mundo) e recolher-me na Igreja e renunciar todo meo estado temporal,...e [o] (CDP, Frois (1560–1580): *Historia do Japam*)

‘I have found a way that seems to me convenient and comfortable in order to avoid all the evils that could follow, which is to shave myself (which is a clear sign in Japan of *a person* leaving and renouncing the world) and retire to the Church and renounce all my worldly estate’

An inclusive example with a typical generalisation from a subjective narrative to a generalisation is (26):

- (26) porque na primeyra noite que chegamos fomos logo roubados de quanto leuauamos, sem nos deixarem nem hua camisa, porque como a casa da prisão era muyto grãde, & muyta a gente que estaua nella (porque segundo nos affirmarão passauão de quatro mil presos) não auia onde *hua pessoa* se pudesse assentar que logo não fosse roubado & cuberto de piolhos. (CDP, Fernão Mendes Pinto (1603): *Peregrinação*)

‘because on the first night we arrived we were immediately robbed of everything we wore, they did not even leave us a shirt, because the prison house was very big and many people were there (as we were told there were more than four thousand inmates) there was nowhere *a person/one* could sit that wouldn’t soon be robbed and covered with lice’

In total, we identified one plausible impersonal use of *uma pessoa* in the 16th century (example 25), possibly two in the 17th century and four in the 18th century.⁶

The picture changes in the 19th century where we see an overall increase (although still with a low frequency) in impersonal uses of *uma pessoa*. There are 73 clearly impersonal occurrences in different syntactic positions, which corresponds to 7.6 occurrences per million words (rounded to one decimal place),

⁶Posio (2021) points to a remark on the early impersonal uses of *uma pessoa* in Nunes (1919). For *a pessoa* we found one case in the 16th and two in the 17th century. All uses were non-episodic and except for (25) inclusive.

5 Portuguese *a pessoa* and *uma pessoa*: Emerging inclusive impersonals

while in the 18th century there are only 1.8 cases per million words. The subject position (including infinitival subjects) prevails (see Table 3).⁷

Table 3: Attestations of impersonal *uma pessoa*, 19th century, CDP, EP and BP, per million words.

<i>Uma pessoa</i>	Subject of an inflected verb	Infinitival subject	Complement (of P and V)
EP	10.26	5.92	3.95
BP	0.80	1.07	0.40

The European subcorpus shows 19.7 impersonal occurrences of both *a pessoa* and *uma pessoa* per million words while the Brazilian subcorpus has 2.7 cases per million words, so a European origin might be plausible.

Impersonal *a pessoa* is attested three times in texts by the Brazilian writer Machado de Assis, in one case as an infinitival subject and twice in complement position, as in (27) (featuring in the 19th-century subcorpus). In (27), *a pessoa* appears in a generalising statement after a narrative sequence as in many other early occurrences (see examples (24) and (25)):

- (27) Paulo correu a pedir socorro. Santos entrou desorientado no quarto, a tempo de ouvir à esposa algumas palavras suspiradas e derradeiras. A agonia começou logo, e durou algumas horas. Contadas todas as horas de agonia que tem havido no mundo, quantos séculos farão? Desses terão sido tenebrosos alguns, outros melancólicos, muitos desesperados, raros enfadonhos. Enfim, a morte chega, por muito que se demore, e arranca *a pessoa* ao pranto ou ao silêncio. (CDP; Machado de Assis (1904): *Esau e Jacó*)

‘Paul ran for help. Santos entered the room disoriented, in time to hear a few sighed and final words from his wife. The agony started right away, and lasted for a few hours. Counting all the hours of agony that have been in the world, how many centuries will it be? Some of these will have been tenebrous, others melancholy, many desperate, rare ones dull. Finally, death arrives, no matter how long it takes, and pulls *the person/one* into tears or silence.’

⁷According to CDP the genre/historical corpus contains approximately the following number of words (in millions): 16th century: 4.4; 17th century 3.4; 18th: 2.2; 19th century: 10.0; 20th century: 10.2 (BP) and 10.5 (EP) (for the exact numbers see <https://www.corpusdoportugues.org/hist-gen/>, page last consulted on 01/05/2022).

What many of these uses have in common is that they appear in Realist novels and in many cases in reported speech or free indirect speech, typically in the previously described generalisation tendency. Such uses can also be found with characters from labouring classes such as servants, whose manner of speaking is also often imitated in these passages. This suggests that there may have been a tendency to use the expressions in oral colloquial speech:⁸

- (28) Pois olha que dúvida! Se se fosse a direito lá por baixo, era mais perto, mas... – Mas foi então pelo prazer de trepar que me trouxeste por aqui? – Não é isso, patrão; mas bem vê V. S.a que o caminho lá por baixo é todo cortado por quintas e campos, e é preciso dar tais voltas, que afinal fica mais longe. Depois, com a chuva que tem caído, faz lá ideia de como estão os riachos por lá! Só o esteiro do almargeal é para *uma pessoa* se afogar. Mas tenha o patrão paciência, que pouco falta agora. (CDP, Dinis, Júlio (1868): *A Morgadinha dos Canaviais*)

‘Well, look what a doubt! If you went straight down there, it would be closer, but... – But was it then for the pleasure of going up that you brought me here? – It’s not that, boss; but you can see that the path below is all cut by farms and fields, and it’s necessary to take such loops, so after all it’s farther away. Then, with the rain that has been falling, you have an idea of how the streams are there! Only the creek of the marsh is for *a person/one* to drown. But have patience, because there is little more now.’

In the following episodic example the speaker orientation and inclusiveness is signaled by *aqui* ‘here’ and the whole passage is emotionally charged:

- (29) Estou eu aqui a chamar há mais de duas horas e vossemecê aparece-me lá quando é muito do seu gosto! Isto atura-se? A culpa tem quem eu sei.. Tu cuidas que mandriar não é roubar? - Mas... - Cale-se! Ouça e cale-se. Tens a língua muito pronta para responder. Ora toma-me cautela, senão vais já, já pela porta fora. Pouca vergonha! *Uma pessoa aqui* aflita, com as coisas por fazer, a querer mandar onde é preciso e não aparece um criado nesta casa! A pagar-se aqui umas soldadas por aí além, e, quando se quer o serviço feito, tem *uma pessoa* de o fazer por suas mãos.. Tu cuidas que isso não é pecado também? Deixa, meu amigo, que tens boas contas a dar de ti. Quem é que lhe deu licença de sair sem ordem de seus amos? Faz favor de

⁸While the earlier examples tend to be from privileged writers and reflects their perspective, these types of use do not appear in non-fiction, academic or technical texts, for example.

me dizer? (CDP Dinis, Júlio (1868): *A Morgadinha dos Canaviais*)

'I've been calling here for over two hours and you come when you like! Is this acceptable? Whoever is to blame... Do you think that to dawdle is not to steal? - But... - Shut up! Listen and shut up. You're always very ready to respond. Now take care, otherwise you'll go right out the door. It's a shame! *A person here* distressed/one here *being distressed*, with things to do, wanting to order where it's needed and not having a servant in this house! To pay here some wages, and, when you want the job done, *a person/one* has to do it with their own hands... Do you think that this is not a sin too? Wait, my friend, you have good explanations to give about yourself. Who gave you permission to leave without orders from your bosses? Will you please tell me?'

Example (30) was interpreted as an episodic use; however, as a reviewer of this contribution has pointed out, this could also be a statement about a generalised, typical behaviour. In any case *uma pessoa* has a clear inclusive reading:

- (30) Joana ia já a retirar-se desconsolada, quando avistou Clara na alameda. Vendo que não era percebida por ela, chamou-a. – Fale à gente. Então que modos são esses agora? Passa por *uma pessoa*, como cão por vinha vindimada! - Não a tinha visto - disse Clara, (CDP; Dinis, Júlio (1863): *As Pupilas do Senhor Reitor*)

'Joana was going to leave saddened when she saw Clara in the alley. Seeing that she was not noticed by her, she called her. – Tell us. So what are these behaviours now? You pass by *a person/one*, like a dog by a harvested vineyard! - I hadn't seen you, said Clara (...).'

A glance at the global frequencies of the forms (not distinguishing impersonal uses) of the 20th-century data of the CDP clearly points to an oral evolution in the 20th century and a difference between EP and BP as for *a pessoa*.⁹

For the 20th century we looked only at Brazilian data until 1969 in the CDP, the time when the first oral data of our sample are documented. We found 11 clear cases of impersonal *uma pessoa* and 5 cases of impersonal *a pessoa*, i.e. uses that have an impersonal interpretation and which do not correspond to the lexical use as in the case of the speaker-inclusive uses or anaphors in the case of the definite *a pessoa*. The latter all appear in a novel (*Meu destino é pecar* from 1945 by Nelson Rodrigues) so they cannot be used as evidence of an increase in frequency.

⁹For a comparison between BP and Spanish, see Amaral (2017).

5.2 A micro-diacronic study of NURC RJ

The rough picture of a purely formal search in the 20th century subcorpus of CDP given in Figure 1 points to a particular oral use of *pessoa*. We assume that the lexical uses are by no means restricted to colloquial language, but rather tend to be tied to more formal written texts, while the higher frequency in oral texts might be due to the incipient impersonal uses, for example in (28) and (29), where the impersonal expression appears in dialogue sequences. This is why we concentrate on oral (partly colloquial) data from the 1970s to the 21st century. We consulted the comparative subcorpus of NURC RJ from Rio de Janeiro’s *Norma Urbana Linguística Culta* Project) with a subselection of data collected from the 1970s and complementary data from the 1990s, with participants with a university degree, born in Rio de Janeiro and children of parents who were preferably from Rio de Janeiro (Barbosa et al. 2021), totalling 38 interviews in the form of dialogues between interviewers and informants. The subcorpus from the 1990s includes recordings with speakers interviewed in the 1970s and interviews with new informants. In our analysis we did not take into account the age or gender of the speakers. In contrast to our previous analysis of written texts in §5.1, we restricted our analysis to the subject position, since various previous studies have shown the higher frequency of impersonals in this position (Amaral & Mihatsch 2019, Posio 2021).

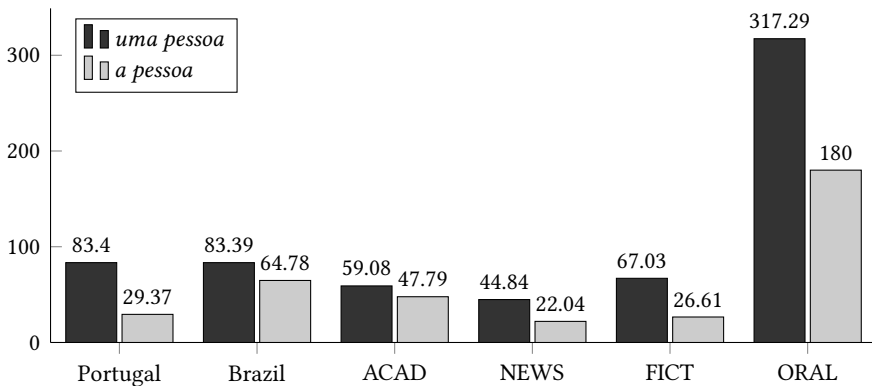


Figure 1: Frequencies of the expressions *a pessoa/uma pessoa* in the 20th century, CDP (per million words) in EP and BP and different genres.

Although the data come from different sources (most importantly, we do not have oral data from prior to the 1970s), the data might reflect an increase from the 19th to the 20th century. Alternatively, the increase in impersonal uses starts

earlier, but this is not evident in the available written data. Interestingly, in the 19th century *uma pessoa* prevails both in EP and BP. In both subcorpora in NURC *a pessoa* clearly prevails, a comparison of the subcorpus from the 1970s and 1990s shows that the proportion of *uma pessoa* diminishes, as shown in Figure 2.

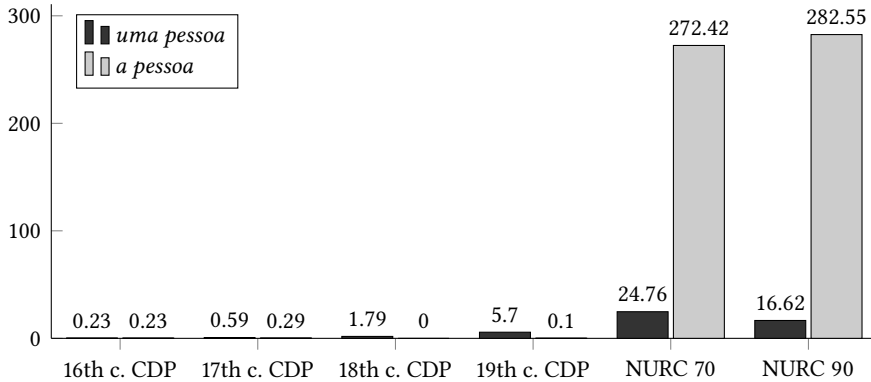


Figure 2: Impersonal *a pessoa* and *uma pessoa* in CDP and NURC RJ for the '70s and '90s (in subject position, per million words).

5.3 Contemporary oral data from Rio de Janeiro and Minas Gerais

The comparison of the oral data from the 1970s, 1990s, 2010s and 2020s show that the proportion between *uma pessoa* and *a pessoa* – with impersonal *a pessoa* being far more frequent than *uma pessoa* – remains relatively stable, with a slight increase of *uma pessoa* from the 1990s to the 21st century, and a more pronounced growth in the use of *a pessoa* in the 21st century in comparison with the 20th. The comparison of *a pessoa* and *uma pessoa* in the samples from the 2010s and 2020s does not show great differences (this is also true for the referential types and contexts, discussed in §6). The analyses of the frequency of *a pessoa* and *uma pessoa* in two contemporary samples of BP also offer important insights in terms of regional variation. There are no significant differences between the 21st-century data from CORPORAPORT Rio de Janeiro and Minas Gerais MOC (see Figure 3). Considering previous studies of *a pessoa* and *uma pessoa* in BP and EP, the results allow us to suppose that perhaps more significant differences can be found between BP and EP than between different Brazilian dialects.

The following examples illustrate contemporary uses. In example (31), the participant comments on the current facilities for students. In this context, he uses *a pessoa* to express the general statement that if one wants to study it is easy

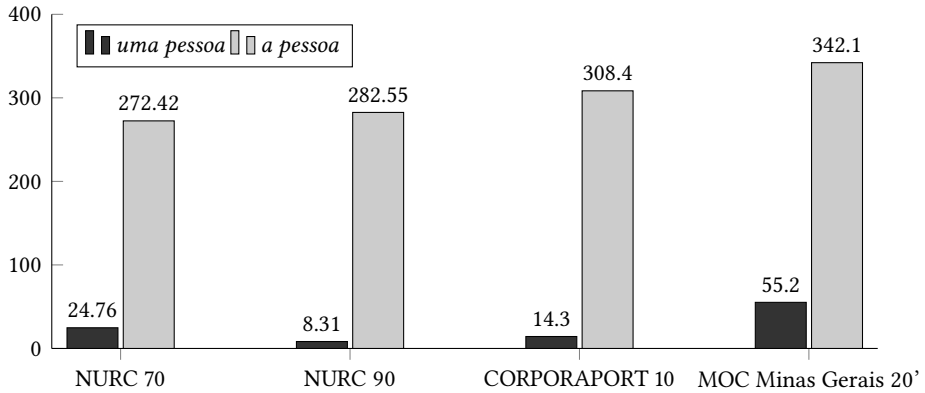


Figure 3: 20th- and 21st-century oral data (impersonal uses in subject position per million words)

because there are many possibilities. Note that after using the formally feminine NP *a pessoa*, the participant anaphorically refers back with the masculine personal pronoun *ele* 'he'. Posio (2021: 13) mentions the same phenomenon when he analyses occurrences of *uma pessoa*. The author has no explanation for the observation that masculine agreement is found only with indefinite form *uma pessoa* but not with the definite *a pessoa* in the corpus of EP he mentions. Masculine agreement is also common with the grammaticalised pronoun *a gente*, which also goes back to a feminine noun, and it is unsurprising that the same kind of agreement can be observed in *pessoa* constructions. Both *gente* ('people') and *pessoa* ('person') are feminine nouns in their original lexical meaning, and the masculine anaphors therefore point to a loss of the feminine gender in the process of pronominalisation (see Lopes 2004 on the changes in agreement of *a gente* in the course of grammaticalisation).

- (31) porque eu trabalho e estudo né eu faço os dois eu não deixei de trabalhar pra poder estudar pros concursos... isso acaba apertando o horário você tem que estudar até mais tarde final de semana você acaba estuDANdo ...
[...]
[...] hoje em dia tem tanto curso tem tanta:: tanta facilidade de acesso à informaÇÃo:: à/ à AUla te/ tem aula telepresenciAL tem: um monte de facilidade ... [...] eu acho que se *a pessoa* quiser estudar mesmo pode ser o que FOR *ele* vai assim você hoje tem MUIto CURso eh:... (COP-A-3-M, male, 31 years old, high educational level)

5 Portuguese *a pessoa* and *uma pessoa*: Emerging inclusive impersonals

‘because I work and study, you know? I do both I didn’t stop working to study for the public contests... this ends up tightening the schedule you have to study late on weekends you end up studying... [...] nowadays there are so many courses there is so much:: it’s so easy to have access to information:: to/ to the CLASS you/ there are non presential classes there is: everything is so easy... [...] I think if *the person* really wants to study anything *he* will, so today you have a LOT of COURSES eh:..’

Examples with a similar interpretation to the ones above appear in the recent data collected in Minas Gerais. In example (32), the interviewer and the participant talk about physical exercise. The participant admits that he does not do any sports, but after that states that it is important for health. To justify that statement, he uses the generalising NP *a pessoa*.

- (32) Interviewer: ...tem algum outro esporte de que o sinhô gosta?
Participant: uah tem esporte que eu gosto só que eu num eu praticamente num tô pratican[d]o quase nenhum caminhada né éh fazer caminhada exercício físico isso é muito bom é necessário *a pessoa* fazer para a saúde (MOC 19 - J.F, male, 59 years old, low educational level)
‘Interviewer: ...is there any other sport you like? Participant: why, there are sports that I like but I don’t practise any like walking, you know, walking physical exercise that’s very good it’s necessary for *the person/one* to do it for one’s health’

In example (33), the inclusion of speaker and addressee can also be observed. The interviewer uses the first person pronoun *nós* and the participant answers with the NP *a pessoa*. Both uses are close to the universal generic function 5 in Table 1.

- (33) Interviewer: ok intão em sua opinião qual é o nosso destino depois da morte? pra onde nós vamos?
Participant: depois da morte *a pessoa* vai ter um lugar de repouso né de descanso né com Deus é a promessa da palavra de Deus após a morte *a pessoa* terá ôta vida cum Deus né (MOC 19 - J.F, male, 59 years old, low educational level)
‘Interviewer: ok so in your opinion what is our fate after death? where are we going to?
Participant: after death *we/one* will have a place to rest, you know, rest with God, it’s the promise of the word of God after death *we/one* will have great life with God, you know’

According to our own observations in informal speech conducted in Brazilian Portuguese,¹⁰ especially with younger speakers, *a pessoa aqui* acquires an interpretation equivalent to the first person pronoun. In all the examples below, written in informal blogs and found through a search in *Corpus do Português – Web/dialects*, the speaker uses this construction as a first person pronoun:

- (34) Ana, querida, ainda não postei seus selinhos, a homenagem e o prêmio dardos, mas é que *a pessoa aqui* tá mesmo enrolada tá? Não é falta de atenção e carinho nao, tá? (<http://anamgs.blogspot.com/2010/06/gente-se-acostuma.html>)

‘Ana, dear, I haven’t posted your stamps, the homage and the Dardos award yet, but *the person here* (= I) is really messed up, right? It’s not a lack of attention and affection, right?’

- (35) No box tinha uma cortina, bem fofa, porque *a pessoa aqui* sempre gostou de coisas fofas né? (<http://cassifamilia.blogspot.com/2013/06/minha-casa-linda-e-com-meu-jeito-meu.html>)

‘In the bathroom stall there was a curtain, very cute, because *the person here* (= I) always liked cute things, right?’

This might point to a use in the direction of a first person pronoun, thus suggesting an interpretation similar to *a gente* (for instance Lopes 2004) and French *on* (see Coveney 2000).

The cases with *uma pessoa* are less frequent, though it is worth citing some examples. In example (36), there is an exclusive interpretation. In example (37), the participant uses the NP *uma pessoa* and also *a pessoa* (possibly anaphorically referring to the impersonal antecedent *uma pessoa*) in a sequence of conditional clauses. Note that these are cases close to the lexical use with a first indefinite NP and a following anaphoric NP:

- (36) não acredito que *uma pessoa* se sinta plena ou realizada vivendo de esmola vivendo de migalhas eu acho que se e/ *elas* deveria(m) ter a oportunidade de ser cidadãos cidadãos né tudo (COP-A-2-M, male, 19 years old, middle educational level)

‘I don’t believe that *a person/one* feels full or fulfilled, living on alms, living on crumbs I think if and/*they* should have the opportunity to be citizens, you know?’

¹⁰The corpora mentioned in Figure 3 include sociolinguistic interviews and are therefore less colloquial than, say, free conversations or occurrences derived from the internet.

- (37) com a reforma mudou muito isso mais se num tiver emprego num tem comida num tem como *uma pessoa* pagar uma água num tem como *a pessoa* pagar uma luz aí começa num estado que *a pessoa* ou então mesmo a comprar *a pessoa* as vez prefere comprar comida dentro de casa do que comprar uma roupa (MOC 07 - A.R, male, 25 years old, high educational level)
- ‘with the reform this has changed a lot, but if you don’t have a job, you don’t have food, *the person/one* doesn’t have money to pay for water there is no way to pay for electricity, then it reaches a state that you are even buying, you sometimes prefers to buy food at home than buy clothes’

6 Diachronic and diatopic tendencies of functional differentiation

The general impression from existing studies on BP (Amaral & Mihatsch 2019) and EP (Posio 2021, Martins 2022) is that *a pessoa* and *uma pessoa* specialise in non-episodic uses, i.e. functions 5 to 7 in Table 1. They also agree on the essentially inclusive or speaker-oriented character and their tendency to occur in generalisation strategies.

We analysed the corpora across the centuries according to the functions in Table 1, isolated uses in inclusive as well as exclusive contexts. Figure 4 illustrates percentages of the respective impersonal uses; as explained above, CORPORA-PORT contains data from the 2010s and MOC from 2020.

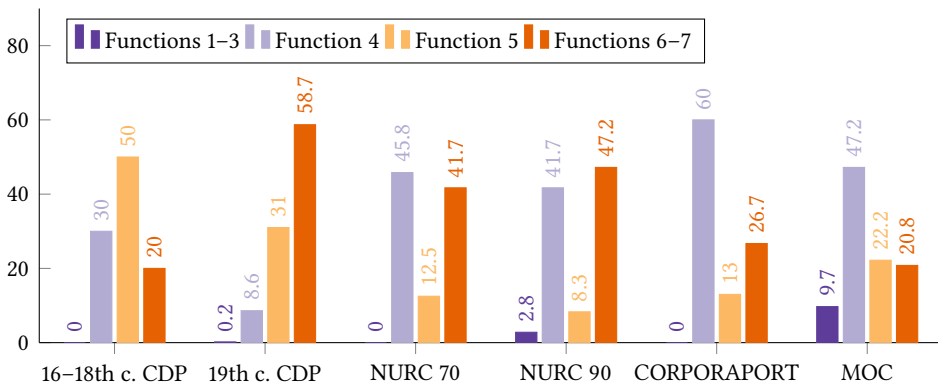


Figure 4: The functional development of *a pessoa* and *uma pessoa* (% of uses) in diachrony in subject position

Functions 4, 5, 6 and 7 (the latter two taken together) appear from the first attestations on, although sporadically. The most frequent non-veridical (mainly hypothetical) uses in the 19th century (functions 6 and 7) decrease in proportion in the 20th and 21st century, the generalising function 4 comes to prevail from the second half of the 20th century and episodic uses first appear in the 19th century and show a slight increase in the contemporary data. This picture shows an increasing functional diversification in the course of grammaticalisation. It is hard to pin down one of the functions as a starting point of the emergence of the impersonal function. General statements and universal truths prevail in the isolated uses before the 19th century, while in the 19th century non-veridical uses dominate. These overlap with function 7, which also links indefinite (non-impersonal) uses and impersonal interpretations, since conditionals tend to allow both interpretations (see §3). Possibly these contexts contributed to an entrenchment of the impersonal function.

Although episodic uses are rare, they have been attested (see Amaral & Mihatsch 2019). Our observations, as well as those of Posio (2021), suggest a transitional context and degrees of episodicity, with narratives in the past (especially what Kitagawa & Lehrer (1990) call “life drama” narrations) occupying an intermediate position between generic and episodic uses. The distance from the moment of speaking possibly leads to a greater degree of vagueness and lesser specificity of episodic impersonals in the past in comparison to uses in the present. The following examples may illustrate this transition zone. In (38) the informant answers the interviewer’s question regarding whether the informant has experienced a potentially lethal situation before, although the use of negation gives this episodic use a hypothetical flavor:

- (38) Interviewer: *intendi... tirando isso você nunca passou por nenhum outro problema não?*
Participant: *não*
Interviewer: *você já presenciou algum acidente?*
Participant: *de moto assim na / nada grave nada que a pessoa saiu de lá muito grave nada não acidente comum de / de esquina esses trem assim* (MOC 13 - A.J, female, 37 years old, low educational level)
‘Interviewer: I see... other than that, you never had any other problems did you?’
Participant: *no*
Interviewer: *Have you ever witnessed an accident?*
Participant: *on a motorcycle like that in / nothing serious, nothing that the person/one came out seriously injured, nothing not, common accident on / around corners and things like that’*

In (39) *a pessoa* refers to the speaker's mother in a specific situation, but one might think alternatively about a general statement about the length of the trip at that time, so here, the episodic character is not entirely clear either, although the perfect rather suggests an episodic use (interestingly *uma pessoa* precedes *a pessoa*):

- (39) meu avô... ele... foi... destacado pro sul... e a minha mãe não podia ir porque ia formar... ia acabar... o... o normal... então ficou pra depois encontrar com eles... eles foram pra depois ela ir... QUANdo a minha mãe foi... eh... não tinha aquela... aquela... () tem hoje né... era um navio bem... bem fulerazinho... fulerazinho... então... coitadinha ela foi... sozinha... foi embora pro sul... e pegou uma tempestade em alto mar... diz ela que a água passava... então ela tinha verdadeiro pavor... não é pra menos... eu acho que *uma pessoa* ... já pensou? à noite... *a pessoa* levou muito tempo pra chegar no sul né... pegou essa tempestade em pleno... alto mar... (NURC 90 complementar)

'my grandfather... he... was... transferred to the south... and my mother couldn't go because she finished her studies... she finished... the... normal course... then she agreed to meet them later... they went so that she should go later... WHEN my mother went... eh... there wasn't that one... that one... () there is today, right... it was a ship quite...quite tiny... tiny... so... she went... alone... she went to the south... and caught a storm on the high seas... she says the water passed... so she was really terrified... no wonder... I think *a person/one* ... have you thought about it? at night... *the person/one* took a long time to reach the south, right... she caught this storm... on the high sea...'

The inclusive quasi-episodic uses do not fit into Gast & van der Auwera's characterisation of exclusive episodic uses – here, rather than showing a development towards a use as a first person pronoun, we think the type of pronominalisation source, linked to inclusive uses in generalising statements, seems to affect the exact sequence of emerging new functions and functional expansion. This is a case of persistence in the course of grammaticalisation following Hopper (1991). However, we cannot exclude an evolution of episodic inclusive uses taking over the function of first person pronouns as in (34) which, according to our research, is possibly more frequent in young people's speech nowadays. Unfortunately the Web/dialects corpus does not give us the age of the writers.

Figure 5 highlights the frequencies of inclusive and exclusive uses.¹¹ The proportion of uses in inclusive contexts decreases over time. Inclusive or speaker-oriented uses characterise the earlier uses leading to the evolution of these impersonals while exclusive uses arise later and tend to increase until the 2010s data, with a slight decrease in 2020.

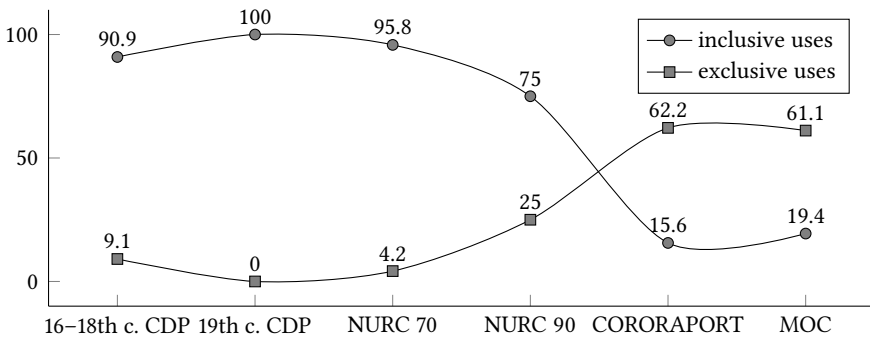


Figure 5: The functional development of *a pessoa* and *uma pessoa* (% of uses in the respective corpus) taken together in diachrony in subject position.

The reversal of frequency of *uma pessoa* to *a pessoa* in the 20th century in BP (see Figure 2) differs from the observations for EP in Posio (2021) and Martins (2022) as well as in Afonso (2008: 147) who argue in favour of a higher frequency of *uma pessoa* in EP. The fact that the impersonal expression *a pessoa* is more frequent might be a result, at least for Brazilian Portuguese, of the greater frequency of anaphoric NPs in longer chains of reference as opposed to the originally indefinite *uma pessoa*.

We have postulated above that *a pessoa* and *uma pessoa* follow the same grammaticalisation path, so we would then also expect comparable functional profiles. Since *uma pessoa* is very infrequent in our corpus data, we conducted a survey based on acceptability judgments (see §4 on the survey design). We used a Likert scale from 1 (perfectly acceptable) to 5 (unacceptable) and tested equivalents of the diagnostic sentences of Gast & van der Auwera (2013) and additional sentences for each function (see Figure 6).

We tested the statistical significance of the differences between the acceptability of the two expressions for each function and only found a significant difference for function 1 (applying the Games-Howell post hoc test¹²). Function

¹¹As pointed out above, the CDP data are based on written texts, while NURC, CORORAPORT and MOC are oral corpora. Thus a comparison of the data up to the 19th century and the 20th and 21st century has to be interpreted with great caution.

¹²Only for condition 1 was *a pessoa* rated significantly higher than *uma pessoa* ($p < 0.001$). See Appendix A for test results.

5 Portuguese *a pessoa* and *uma pessoa*: Emerging inclusive impersonals

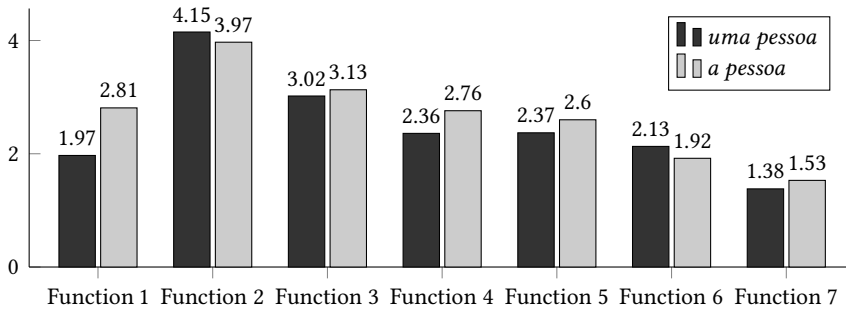


Figure 6: Degrees of acceptability of *a pessoa* and *uma pessoa* in diagnostic functions (high values indicate lower acceptability).

1 also allows an indefinite interpretation of *uma pessoa*, so the participants probably did not judge the impersonal use, despite the instruction to consider only impersonal interpretations. The near identity of both forms in all other cases is a striking confirmation of the functional equivalence in BP. The data also show an increasing acceptability from function 2 to function 7, which backs our corpus analysis, i.e. the preference for non-episodic inclusive uses. The position of function 1 is harder to explain. It is possible that the informants interpreted the sentences as having a personal/referential interpretation.

Finally, when discussing the evidence of the grammaticalisation of *a/uma pessoa*, Posio (2021: 12) mentions the issue of the repetition of the noun phrase to refer to itself, instead of using a personal pronoun. The author comments that in his data there are no cases where the personal pronoun *ela* ‘she’ is used to refer back to the NP. Contrary to what the author observes for EP, our data present occurrences with this context (40), as well as occurrences where the whole NP is repeated (41). However, unlike what the author finds for EP, according to the first author of this paper *uma pessoa* does not allow this kind of repetition in BP (although we detected one example in C-ORAL-Brasil, see Amaral & Mihatsch 2019). Therefore, an example like (41) seems to be unacceptable in BP, while (40) is acceptable. The samples thus corroborate the existence of differences between the two varieties of Portuguese, especially when analysing the frequency and the paths of grammaticalisation of *uma pessoa*.

- (40) às vezes *a pessoa* ela gosta de assistir um jornal e ela não tem... é costume de assistir um outro... (Amaral 2015)
 ‘sometimes *one* (lit. ‘the person’, feminine) she likes to watch a news programme and she is not used to watch another one.’

- (41) Mas esse peixe, já *uma pessoa* às vezes não o conhece. Não sabe de que peixe é, não é? Se *uma pessoa* visse a figura do peixe, já *uma pessoa* dizia: “Olha, pode ser a sardinha, pode ser carapau”. (CORDIAL-SIN, VPA-30)
- ‘But that fish, sometimes *a person* does not even recognise it. They don’t know what fish it is, right? If *a person* saw the form of the fish, *a person* would say: “Look, it can be a sardine, it can be a mackerel.”’ (Posio 2021: 13)

7 Conclusion

In this paper we discussed the functions of inclusive impersonal uses of *a pessoa* and *uma pessoa* in Brazilian Portuguese in order to shed light on the diachronic development of the restrictions determining impersonal uses and the differences and parallels between the two expressions.

First, it should be highlighted that the development of impersonal uses detectable in our data – as well as their synchronic properties – point to parallels in the origins of *a pessoa* and *uma pessoa*, although the grammaticalisation process then leads to a subsequent, mainly diatopical differentiation. In our Brazilian oral data *a pessoa* very clearly prevails. Their impersonal function becomes evident when trying to translate the impersonal attestations into other languages that have nouns derived from Latin *persona* such as Spanish *persona*, French *personne* or English *person*. The translation requires an impersonal pronoun such as English *one* or French *on*.

The first impersonal occurrences in the 16th and 17th centuries are scarce. The situation changes in the 19th century, when there is an increase, especially in the incidence of *uma pessoa*. In the second half of the 20th century Brazilian Portuguese, however, clearly prefers the impersonal *a pessoa*, although *uma pessoa* is also used. A classification of the occurrences from two different varieties of BP showed a similar pattern of behaviour of the two expressions.

From the 16th century to the 21st century they have shown a clear preference for inclusive, non-episodic uses up to the 21st century, although our corpus data threw up results that reveal a functional shift and a certain functional expansion, notably exclusive and episodic uses, in the course of grammaticalisation of the two expressions. Some occurrences even show a tendency toward first person pronoun uses, so it is certainly worth continuing this research with recent data from young speakers.

Appendix A Games-Howell post hoc test results

Table 4: Games-Howell post hoc test results. “H” rating indicates higher rating, “L” lower.

<i>a pessoa</i>			<i>uma pessoa</i>		
Comparison	Rating	<i>p</i>	Comparison	Rating	<i>p</i>
8:1	H	<0.001	2:1	H	<0.001
8:3	H	0.001	2:3	H	<0.001
8:4	H	<0.001	2:4	H	<0.001
8:5	H	<0.001	2:5	H	<0.001
8:6	H	<0.001	2:6	H	<0.001
8:7	H	<0.001	2:7	H	<0.001
8:9	H	<0.001	2:10	H	<0.001
8:10	H	<0.001	2:11	H	<0.001
8:11	H	0.028	8:1	H	<0.001
2:1	H	<0.001	8:4	H	<0.001
2:4	H	<0.001	8:5	H	0.001
2:5	H	0.002	8:6	H	<0.001
2:6	H	<0.001	8:7	H	<0.001
2:7	H	<0.001	8:10	H	0.022
2:9	H	<0.001	8:11	H	<0.001
2:10	H	0.005	3:1	H	0.001
6:1	H	<0.001	3:2	L	<0.001
6:2	H	<0.001	3:6	H	0.008
6:3	H	<0.001	3:7	H	<0.001
6:4	H	0.004	3:11	H	<0.001
6:8	H	<0.001	10:1	H	0.001
6:10	H	0.005	10:2	L	<0.001
6:11	H	0.003	10:6	H	0.042
7:1	L	<0.001	10:7	H	<0.001
7:2	L	<0.001	10:8	L	0.022
7:3	L	<0.001	10:11	H	0.002
7:4	L	<0.001	7:1	L	0.018
7:5	L	0.017	7:2	L	<0.001
7:8	L	<0.001	7:3	L	<0.001
7:9	L	<0.001	7:4	L	<0.001
7:10	L	<0.001	7:5	L	0.036
7:11	L	<0.001	7:6	L	0.007
			7:8	L	<0.001
			7:10	L	<0.001

Acknowledgements

We would like to thank two anonymous reviewers for their valuable suggestions, Désirée Kleineberg for her help with the statistical analysis of the judgment data, Welber Nobre dos Santos for the data collection for the MOC corpus and Ina Berner for helping to prepare the NURC data for analysis.

References

- Achard, Michel. 2015. *Impersonals and other agent defocusing constructions in French*. Amsterdam/Philadelphia: John Benjamins.
- Afonso, Susana P. C. 2008. *The family of impersonal constructions in European Portuguese: An onomasiological constructional approach*. University of Manchester. (Doctoral dissertation).
- Amaral, Eduardo Tadeu Roque. 2015. *Corpus of the project “O uso de nomes gerais nos falares mineiros”*. Belo Horizonte. <http://letras.ufmg.br/nomesgerais>.
- Amaral, Eduardo Tadeu Roque. 2017. Estudio contrastivo de nombres generales para humanos en español y en portugués. *Lingüística y literatura* 72. 54–79.
- Amaral, Eduardo Tadeu Roque & Wiltrud Mihatsch. 2016. Le nom français *personne* en comparaison avec le portugais brésilien *pessoa* et l’allemand *person*: Des noms en voie de pronominalisation. *SHS Web of Conferences* 27. DOI: 10.1051/shsconf/20162712015.
- Amaral, Eduardo Tadeu Roque & Wiltrud Mihatsch. 2019. Incipient impersonal pronouns in colloquial Brazilian Portuguese. In Peter Herbeck, Bernhard Pöll & Anne C. Wolfsgruber (eds.), *Semantic and syntactic aspects of impersonality* (Linguistische Berichte, Sonderhefte 26), 149–185. Hamburg: Buske.
- Barbosa, Afranio G., Célia Regina dos Santos Lopes & Dinah M. I. Callou. 2021. *Projeto norma linguística urbana culta (NURC)*. Rio de Janeiro. <https://nurcrj.letras.ufrj.br/>.
- Barsalou, Lawrence W. 1983. Ad hoc categories. *Memory and Cognition* 11(3). 211–227.
- Cabredo Hofherr, Patricia. 2006. “Arbitrary” pro and the theory of pro-drop. In Peter Ackema, Patrick Brandt, Maaike Schoorlemmer & Fred Weermann (eds.), *Agreement and arguments*, 230–258. Oxford: Oxford University Press.
- Cabredo Hofherr, Patricia. 2008. Les pronoms impersonnels humains: Syntaxe et interpretation. *Modèles linguistiques* 57. 35–56. DOI: 10.4000/ml.321.
- Cabredo Hofherr, Patricia. 2017. Voice and voice alternations. In Andreas Dufter & Elisabeth Stark (eds.), *Manual of Romance morphosyntax and syntax* (Manuals of Romance Linguistics Series), 230–271. Berlin: De Gruyter.

- Cappeau, Paul & Catherine Schnedecker. 2015. *(les/des) gens vs (les/des) personnes*: Évolution diachronique et comparaison diamésique. Des SN en voie de pronominalisation. In K. Jeppesen Kragh & J. Lindschouw (eds.), *Les variations diasystématiques et leurs interdépendances dans les langues romanes*, 449–463. Strasbourg: Editions de linguistique et de philologie.
- Carvalho, Janayna. 2020. Sentenças impessoais do português brasileiro. *Cuadernos de la Alfil* 12(2). 251–277.
- Company Company, Concepción & Julia Pozas Loyo. 2009. Los indefinidos compuestos y los pronombres genéricos-impersonales *omme* y *uno*. In Concepción Company Company (ed.), *Sintaxis histórica de la lengua española, Segunda parte: La frase nominal*, 1073–1222. Ciudad de México: Fondo de Cultura Económica y Universidad Autónoma de México.
- Coveney, Aidan. 2000. Vestiges of *nous* and the 1st person plural verb in informal spoken French. *Language Sciences* 22. 447–481. DOI: 10.1016/S0388-0001(00)00014-0.
- Egerland, Verner. 2003. Impersonal pronouns in Scandinavian and Romance. *Working Papers in Scandinavian Syntax* 71. 75–102.
- Elvira, Javier. 1994. *Un(o)* en español antiguo. *Verba: Anuario galego de filoloxia* 21. 167–182.
- Ferreira, Marcelo Barra & Clara Nunes Correia. 2016. The semantics of DPs. In W. Leo Wetzels, João Costa & Sergio Menuzzi (eds.), *The handbook of Portuguese linguistics*, 1st edn., 356–373. Hoboken, NJ: John Wiley.
- Gast, Volker & Johan van der Auwera. 2013. Towards a distributional typology of human impersonal pronouns, based on data from European languages. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries: Studies in the memory of Anna Siewierska*, 119–158. Berlin: De Gruyter.
- Giacalone Ramat, Anna & Andrea Sansò. 2007. The spread and decline of indefinite man-constructions in European languages: An areal perspective. In Paolo Ramat & Elisa Roma (eds.), *Europe and the Mediterranean as linguistic areas: Convergences from a historical and typological perspective*, 95–131. Amsterdam: John Benjamins.
- Giacalone Ramat, Anna & Andrea Sansò. 2011. L’emploi indéfini de *homo* en latin tardif: Aux origines d’un européanisme. In Michèle Fruyt & Olga Spevak (eds.), *La quantification en latin*, 93–115. Paris: Harmattan.
- Givón, Talmy. 1978. Definiteness and referentiality. In J. Greenberg (ed.), *Universals of human language*, vol. 4: Syntax, 291–330. Stanford University Press.
- Givón, Talmy. 1981. On the development of one as an indefinite marker. *Folia Linguistica Historica* 11. 35–53.

- Haspelmath, Martin. 2001. The European linguistic area: Standard Average European. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals/Sprachtypologie und sprachliche Universalien/La Typologie des langues et les universaux linguistiques: An international handbook/Ein internationales Handbuch/Manuel international*, 1492–1510. Berlin: Walter de Gruyter.
- Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Michael Tomasello (ed.), *The new psychology of language*, vol. 2, 211–242. Mahwah, NJ: Lawrence Erlbaum.
- Hopper, Paul J. 1991. On some principles of grammaticization. In Elizabeth C. Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, vol. 1, 17–36. Amsterdam: John Benjamins. DOI: 10.1075/tsl.19.1.04hop.
- Kato, Mary A. 2000. The partial pro-drop nature and the restricted VS order in Brazilian Portuguese. In Mary A. Kato & Esmeralda V. Negrão (eds.), *Brazilian Portuguese and the null subject parameter*, 223–258. Frankfurt am Main, Madrid: Vervuert Verlagsgesellschaft.
- Kato, Mary A. & Esmeralda V. Negrão (eds.). 2000. *Brazilian Portuguese and the null subject parameter*. Frankfurt am Main & Madrid: Vervuert. DOI: 10.31819/9783964561497.
- Kitagawa, Chisato & Adrienne Lehrer. 1990. Impersonal uses of personal pronouns. *Journal of Pragmatics* 14. 739–759.
- Kluge, Bettina. 2016. Generic uses of the second person singular: How speakers deal with referential ambiguity and misunderstandings. *Pragmatics* 26(3). 501–521. DOI: 10.1075/prag.26.3.07klu.
- Koenig, Jean-Pierre & Gail Mauner. 1999. A-definites and the semantics of implicit arguments. *Journal of Semantics* 16. 207–236.
- Krifka, Manfred & Claudia Gerstner-Link. 1993. Genericity. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (eds.), *Syntax: An international handbook of contemporary research*, 966–978. Berlin: Walter de Gruyter.
- Krifka, Manfred, Francis Jeffrey Pelletier, Gregory N. Carlson, Alice ter Meulen, Gennaro Chierchia & Godehard Link. 1995. Genericity: An introduction. In Greg N. Carlson & Francis J. Pelletier (eds.), *The generic book*, 1–124. Chicago, London: The University of Chicago Press.
- Lamoglia Duarte, Maria Eugenia. 2000. The loss of the “avoid pronoun” principle in Brazilian Portuguese. In Mary A. Kato & Esmeralda V. Negrão (eds.), *Brazilian Portuguese and the null subject parameter*, 17–36. Frankfurt am Main, Madrid: Vervuert Verlagsgesellschaft. DOI: 10.31819/9783964561497-002.

5 *Portuguese a pessoa and uma pessoa: Emerging inclusive impersonals*

- Lopes, Célia Regina dos Santos. 2003. *A inserção de a gente no quadro pronominal do português: Percurso histórico*. Frankfurt am Main/Madrid: Vervuert/Iberoamericana. DOI: 10.31819/9783865278494.
- Lopes, Célia Regina dos Santos. 2004. A gramaticalização de a gente em português em tempo real de longa e de curta duração: Retenção e mudança na especificação dos traços intrínsecos. *Fórum Linguístico* 4(1). 47–80.
- Martins, Ana Maria. 2022. Autorreferência generalizadora com a pessoa, uma pessoa e outros nomes humanos referencialmente vagos. *Revista de Estudos Linguísticos da Universidade do Porto* 1. 109–142. DOI: 10.21747/16466195/ling2022v1a5.
- Mauri, Caterina. 2017. Building and interpreting ad hoc categories: A linguistic analysis. In Joanna Blochowiak, Cristina Grisot, Stephanie Durrleman & Christopher Laenzlinger (eds.), *Formal models in the study of language*, 297–326. Cham: Springer. DOI: 10.1007/978-3-319-48832-5_16.
- Mihatsch, Wiltrud. 2017. Les noms d’humains généraux aux limites de la grammaticalisation. *Syntaxe et Sémantique* 18(1). 67–99.
- Moltmann, Friederike. 2010. Generalizing detached self-reference and the semantics of generic *one*. *Mind and Language* 25. 440–473.
- Nunes, José Joaquim. 1919. *Compêndio de gramática histórica portuguesa: Fonética e Morfologia*. Lisboa: Livraria Clássica.
- Ornelas de Avelar, Juanito. 2023. On pronominal uses of *geral* in Brazilian Portuguese. In Pekka Posio & Peter Herbeck (eds.), *Referring to discourse participants in Ibero-Romance languages*, 177–202. Berlin: Language Science Press. DOI: 10.5281/zenodo.8124498.
- Posio, Pekka. 2012. Who are ‘we’ in spoken Peninsular Spanish and European Portuguese? Expression and reference of first person plural subject pronouns. *Language Sciences* 34(3). 339–360. DOI: 10.1016/j.langsci.2012.02.001.
- Posio, Pekka. 2017. Entre lo impersonal y lo individual: Estrategias de impersonalización individualizadoras en el español y portugués europeos. *Spanish in Context* 14(2). 209–229. DOI: 10.1075/sic.14.2.03pos.
- Posio, Pekka. 2021. *A pessoa and uma pessoa: Grammaticalization and functions of a human impersonal referential device in European Portuguese*. *Journal of Portuguese Linguistics* 20(1). DOI: 10.5334/jpl.254.
- Santos, Welber N. 2021. *A concordância nominal de número no português falado em Montes Claros, MG: A influência dos nomes gerais em uma análise variacionista*. Belo Horizonte: Federal University of Minas Gerais. (MA thesis).
- Siewierska, Anna. 2011. Overlap and complementarity in reference impersonals: Man-constructions vs. third person plural-impersonals in the languages of Europe. In Andrej L. Malchukov & Anna Siewierska (eds.), *Impersonal construc-*

tions: A cross-linguistic perspective, 57–89. Amsterdam: John Benjamins. DOI: 10.1075/slcs.124.03sie.

Siewierska, Anna & Maria Papastathi. 2011. Towards a typology of third person plural impersonals. *Linguistics* 49(3). 575–610.

Silva, Luis Filipe L. & Sueli M. Coelho. 2020. A gramaticalização de *geral* no português brasileiro: De adjetivo a pronome indefinido/quantificador. *Revista (Con)Textos Linguísticos* 14(28). 159–176.

Chapter 6

Personal uses of impersonalizing strategies: Hybrid constructions with *a gente* and *se* in rural Madeiran Portuguese varieties

Yoselin Henriques Pestana

Universität Zürich

The varieties of Portuguese spoken in Madeira present a predominant use of *a gente*, a grammaticalized first person plural pronoun, derived from the noun phrase ‘the people’, instead of the traditional pronoun *nós*. They also exhibit constructions where *a gente* cooccurs with the impersonal clitic *se*. In a pioneering study, Martins (2009) provides a detailed description of what she calls “double subject impersonal constructions” and proposes that *a gente* restricts the generic interpretation of the clitic *se*. Based on spoken data from semi-directed interviews and free-speech conversations with elderly speakers of rural Madeiran Portuguese, this chapter provides a quantitative and qualitative approach to the $[(a\ gente) + se]$ construction. The goal of this study is twofold. First, a depiction of the broad referential range of this hybrid structure is presented. Its possible interpretations cover a scope similar to that of first person plural pronouns reaching from indefinite readings to deictic ones (referring to participants of the speech act). Second, a description of the syntactic features of this innovative construction will show that the element *se* is being reanalyzed as a dependent person marker in rural Madeiran Portuguese varieties.

1 Introduction

In some Portuguese varieties, the traditional first person plural (henceforth 1PL) pronoun *nós*, illustrated in example (1), coexists with a newer 1PL pronoun *a gente*, as shown in example (2). The latter originates from a noun phrase consisting of



the definite feminine article *a* ('the') and the generic noun *gente* ('people'). By losing its nominal properties, this noun phrase gave rise to an indefinite pronoun *a gente* displaying generic readings (Lopes 2007, 2003, 1999, de Omena 2003, among others). This grammaticalization process first resulted in a referential shift from speaker-exclusive readings to speaker-inclusive ones. From the 19th century onward, *a gente* became a new 1PL pronoun with specific interpretation and has even replaced the pronoun *nós* in some Portuguese varieties. The nominal origin and gradual grammaticalization of *a gente* led to a mismatch between its semantic and syntactic properties resulting in mixed agreement patterns (cf. §3).¹

- (1) Nós, nessa altura, não tínhamos luz.
PRON.SBJ.1PL in-that time NEG have-IPFV.1PL electricity
'Back then, we did not have electricity.'
- (2) Sabe onde é que a gente vai dar com ele?
know.PRS.3SG where is that PRON.SBJ.1PL go.PRS.3SG give-INF with him
'Do you know where we find him?'

In rural Madeiran Portuguese (hereafter MP) varieties, the 1PL pronoun *a gente* seems to have largely replaced the canonical *nós*. However, there is another variant to these pronominal 1PL expressions illustrated in (3). In these constructions, the 1PL pronoun *a gente* cooccurs with what seems to be an impersonal *se* (henceforth *se-IMP*²), found in most Romance languages.

- (3) A gente contava-se os dias [...]
PRON.SBJ.1PL count-IPFV.3SG=SE the days
'We counted the days [...]

The element *se* appears cliticized to the verb, whereas *a gente* is identifiable as the subject. I will call it "hybrid construction" due to its nature of combining a personal pronoun with an impersonal marker. The qualitative analysis shows that these constructions present a referential scope reaching from indefinite to deictic readings which is thus congruous with the range of interpretations of the other 1PL pronouns *nós* and *a gente* available in the varieties under study.³

¹The glossing of the language examples follows the Leipzig Glossing Rules.

²The label *se-IMP* is used in this chapter as an umbrella term for impersonal and passive *se* constructions. Note that all examples found in the corpus are instances of non-agreeing and thus impersonal *se* constructions.

³The fact that the pronoun *a gente* is still frowned upon in the context of school education contributes to the existence of discrepancies on the stylistic level. Thus, the pronoun *nós* is

So far, very little attention has been paid to these constructions, except for the research by Martins (2003, 2005, 2009). Based on the assumption that the two elements share subject features, the author refers to them as “impersonal subject doubling constructions” (Martins 2009). However, the data analyzed in the present study bring to fore other aspects of the referential properties of these constructions. Thus, they provide evidence that these constructions may display interpretations that go far beyond the speaker-inclusive impersonality thoroughly described in Martins (2009), as illustrated in the example (4).

- (4) Amanhã a gente vai-se limpar o escritório.
 Tomorrow PRON.SBJ.1PL go-PRS.3SG=SE clean-INF the office
 ‘Tomorrow, we will clean the office.’

In light of these new insights, this chapter aims to review the referential and syntactic properties of the hybrid construction. Its particular properties are derived under the hypothesis that *se* might be reanalyzed as a person-marking item associated with the 1PL pronoun *a gente* in rural MP varieties.

The present study is structured as follows: §2 describes the data under survey. §3 presents a brief overview of the variation of 1PL pronominal subject expressions in Portuguese varieties. §4 describes the referential range of the construction under focus and its constituting elements in Portuguese. §5 analyzes some of the syntactic features the hybrid construction [(*a gente*)+*se*] displays. Lastly, §6 provides some conclusions along with observations for future research.

2 Data and methodology

Portuguese dialectology grapples with a scarcity of corpora and data that impedes a detailed and thorough analysis of the numerous morphosyntactic phenomena still little-known to linguists. So far, dialectal studies on Madeiran varieties have used either the dialectal corpus *CORDIAL-SIN* (e.g., Martins 2021, among others) or the *Corpus de Concordância* (e.g., Bazenga 2019, 2015) as their primary data sources. However, apart from presenting only few excerpts of spontaneous speech of the dialect under analysis, the *CORDIAL-SIN* raises other problems, such as the communicative asymmetry between interviewer – a speaker of the standard Portuguese variety – and a local informant, which results in

perceived as corresponding to more formal contexts by speakers with higher educational levels. A similar tendency is found in French in the use of *nous* as opposed to the newer 1PL pronoun *on* (Coveney 2000). Future research on MP based on corpora displaying different degrees of formality will allow to confirm or rule out this trend.

auto-correction and neutralization of certain phenomena. Furthermore, the thematic domains addressed in the spontaneous and semi-directed speech samples of the available corpora – i.e., local traditions and customs – do not enable a broad referential scope such as the one displayed by the hybrid construction under discussion. Therefore, we compiled a corpus composed of two sub-samples – semi-directed sociolinguistic interviews and free conversation samples – to tackle these shortcomings.⁴

The sociolinguistic profiles of the speakers chosen as informants roughly meet the standards introduced by Chambers & Trudgill (1980), also known by the acronym NORM (“non-mobile, older, rural, males”), except for their gender.⁵ The informants have a low level of education ranging from zero to four years of primary school and represent an age scale ranging from 54 to 84.

Part of the data stems from 13 semi-directed interviews in different rural sites across Madeira Island.⁶ To increase the occurrence of constructions displaying the broadest possible referential range, our data include classic dialectological interview questions and questions concerning the social environment and private lives of the informants. The latter category proved to be crucial for triggering more specific uses of the hybrid construction under analysis. Furthermore, although most of the interviews were conducted with one interviewer and one informant, the inclusion of several participants produced occurrences of completely deictic expressions, i.e., constructions that refer exclusively to the speech-act participants.

⁴The corpus analyzed in this study is, in comparison to the available corpora, in many respects broader in coverage. Firstly, it includes speech samples from different rural sites of Madeira Island, whereas the *Corpus the Concordância* focuses exclusively on the variety spoken in the capital city of the island, Funchal. The *CORDIAL-SIN* includes data from four localities, two of which are situated on the archipelago’s main island: Câmara de Lobos and Caniçal. Secondly, the sub-sample of semi-directed interviews contains, in addition to questions used in classic dialectological interviews, topics which allude to the personal and family lives of the informants. The inclusion of such topics enables the mention of more delimited groups. Thirdly, the sub-sample of free conversations between family members, neighbors or friends not only promotes the occurrence of 1PL expressions with deictic interpretations, but also presents a highly natural communicative environment. Moreover, the fact that the interviewer is perceived by the informants as a member of their speech community, also contributes to counteracting the asymmetrical situation of sociolinguistic interviews.

⁵Our corpus consists of language produced predominantly by female speakers. This is due to the extremely limited number of local male speakers of the target age group who have not emigrated for an extended period.

⁶The interviews were conducted in Estreito de Câmara de Lobos, Câmara de Lobos, Curral das Freiras, Maroços, Canhas, Santa, Tabúa, Campanário, Camacha (Santa Cruz), Santo António (Santana), and São Vicente.

In addition to the 13 semi-directed interviews, the second part of our corpus contains free conversation samples. These latter samples complement the above interviews in two different aspects. Although the interviewer was present during the recording, by taking place in a familial context, the free conversations yield “immediate” speech samples (following Koch & Oesterreicher 1990). These circumstances are invaluable for the study of morphosyntactic variation. Secondly, since the discourse participants are familiar with each other and therefore share social networks, the free conversation samples include several specific uses of the hybrid construction as the informants often produce utterances that refer to particular groups to which they belong. This type of use is less common in semi-directed interviews in which speech act participants do not know each other. Conversely, informants of semi-directed interviews provide more clues so that the addressee correctly infers the intended 1PL reference, which is a complex and challenging task for linguistic analysis.

In total, the data under survey contain 827 examples of 1PL subjects. For the analysis, all occurrences were coded manually for an array of grammatical and referential properties, including expression or omission of the subject pronoun, verbal agreement, coreference, and referential range.

3 First person plural (1PL) in Portuguese varieties

This section provides a brief overview of the 1PL pronominal subject expressions *a gente* and *nós* in Portuguese. It then describes the differences in the usage of *a gente* and *nós* in rural MP varieties.

Due to its nominal origins described in the introduction to this chapter, *a gente* presents discrepancies between its semantic and syntactic properties, leading to varying agreement patterns in terms of verbal inflection, as can be observed in the contrast between examples (5) and (6). Likewise, *a gente* triggers varying adjectival (or participial) agreement as shown in example (7) (cf. Pereira 2003, Costa & Pereira 2013, 2005).

- (5) A gente fizemos uma fogueira.
 PRON.SBJ.1PL make-PST.1PL a bonfire
 ‘We made a bonfire.’
- (6) A gente ia lavar a roupa aqui.
 PRON.SBJ.1PL go.IPFV.3SG wash-INF the clothes here
 ‘We came here to wash our clothes.’

- (7) A gente era pequenos.
PRON.SBJ.1PL be.IPFV.3SG small-MASC.PL
'We were little.'

A considerable number of studies show variation between *a gente* and *nós* in different varieties of Brazilian Portuguese (BP) (cf. Vianna 2011 for an overview). It is observable that the newer pronoun *a gente* "is the more productive of the forms" (Travis & Silveira 2009: 22) and thus, appears to replace *nós* progressively. This canonical pronoun and its associated 1PL verbal marking – the desinence *-mos* illustrated in example (8) – are subject to a restricted distribution. Travis & Silveira (2009) observe the retention of these morphological forms in high-frequency verbs such as *ter* ('to have') or *ser* ('to be') and in cohortative constructions⁷ illustrated in (9). According to the authors, these are some of the few domains to which *a gente* has not extended.

- (8) onde nós vivíamos
where PRON.SBJ.1PL live-IPFV.1PL
'where we lived'
- (9) Então, vamos à minha casa!
so go.PRS.1PL to-the my house
'So, let's go to my place!'

Limited studies have addressed the variation of 1PL expression in European Portuguese (EP) varieties. Contrary to the traditional belief that the newer form *a gente* is commonly found in central and southern EP varieties, recent studies have shown that this pronoun exhibits a high usage rate throughout continental EP varieties. For instance, while contrasting the use of the two 1PL pronominal variants in Brazilian and European varieties, de Paiva Sória (2013) found that the pronominal expression *a gente* is highly productive in most of the 31 local EP varieties accounted for. The author concludes that this pronoun is not only firmly established but also most commonly used throughout the observed EP varieties (de Paiva Sória 2013). Similarly, Posio (2012) observes a relatively high application rate of *a gente* in his contrastive study on 1PL subject expression in EP and Peninsular Spanish. Regarding the referential scope of *a gente*, the author acknowledges its speaker-inclusive impersonal traits attributing this fact to its impersonal origin.

⁷There are numerous denominations for these constructions. Travis & Silveira (2009), for instance, use the label "hortative constructions". Others use a more transparent terminology such as "inclusive imperative" (Dobrushina & Goussev 2005). Following Posio (2012) the term "cohortatives" will be used in this chapter to refer to these constructions.

While the construction with *a gente* was included in the current study as a way to create first person plural reference, examining the use of *a gente* in context reveals that in the EP data it is very seldom used in contexts where only a personal (i.e. inclusive or exclusive) reference is possible. In most cases, the referential range of *a gente* can be described as speaker-inclusive impersonal or allowing both impersonal and first person plural interpretations. (Posio 2012: 348)

As for the varieties under survey, the data suggest that the newer pronoun *a gente*, found in all but three examples, has largely replaced the canonical subject pronoun *nós*. Given this markedly high use rate of *a gente* in the data of rural MP varieties, it can be hypothesized that this newer 1PL pronoun is more grammaticalized in some EP varieties than previously believed (cf. Posio 2012) and thus may occur in less restricted referential contexts. As expected, the data show divergent agreement patterns triggered by *a gente*, with a clear predominance of 3SG verb forms. Table 1 summarizes the quantitative analysis concerning the subject-verb agreement involving 1PL subject pronoun variants.⁸

Table 1: Subject and verb agreement with 1PL pronouns (*a gente* and *nós*) in rural MP varieties

	3SG		1PL	
	<i>n</i>	%	<i>n</i>	%
∅	5	9.1	50	90.9
<i>a gente</i>	160	87.9	22	12.1
<i>nós</i>	0	0	3	100

As shown in Table 1, the data under survey contain 75 examples of 1PL verb forms, the majority of which are found in clauses lacking an overt subject. The low occurrence of 3SG verbs without an expressed subject is due to the ambiguity of this form, referring either to a 3SG or to a 1PL subject associated with the pronoun *a gente*. Therefore, ambiguous examples such as (10) were excluded from the analysis.

⁸Although the hybrid construction [(*a gente*)+*se*] is considered a variant of the subject pronouns *a gente* and *nós*, it was excluded from the quantitative analysis resumed in Table 1. Moreover, 21 tokens of non-finite verbs occurring with the pronoun *a gente* were excluded, due to their well-known distinctive person marking behavior.

- (10) A gente_i vê ele_j quando passa_{i,j} por ali.
 PRON.SBJ.1PL see.PRS.3SG to my pass-PRS.3SG through there
 ‘We see him when (he/we) pass(es) by.’

The only occurrences of verb agreement in 3SG lacking an overt subject considered in the analysis are those displaying coreference with a preceding 1PL pronoun.⁹ Example (11) illustrates the coreference between a 3SG verb form and *a gente*.

- (11) Mas a gente não deitava aquela [carne] fora;
 But PRON.SBJ.1PL NEG throw-IPFV.3SG that [meat] away
 ‘But we did not throw that meat away;’
 tirava um bocadinho e cozia.
 take-IPFV.3SG a bit and cook-IPFV.3SG
 ‘(we) took out a little bit and (we) cooked (it).’
 Depois partia um bocadinho a cada um.
 Afterwards cut-IPFV.3SG a bit for each one
 ‘Afterwards (we) cut a little bit for each one.’

Interestingly, despite not displaying an overt subject pronoun *a gente* in the immediate co-text, example (12) is not ambiguous. What establishes the 1PL reference of the second verb form *tinha* is its coreferentiality with the element *se* cliticized to the first 3SG verb *tinha-se*. Furthermore, the coordination with the 1PL form *temos* also indicates coreference between the two preceding null subject verb forms. Based on their coreference, these three verb forms, including the first verbal form bearing the element *se* (*tinha-se*) in example (12) are variants of 1PL expression lacking an overt subject.

- (12) Tinha-se uma fonte; tinha e temos.
 have-IPFV.3SG=SE a fountain have-IPFV.3.SG and have-PRS.1PL
 ‘(We) used to have a fountain, (we) still have (one).’

As discussed in the introduction to this chapter, MP varieties display – in addition to *a gente* and *nós* – what we call a hybrid construction. This 1PL expression

⁹There is an ongoing discussion on the existence of a null subject associated with *a gente*. For instance, Pereira (2003) and Martins & Nunes (2021) state that its mixed verbal agreement patterns hinder the existence of a null subject associated with this pronoun. De Paiva Sória (2013), in the same vein, argues that the omissions of *a gente* cannot be considered proper cases of null subjects due to the fact that it is only permissible in restricted syntactic contexts in which there is an overt *a gente* in the immediate discourse.

consists of the newer pronoun *a gente* and the originally impersonal marker *se*. In this paper, we argue that *a gente* takes the role of a 1PL subject, which is omissible in these Portuguese varieties.¹⁰ The clitic *se* associated with this pronoun is reanalyzed as a dependent person form encoding 1PL marking. As a result, the omission of *a gente* is more permissible in the varieties under focus than in continental Portuguese varieties (cf. Pereira 2003, Posio 2012), for the clitic *se* manages to disambiguate 3SG verb forms and establish 1PL specific reference, as illustrated in example (12) above and in (13).

- (13) A: As mulheres trabalhavam na fazenda ou bordavam?
 the women work-IPFV.3.PL in-the field or sew-IPFV.3PL
 ‘Did the women work in the fields or sew?’
- B: Bordava-se.
 sew-IPFV.3SG=SE
 ‘(We) sewed.’

In terms of frequency, the data indicate that within the different forms of 1PL expression – e.g., pronominal forms *a gente* and *nós* as opposed to the hybrid construction [(*a gente*) + *se*] – the latter is far more frequent than the former, as the results in Table 2 show.

Table 2: First person plural expression in rural MP varieties

	3SG		1PL	
	<i>n</i>	%	<i>n</i>	%
[(<i>a gente</i>) + <i>se</i>]	566	98.95	6	1.05
<i>a gente</i> / <i>nós</i> ^a	160	86.49	25	13.51

^aThe pronoun *nós* only appears three times in the data under study. All of these occurrences occur with 1PL verb forms.

Furthermore, the low rate of 1PL verbal forms and the virtual substitution of the canonical pronoun form *nós* are likely to be symptomatic of a possible ongoing realignment of the pronominal and verbal paradigm in the insular Portuguese variety under focus.

¹⁰Due to the possible omission of *a gente* the hybrid construction under analysis will be represented as “[(*a gente*) + *se*]” in this chapter.

Although these hybrid constructions have received little attention in research on morphosyntactic variation so far, Martins (2009) provided a seminal study in this area. Following a generative framework, Martins (2009: 179) argues that these structures fall under the denomination of “double subject impersonal *se* construction”, as a strong subject pronoun or determinant phrase (DP) appears with what the author considers to be an “impersonal subject” (*se*). According to Martins, this construction is present in both insular and continental Portuguese dialects, particularly in the center-south region of continental Portugal (Martins 2009: 180, fn. 2). In the Portuguese varieties taken into account by the author, *se* can cooccur not only with 1PL pronouns (*nós* and *a gente*) but also with 3PL pronouns and “less commonly with full determinant phrases” (Martins 2009: 179). The interpretation of these constructions depends on the semantics of the subject expression, in the sense that the “doubling strong pronoun” semantically restricts the denotative scope of *se*. Thus, an inclusive interpretation – including the speaker in a non-specific group of humans – arises when the strong pronoun or DP is 1PL. In contrast, 3PL strong pronouns in combination with *se* usually trigger an exclusive interpretation, excluding the speaker from the referent group.

In contrast to these findings, our corpus shows a significant number of other uses that are not included in Martins (2009), i.e., denoting specific sets of referents. In a footnote, the author states that only two cases of a specific reading were attested and that it “appears to be infrequent” (Martins 2009: 186, fn. 10). The lack of such data in Martins (2009) is likely to be a corpus effect. The *CORDIAL-SIN* includes classic dialectological interviews and short free speech samples. The topics addressed therein cover a thematic range of aspects of the language community’s cultural life (e.g., traditions, customs, fishing and farming practices, etc.), thus favoring the mention of unspecific groups. The fact that our semi-directed interviews also include questions on the informant’s personal life contributes to the allusion to specific referents and specific groups. Our data thus suggest that the analyzed construction manifests a broader referential scope than previously assumed. This wide scope of possible interpretations ranging from impersonal to personal can be observed in examples (14) and (15) respectively:

(14) A gente diz-se assim: uma traçada.

PRON.SBJ.1PL say-PRES.3SG=SE so: a bundle

‘We say it like this: a bundle.’

(15) Mãe, o que é que a gente vai-se fazer?

mother what PRON.SBJ.1PL go-PRS.3SG=SE do-INF

‘Mom, what are we going to do?’

While (14) is consistent with the impersonal and speaker-inclusive interpretative effects described by Martins (2009: 186–188), (15) illustrates a personal interpretation of the construction due to its use in a directive speech act addressed to the interlocutor. The readings conveyed by [(*a gente*)+*se*] are further discussed in the following section.

4 Referential scope

This section provides an overview of the referential scope of the hybrid construction and its constituents, namely the pronoun *a gente* and the clitic *se*. This detailed description highlights the possible reference overlaps of both elements, which have made possible the conjoint construction found in rural MP varieties. As mentioned, I consider [(*a gente*)+*se*] to be a variant of pronominal 1PL expressions in the varieties under focus. As such, it shows a complex reference, which Posio considers as being able to “include any human beings from the addressee to a third person or persons, an institution, or even the whole humankind” (2012: 342).

The first part of this section will deal with the referential range of *se*-IMP constructions in Portuguese varieties. The second part is dedicated to the referential aspects of *a gente*. Finally, the last section uses the observations of the first two sections to phrase possible interpretations of the hybrid construction under analysis.

4.1 *se*-IMP and its referential properties

In most Romance languages, *se*-IMP constructions are a common agent-defocusing strategy in which the reference of the agent is interpreted as unspecific and human. Due to its properties of conveying a reduction in referentiality regarding the intended subject, recent studies on impersonalizing strategies have referred to these as “R-impersonals” (Siewierska 2011). In terms of its formal characteristics, the clitic *se* attaches to a verb in third person, singular or plural. With transitive verbs, a plural NP bearing the semantic role of patient can trigger plural agreement with the verb, thus manifesting both object and subject properties.¹¹

¹¹The patient NP is typically placed after the verb. Preverbal patient NPs, display topic status, and thus appear in canonical subject position. However, both preverbal as well as postverbal plural patient NPs may trigger agreement on the verb. Posio & Vilkkuna (2013: 187) state that the postverbal position is most commonly found in their dialectal data.

Following this possible agreement, traditional Portuguese grammars usually distinguish between two formally different *se* constructions: an agreeing construction often referred to as “passive”, and a non-agreeing one often referred to as “impersonal” (Naro 1976). Cinque (1988), who provides a detailed seminal description for *si* constructions in Italian, proposes that depending on its agreement, the intended subjects have different interpretations: “quasi-existential” in agreeing constructions and “quasi-universal” in non-agreeing structures. While observing the same types of constructions in Portuguese, Raposo & Uriagereka (1996: 750) adopt the labels “indefinite SE construction” and “generic SE construction” respectively.

The grammatical status of *se-IMP* has been prone to polemic in linguistic studies. Some authors have considered *se* to display subject properties (cf. Martins 2009, 2005, 2003, Raposo & Uriagereka 1996). Others have highlighted its functionality in discourse and considered it a grammaticalized impersonality marker (Posio & Vilkkuna 2013).

As far as the referential properties of impersonal *se* are concerned, few studies on Portuguese have dealt with it extensively. For instance, Naro (1976) observes that in standard EP, *se* might incidentally include the speaker in its referential scope. Regarding dialectal Portuguese varieties, Posio & Vilkkuna (2013) find that the default readings of *se-IMP* tend to be speaker-inclusive impersonal. They might even alternate with the 1PL pronouns *a gente* and *nós* in impersonal contexts¹², according to the speaker-inclusive semantic properties they share. A description that ascribes a more specific reference property to the element *se* can be found in Casteleiro (1975). Considering 1PL expression items in nonstandard varieties of continental Portuguese, the author acknowledges that, apart from *a gente*, *se* frequently alludes to 1PL referents (Casteleiro 1975: 65). However, the examples proposed to support this idea do not present enough context to univocally infer a specific referent. Hence, according to these observations on the semantic properties of the element *se*, there seems to be a consensus that – per its agent demotion properties – *se* manifests an overall indefinite interpretation that incidentally may include the speaker in its scope.

These results regarding the predominance of speaker-inclusive readings of impersonal *se* are consistent with the data under survey in this study. All of the 258 *se-IMP* constructions found in the corpus have speaker-inclusive readings. Furthermore, the impersonal interpretation is blocked in episodic clauses (e.g., fea-

¹²Given the fact that the examples used by Posio & Vilkkuna (2013: 211–213) to illustrate this alternation stem exclusively from Madeiran informants, one might be tempted to hypothesize that this alternation is a possible hybrid construction used in coreferential contexts, without the expressed subject pronoun *a gente*.

turing predicates anchored in time). The fact that the only verbs found in *se-IMP* constructions are imperfect or present tense underlines this aspect. The cooccurrence with perfective predicates would trigger a specific reading anaphorically associated to a 1PL referent, as illustrates the difference between examples (16) and (17).

- (16) Deita-se sal na carne.
 put-PRES.3SG=SE salt in-the meat
 ‘One puts salt on the meat.’
- (17) Teve-se uma viagem maravilhosa.
 have-PST.3SG=SE a trip wonderful
 ‘(We) had a wonderful trip.’

The preferably speaker-inclusive interpretations of *se-IMP* found in previous studies, in addition to the rural MP data analyzed in this study, might facilitate the specific, and even deictic, readings displayed by the hybrid construction under focus.

4.2 The pronoun *a gente* and 1PL reference

The nominal origin of *a gente* (Lopes 2003, 1999) has resulted in a mismatch between semantic and syntactic properties, a phenomenon well studied in previous research (cf. for EP, Costa & Pereira 2013, 2005). Thus, the deviation between notional person (1PL) and the grammatical person (3SG) often incentivizes debates about the pronominal status of *a gente* (e.g., Taylor 2009). While its pronominal properties are still subject to ongoing discussion, there seems to be consensus on its referential properties ranging from impersonal readings to personal ones. As mentioned in §3, previous studies show that the variation between *nós* and *a gente* is present in both BP and EP varieties. However, it has been shown that EP varieties display lower usage rates of *a gente* than BP varieties, where there is a notorious expansion of *a gente* into more formal discursive contexts (Callou & Lopes 2004). Along these lines, the research on *a gente* indicates that increasing application rates correlate with increasing grammaticalization (cf. de Omena 2003). These empirical studies show that *a gente* can be used as a subject pronoun with speaker-inclusive impersonal reference and specific reference alike, despite the generic origins of *a gente* (cf. de Omena 2003, Travis & Silveira 2009).

The data under analysis here suggest a high usage rate of the pronoun *a gente* (cf. Table 1), which seems to have almost entirely replaced the canonical pronoun

nós in rural MP varieties. These findings might indicate a grammaticalized pronoun state comparable to the one found in BP varieties. Moreover, its referential scope is compatible with other 1PL pronominal expressions, ranging from an unspecified group of persons (or people in general) including the speaker, to purely deictic uses, referring to speech-act participants.¹³

Considering that the speaker-inclusive impersonal reference constitutes the common denominator of both 1PL pronouns and *se*-IMP, this intersection may have been the basis of the junction of *a gente* and *se* in the varieties under focus.

4.3 The referential scope of the hybrid construction [(*a gente*) + *se*]

The reference of the hybrid constructions are generally consistent with the range of possible references displayed by 1PL pronouns. Not only does [(*a gente*)+*se*] display the impersonal speaker-inclusive references described by Martins (2009: 186–188), it also shows purely deictic interpretations referring to speech-act participants. The contrast between examples (18) and (19) spans the wide variety of possible interpretations.

- (18) Pega-me às costas! A gente chega-se
pick-IMP.2SG=ACC.1SG to-the back PRON.SBJ.1PL arrive-PRES.3SG=SE
lá num instante.
there in-an instant
'Pick me up on your back! We'll get there in no time.'
- (19) A gente aqui chama-se abóbora moira.
PRON.SBJ.1PL here call-PRES.3SG=SE pumpkin <word>
'We call it moira-pumpkin here.'

The deictic reading of the hybrid construction in (18) is connected with the imperative – which refers to the addressee – and the accusative clitic (*-me*) in the first sentence. The group referred to (i.e., 2SG imperative *pega* and 1SG clitic *-me*) is construed simultaneously as the sentence is uttered, which leads to the deictic interpretation of [(*a gente*)+*se*] in the following sentence. Contrary to that, (19) shows an impersonal interpretation, referring to the speaker community in general.

¹³Following Posio (2012: 342), this flexibility of reference “is what makes possible also the use of first person plural as an impersonalizing strategy”.

In between these two poles of the referential continuum – speaker-inclusive impersonality and hearer-inclusive reference¹⁴ – there are several intermediate interpretations, highly dependent on various co-textual aspects, to which I will return later on in this chapter.

The results of the quantitative analysis (cf. §5.1) show that the majority of the constructions being analyzed lack an overt subject. Interestingly, the omission of *a gente* materializes in a construction in which *se* can establish personal interpretation, as illustrated in the affirmative verbal response¹⁵ to the question in (20).

- (20) A: Vocês já estão em casa?
 PRON.SBJ.2PL already be-PRES.3PL at home?
 ‘Are you already home?’
 B: Está-se.
 be-PRES.3SG=SE
 ‘Yes, we are.’

The specific interpretation in (20) is based on the contrast between the 2PL pronoun *vocês* and the verb *estar* bearing the element *se*. The speaker’s utilization of this construction as an affirmative verbal answer (Martins 2013, 2016b) further supports the hypothesis of *se* functioning as a 1PL person marker.

In direct comparison, these *se* interpretations differ strikingly from those of *se-IMPS* described in §4.1. The fact that an impersonal interpretation of *se* is impossible in contexts where it refers to speech-act participants seems to confirm the hypothesis regarding the reanalysis of *se*. This reanalysis becomes even more evident in cases like (21) below, where *se*, controlled by another subject NP with 1PL reference rather than *a gente*, seems to add number- and person-marking to the 3SG verb form.¹⁶

- (21) Eu mais meus primos ia-se buscar lenha.
 PRON.SBJ.1SG with my cousins go.IPFV.3SG=SE get-INF firewood
 ‘Me and my cousins used to go get firewood.’

¹⁴Posio (2012: 342) utilizes clusivity as the differentiating factor between “hearer-inclusive” and “impersonal (speaker-inclusive)” readings. The two extremes of the continuum proposed here are defined following Posio’s categorization.

¹⁵In Portuguese, there are several strategies to respond affirmatively to a polar question. One of these strategies consists of repeating the finite verb by adapting person and number features (cf. Martins 2013, 2016b)

¹⁶Note that there are no occurrences of lack of agreement between a coordinate preverbal subject and the verb in the data under analysis.

There are at least two possible analyses for examples like (21). One analysis would consider the preverbal coordinate NP (*Eu mais meus primos*) to be a topic, which precedes the sentence displaying a null subject *a gente* (*Eu mais meus primos, Ø ia-se buscar lenha*). A second possibility is to assume the preverbal NP (*Eu mais meus primos*) to be the subject. In light of the second analysis, example (21) could illustrate a further step in the grammaticalization path of *se* as a 1PL person marker.

There are several possible intermediate interpretations of the hybrid construction located between the impersonal and deictic poles of the proposed referential scale. These specific interpretations of [(*a gente*)+*se*] are determined by a vast array of co-textual factors.

(22) *Ia-se* *as duas*.
 go.3SG=SE the two
 ‘The two of us used to go.’

(23) *Ia-se* *todos para lá*.
 go.3SG=SE all to there
 ‘All of us went there.’

Examples (22) and (23) illustrate the graduality of possible specific interpretations of *se*. For instance, the specific reading of (22) relies on the cooccurrence with numerals, thus on the cardinality of the group. The interpretation of (23) is slightly less specific than the one triggered by (22). The cooccurrence of *todos* (‘all’) implies that there is a specific number of members in the set of referents which consequently evokes a specific rather than an impersonal reference.

The examples clearly illustrate the broad scope of references covered by the hybrid construction under analysis. It appears in contexts where possible references of *a gente* and *se-IMP* converge. This suggests that the speaker-inclusive impersonality shared by both constituents is the common denominator and may be where this hybrid construction originated. The fact that this originally *se-IMP* occurs in specific or even deictic contexts shows that it no longer requires *a gente* to establish personal reference, which could be a symptom of the reanalysis of *se* as a person marker. A possible syntactic catalyst for this reanalysis is described by Posio & Vilkuna (2013). The authors observe that, while in EP varieties, the patient NP can be reanalyzed as the subject of *se-IMP* constructions, “in Madeira and Porto Santo dialects the Patient has been reanalyzed as a direct object” (2013: 213). This reanalysis of the patient NP as a direct object can be observed in ex-

amples of *se*-IMP constructions featuring accusative clitics,¹⁷ the combination of which is considered ungrammatical in standard EP (Naro 1976: 786). The reanalysis of the patient NP as a direct object and thus eliminating it from the list of possible clausal subjects, might have served as a catalyst for the latter reanalysis of *se* as a person marker.

5 Syntactic features of [(*a gente*) + *se*]

The new insights into the referential scope of the hybrid construction, briefly introduced in the previous section, have substantial repercussions on the analysis of the syntactic properties of this particular phenomenon. Given that the hybrid construction [(*a gente*)+*se*] not only exhibits impersonal readings, in which they partially overlap with the semantics of canonical *se*-IMP constructions, but also allows for specific and even deictic readings (e.g., the reference to speech act participants), it is necessary to reconsider the syntactic and semantic properties of the formerly impersonal marker *se* in the studied variety.

This section describes and discusses different syntactic properties displayed by the construction under study here in contrast with the findings of previous studies. §5.1 deals with variable subject expression of *a gente* in these contexts. §5.2 describes verbal agreement patterns in clauses where [(*a gente*)+*se*] accounts for subject person marking. §5.3 provides evidence for the ability of *se* in these contexts to trigger adjectival agreement. §5.4 connects with the former two and adds a descriptive insight into the construction's behavior in coreferential contexts.

5.1 Variable subject expression

As it has already been stated in §3, the construction under analysis can occur with the subject pronoun *a gente* (24), with subject NPs as its antecedents (25), or in clauses without an overt subject or antecedent (26).

- (24) Mas *a gente* não *se* fazia bacalhau.
 But PRON.SBJ.1PL NEG SE=make-IPFV.3SG codfish
 'But we didn't make codfish.'

¹⁷Following Posio & Vilkuna (2013: 214) the reanalysis of the patient NP as a direct object can be observed in their example (24) partially reproduced here as (i).

- (i) Em sendo para a latada, deixa-se-a crescer [...]
 in be-GER for the trellis leve-3SG=SE=ACC.3SG grow INF
 'Being for the trellis, you let them grow [...].'

- (25) Eu e Alicinha, cada uma fazia a sua
 PRON.SBJ.1SG and Alicinha each one make-IPFV.3SG the POSS.F-SG
 semana. Cosia-se uma semana inteira o almoço.
 week cook-IPFV.3SG=SE one week whole the lunch
 ‘Alicinha and I each made her own week. (We) cooked lunch for a whole
 week.’
- (26) Se ele fosse preciso ser operado,
 If PRON.SBJ.3SG be.SBJV.IPFV.3SG necessary be.INF operated
 ficava-se lá.
 stay-IPFV.3SG=SE there
 ‘If it would be necessary for him to get surgery, (we) would stay there.’

The high frequency of hybrid constructions without an expressed subject, in which the clitic *se* is the primary element encoding 1PL reference on the verb, indicates the degree of grammaticalization of this expression in rural MP varieties. Indeed, the vast majority of the analyzed clauses in the data do not occur with the subject pronoun *a gente*.

Out of the 566 clauses containing [(*a gente*)+*se*], 177 occur with the overt subject pronoun. The remaining 389 cases are occurrences of 3SG verbal forms with the clitic *se* displaying personal (i.e. specific or deictic) interpretations.

Table 3: First person plural pronominal expression in rural MP varieties

	overt		null	
	<i>n</i>	%	<i>n</i>	%
[<i>a gente</i> + <i>se</i>]	177	31.3	389	68.7
<i>a gente/nós</i>	208	79.1	55	20.9

Contrary to these findings, Posio states for peninsular EP varieties that *a gente* “is usually expressed even in contexts that strongly favor the omission,” such as coreferential contexts within coordinated clauses (2012: 345). This is a straightforward consequence of the ambiguity conveyed by the omission of *a gente* with 3SG verb forms. Thus, as illustrated below, the presence of *se* results in higher permissibility of the 1PL subject omission. Example (27), for instance, illustrates an occurrence of *se* lacking an overt subject pronoun. Its specific interpretation is, again, determined by co-textual factors.

- (27) A: *Quantos filhos é que a sua mãe teve?*
 ‘How many children did your mother have?’
 B: *Era-se dez.*
be.IPFV.3SG=SE ten
 ‘(We) were ten.’ (=‘There were ten of us.’)

The fact that null pronominal subjects are so common in these contexts is relevant in two different respects. First, it shows that *se* is able to disambiguate 3SG verb forms while establishing 1PL reference. Second, it suggests that the element *se* is being reanalyzed as a 1PL person marker.

5.2 Variable verbal agreement patterns

In her analysis of the verbal agreement in constructions comprising a strong pronoun and the clitic *se*, Martins (2009: 185) found that it is the former that triggers agreement on the verb due to the presumably person-less nature of *se*. She draws this conclusion from the fact that the same variable agreement patterns – namely 3SG, 1PL, and 3PL – induced by the pronoun *a gente* can also be found encoded in the verbal forms associated with these “double subject impersonal *se* constructions” (Martins 2009: 185–186). In our data, as shown in Table 2, 13.5% of the clauses in which *a gente* or *nós* assumes the role of pronominal subject have the verb in 1PL. However, in regard to verbal agreement induced by the hybrid construction [(*a gente*)+*se*], the data display a clear predominance of 3SG verb forms and only six cases of verbal agreement with 1PL verb forms (1.05%). It is worth noting that in all these cases *se* appears in the proclitic position, thus occurring in finite subordinate clauses (28), and in principal clauses featuring negative polarity items (29) or other proclisis-inducing elements such as focalizing *já* (30).¹⁸

- (28) *porque se fomos as mais velhas*
 because SE=be.PST.1PL the more old
 ‘because (we) were the eldest’
- (29) *Não se morremos de fome.*
 NEG SE=die-PST.1PL of hunger
 ‘We didn’t die of hunger.’

¹⁸European Portuguese varieties display complex clitic positioning patterns. This issue, however, goes far beyond the scope of the present chapter. For a seminal description on this issue see Martins (2016a). Furthermore, a recent study by the same author (Martins 2021) highlights the clitic positioning in insular Portuguese varieties of the Madeira and Azores archipelagos.

- (30) Já se criámos dois de cada vez.
even SE=raise-PST.1PL two at each time
'We even raised two at a time.'

The proclitic position of *se* could explain the preferable omission of *a gente* and the use of the person marking morpheme *-mos* in these contexts. The combination of the three person marking items – *a gente*, morphological 1PL marking *-mos* and the element *se*¹⁹ – could lead to over-specification of 1PL person-encoding on the verb.²⁰

5.3 Adjectival agreement

For standard EP, Martins (2009: 191-192) states that the element *se* in impersonal constructions cannot establish adjectival agreement in predicative contexts. Comparing with the adjectival agreement properties found in dialectal EP varieties, the author differentiates between two types of “impersonal *se*”: one found in standard EP varieties, whose number feature corresponds to “singular”; the other one, found in EP dialects, manifesting the construction under analysis, displays the number feature “plural” and therefore allows “plural agreement between *se* and an adjectival predicate” (Martins 2009: 192). Additionally, the author proposes two examples illustrating the ungrammaticality of plural adjectival agreement in *se-IMP* constructions in standard EP, reproduced here as (31) and (32):

¹⁹There are no examples manifesting all three person marking items in the corpus under study. For illustrative purposes, consider the following fabricated example:

- (i) * A gente trabalhamos-se muito.
PRON.SBJ.1PL work-PRS.1PL=SE a lot
'We work a lot.'

²⁰Despite the overall high productivity of $[(a\ gente)+se]$, reflexive/reciprocal verbs seem to restrict the use of this construction. It must be noted that *se* is homonymous to the reflexive/reciprocal 3SG and 3PL clitic in Portuguese. Moreover, in some varieties with the pronoun *a gente* – including those mentioned here – it formally coincides with the 1PL reflexive/reciprocal clitic (Martins 2009: 185). Thus, in our data, reflexive/reciprocal verbs tend to block the use of $[(a\ gente)+se]$, due to the unacceptability of the sequence **se-se* discussed in Martins (2009: footnote 18). The ungrammaticality of the sequence **se-se* also accounts for the well-attested incompatibility of reciprocal/reflexive verbs in *se-IMP* constructions. In light of these observations, further analyses are required to understand the use of $[(a\ gente)+se]$ with other clitic pronouns.

- (31) Quando se é novo...
 When SE=be.PRS.3SG young-M.SG
 ‘When one is young...’
- (32) * Quando se é novos...
 When SE=be.PRS.3SG young-M.PL
 ‘When one is young...’

In terms of the hybrid construction [(*a gente*)+*se*], the data under analysis here confirm the tendencies described in Martins (2009). Thus, the hybrid construction exclusively triggers plural agreement in predicative contexts. Moreover, our data include cases of agreement reflecting the gender of the intended referents, as shown in examples (33) and (34):

- (33) Quando se era pequenos?
 when SE=BE.IPFV.3SG little-M.PL
 ‘When (we) were little?’
- (34) Era-se pequenas.
 be.IPFV.3SG=SE little-F.PL
 ‘(We) were little.’

The agreement contrast between these two examples stems from the fact that the group alluded to in example (34) is exclusively female (the informant is referring to herself and the neighbor’s daughter). Example (33), however, relates to the informant’s brothers and sisters, thus displaying default masculine and plural adjectival agreement.

The data under analysis suggest that the hybrid construction [(*a gente*) + *se*] displays non-variable plural adjectival agreement in predicative contexts. However, in terms of gender agreement, variable patterns can be found. These variable gender agreement patterns are consistent with those attested for other 1PL person marking items in Portuguese (Costa & Pereira 2013, 2005, Pereira 2003). Furthermore, the fact that the element *se* of hybrid constructions exclusively triggers plural agreement on the predicate might further endorse its status as a 1PL person marker.

5.4 Coreference

Coreference has been identified as an essential contributing factor for the expression or omission of subject pronouns in pro-drop languages. There is a broad consensus that coreference with a previous subject favors the omission of subject

pronouns (Silva-Corvalán 1982, among others). This section focuses on the role of coreference regarding the structural and semantic features of [(*a gente*)+*se*]. Even though this study does not claim to contemplate all the factors that enable subject omission in these contexts, the examples clearly illustrate that, even in contexts lacking an overt subject and those without an immediate 1PL subject antecedent, *se* establishes 1PL reference.

Previous studies on the variation of 1PL pronominal expression in Portuguese varieties have shown that the newer pronoun *a gente* occurring with 3SG verb forms can only be omitted in a restricted number of contexts. One requirement is coreference with the overt antecedent *a gente* (de Paiva Sória 2013). The quantitative analyses on rural MP varieties (cf. §3) confirm the tendencies found in previous research: only five examples of null subject *a gente* were found in the corpus. This is in line with Posio’s findings for continental EP varieties, where overt *a gente* even appears “in contexts that strongly favor the omission” (2012: 345).

In the context of the hybrid construction [(*a gente*)+*se*], the omission rate is much higher, occurring in 68.7% of the cases displaying 1PL reference (cf. Table 3). The data under analysis suggest that coreference is a key factor affecting the omission of the subject pronoun in hybrid constructions, resulting in a 3SG verb form and the clitic *se*. Furthermore, discourse connectedness (in terms of Paredes Silva 1993) appears to determine whether a given occurrence of *se* is to be interpreted as personal rather than impersonal. Example (35) shows the ability of the clitic *se* to maintain 1PL-specific reference when the pronoun *a gente* is omitted.

- (35) A gente era-se costumadas ambas.
 PRON.SBJ.1PL be.IPFV.3SG=SE accustomed both
 ‘We were used to each other.’
 Ia-se para a escola,
 go.IPFV.3SG=SE to the school
 ‘(We) used to go to school’
 ia-se as duas passava-se ali...
 go.IPFV.3SG=SE the two.FPL pass-IPFV.3SG=SE there
 ‘(we) used to go together, (we) would pass by..’

Cameron (1995) proposes that a 1PL expression is usually introduced into discourse only after its reference – or parts of the referent set – has previously been established in the discourse. This can be seen in example (36) where the informant starts the utterance with *eu* and then goes on to refer to herself and

her spouse, which can be inferred from the semantics of the verb *casar*. Partial coreference is thus established between the clitic *se* and the first person singular personal pronoun *eu*.²¹

- (36) Eu quando casei, criava-se dois [porcos].
 PRON.SBJ.1SG when marry-PST.1SG raise-IPFV.3SG=SE two [pigs]
 ‘When I got married, (we) raised two [pigs].’

The importance of coreference between a 1PL antecedent and the element *se* becomes even more evident when discourse-initial contexts are considered.

- (37) Matava-se um porco, era tudo salgado.
 kill-IPFV.3SG=SE a pig be.IPFV.3SG everything salted
 ‘One used to kill pigs, everything had to be salted.’
 Comprava-se uma salga para salgar o porco.
 buy-IPFV.3SG=SE a <name> to salt-INF the pig.
 ‘One bought a salting vessel to salt the pig.’

In discourse-initial contexts, the absence of *a gente* or another 1PL referent renders the personal interpretation of *se* improbable or at least impossible to determine. Consider the contrast between the previous example (37) and example (38) below:

- (38) Em princípio, a gente foi-se bebés.
 in beginning PRON.SBJ.1PL be.PST.3SG=SE babies
 ‘First, we were babies.’
 Depois cresceu-se, foi-se para a Escola das Irmãs.
 then grow.PST.3SG=SE go.PST.3SG=SE to the <name of the school>
 ‘Then (we) grew up, (we) went to Escola das Irmãs.’
 Mas como os meus pais não tinham a possibilidade de pôr a gente a estudar,
 ‘But, since my parents did not have the possibility to let us go to school, apenas se deu a terceira classe.
 only SE=give.PST.3SG. the third class
 ‘(we) only completed the third grade.’

²¹This partial coreference is connected to what Gelbes (2008: 522–524) calls “correferencia inclusiva” (‘inclusive correferentiality’).

Example (38) shows another extract of a discourse-initial context. The informant answers a question on how many siblings she has and how they were brought up. She retrieves the set of referents – the informant and her siblings – by using the pronoun *a gente* in the first sentence. What follows is a chronological depiction of the events with coreferential null subjects. Interestingly, in (38) the pronoun *a gente* is omitted in the fourth sentence (*apenas se deu a terceira classe*), even though there is discontinuity regarding the previous subject *os meus pais*. This shows, in part, that coreference cannot fully account for the variation between expression and omission of the subject in hybrid constructions. The correct interpretation of the 1PL subject, in this case, is most probably established by the perfective past (*deu*) describing an event anchored in time, thus favoring a personal interpretation (Siewierska 2011).

As anticipated in the introductory lines to this section, there are particular contexts in which the omitted subject of the hybrid construction is not coreferential to the previous subject. The following examples, for instance, show contexts that strongly favor the omission of the subject. Hence, the element *se* is used to encode 1PL person marking on its own.

- (39) Teresinha, vai-se brincar!
 <name> go.PRES=SE play-INF
 ‘Teresinha, let’s go play!’
- (40) A: A senhora brincava com os seus irmãos?
 the Mrs play-IPFV.3SG with the your siblings
 ‘Did you play with your siblings?’
 B: Brincava-se ao domingo.
 play-IPFV.3SG=SE on-the Sunday
 ‘Yes, (we) played on Sundays.’

Example (39) shows a cohortative construction that “expresses the exhortation to the addressee to carry out an action together with the speaker” (Dobrushina & Goussev 2005: 179). Hence, the interpretation of the referent is inherently personal (i.e., deictic). Another context in which *se* establishes personal 1PL reference is found in affirmative verbal responses, such as (40). In these contexts, the element *se* retrieves the set of referents defined in the question (i.e., the informant and her siblings) thus assuming the role of a 1PL person marker. These examples can be considered crucial evidence for the reanalysis of *se* of the hybrid constructions in rural MP varieties.

6 Conclusion

Rural Madeiran Portuguese varieties manifest two predominant 1PL pronominal expressions: *a gente* and the more common variant [(*a gente*)+*se*]. Furthermore, the quantitative analysis indicates that the presence of *se* allows for substantial variation in terms of the presence and absence of *a gente* or other subjects displaying 1PL reference (cf. Table 3). To account for this fact, a hypothesis was anticipated that the clitic *se* seems to display 1PL marking features in the absence of other person markers. There is evidence within the syntactic properties of [(*a gente*)+*se*] outlined in this study that seems to support this tentative hypothesis:

1. There is a meager rate of 1PL verb forms in the context of these constructions, which might instigate that *se* suffices to establish 1PL reference.
2. The hybrid construction (with and without the overt subject) can trigger variable gender agreement according to the constellation of the alluded group. However, in terms of number adjectival agreement in predicative contexts it exclusively triggers plural agreement.
3. Independently of its coreference with a 1PL antecedent, *se* can trigger personal interpretations in contexts lacking an overt subject. Thus, it can be found in verbal affirmative answers and cohortative constructions, both of which favor the omission of the subject.

In terms of its referential properties, several observations can be made. The hybrid construction, whose constituents originate from impersonalizing strategies, might imply not only specific interpretations but also deictic ones, even in the contexts mentioned above lacking an overt subject. The fact that *se* can refer to speech-act participants is the most straightforward argument supporting the initial tentative hypothesis. However, more research is needed to determine whether or not discourse-initial antecedentless contexts are the only restriction for the occurrence of *se* referring to a specific 1PL subject. Moreover, the analysis of diachronic data could offer more substantial insights into the possible origins of this hybrid construction.

Acknowledgments

I would like to express my gratitude to all the informants who agreed to participate. Moreover, I wish to thank Carlota de Benito Moreno, Johannes Kabatek, two anonymous reviewers and the editors of this volume for their helpful comments, suggestions and insights.

References

- Bazenga, Aline. 2015. A concordância de terceira pessoa plural: A variedade insular do PE (funchal). In Silvia Rodrigues Vieira (ed.), *A concordância verbal em variedades do Português: A interface fonética-morfossintaxe*, 76–103. Rio de Janeiro: Vermelho Marinho.
- Bazenga, Aline. 2019. A variação entre ter e haver em construções existenciais numa variedade insular do PE (Funchal). In Ernestina Carrilho, Ana Maria Martins, Sandra Pereira & João Paulo Silvestre (eds.), *Estudos linguísticos e filológicos oferecidos a Ivo Castro*, 181–216. Lisboa: Centro de Linguística da Universidade de Lisboa.
- Callou, Dinah & Célia Regina dos Santos Lopes. 2004. Contribuições da sociolinguística para o ensino e a pesquisa: A questão da variação e mudança linguística. *Revista do GELNE (UFC)* 5. 63–74.
- Cameron, Richard. 1995. The scope and limits of switch reference as a constraint on pronominal subject expression. *Hispanic Linguistics* 6(7). 1–27.
- Casteleiro, João Malaca. 1975. Aspectos da sintaxe do Português falado no interior do país. *Boletim de Filologia* XXIV. 57–74.
- Chambers, Jack K. & Peter Trudgill. 1980. *Dialectology*. Cambridge: Cambridge University Press.
- Cinque, Guglielmo. 1988. On *si* constructions and the theory of Arb. *Linguistic Inquiry* 19. 521–581.
- Costa, João & Sandra Pereira. 2005. Phases and autonomous features: A case of mixed agreement. In Martha McGinnis & Norvin Richards (eds.), *Perspectives on phases* (MIT Working Papers in Linguistics 49), 115–124. Cambridge, MA: MIT Press.
- Costa, João & Sandra Pereira. 2013. *A gente*: Pronominal status and agreement revisited. *The Linguistic Review* 30. 161–184. DOI: 10.1515/tlr-2013-0006.
- Coveney, Aidan. 2000. Vestiges of *nous* and the 1st person plural verb in informal spoken French. *Language Sciences* 22. 447–481. DOI: 10.1016/S0388-0001(00)00014-0.
- de Omena, Nelize Pires. 2003. A referência à primeira pessoa do discurso no plural. In Maria da Conceição de Paiva & Maria Eugênia Lamoglia Duarte (eds.), *Mudança linguística em tempo real*, 63–80. Rio de Janeiro: Contra Capa/FAPERJ.
- de Paiva Sória, Máira Vasconcelos. 2013. *Nós, a gente e o sujeito nulo de primeira pessoa do plural*. Universidade de Lisboa. (MA thesis). <http://hdl.handle.net/10451/12218>.

- Dobrushina, Nina & Valentin Goussev. 2005. Inclusive imperative. In Elena Filimonova (ed.), *Clusivity: Typology and case studies of the inclusive-exclusive distinction*, 179–211. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/tsl.63.09dob.
- Gelbes, Silvia Ramírez. 2008. Correferencia y aspectualidad: La correferencia inclusiva y los verbos télicos. *Zeitschrift für Romanische Philologie* 124. 514–530. DOI: 10.1515/zrph.2008.047.
- Koch, Peter & Wulf Oesterreicher. 1990. *Gesprochene Sprache in der Romania*. Tübingen: Max Niemeyer Verlag.
- Lopes, Célia Regina dos Santos. 2003. *A inserção de a gente no quadro pronominal do português: Percorso histórico*. Frankfurt am Main/Madrid: Vervuert/Iberoamericana. DOI: 10.31819/9783865278494.
- Lopes, Célia Regina dos Santos. 1999. *A inserção de a gente no quadro pronominal do português: Percorso histórico*. Universidade Federal do Rio de Janeiro. (Doctoral dissertation).
- Lopes, Célia Regina dos Santos. 2007. Pronomes pessoais. In Silvia Figueiredo Brandão & Silvia Rodrigues Vieira (eds.), *Ensino de gramática: Descrição e uso*, 103–114. São Paulo: Contexto.
- Martins, Ana Maria. 2003. Construções com *se*: Mudança e variação no português europeu. In Ivo Castro & Inês Duarte (eds.), *Razões e emoção: Miscelânea de estudos em homenagem a Maria Helena Mateus*, 163–178. Lisboa: Imprensa Nacional: Casa da Moeda.
- Martins, Ana Maria. 2005. Passive and impersonal *se* in the history of Portuguese. In Klaus Pusch, Johannes Kabatek & Wolfgang Raible (eds.), *Romance corpus linguistics II: Corpora and diachronic linguistics*, 411–430. Tübingen: Gunter Narr.
- Martins, Ana Maria. 2009. Subject doubling in European Portuguese dialects: The role of impersonal *se*. In Enoch O. Aboh, Elisabeth van der Linden, Josep Quer & Petra Sleeman (eds.), *Romance languages and linguistic theory: Selected papers from Going Romance, Amsterdam 2007*, 179–200. Amsterdam: John Benjamins. DOI: 10.1075/rllt.1.
- Martins, Ana Maria. 2013. Emphatic polarity in European Portuguese and beyond. *Lingua* 128. 95–123. DOI: 10.1016/j.lingua.2012.11.002.
- Martins, Ana Maria. 2016a. A colocação dos pronomes clíticos em sincronia e diacronia. In Ana Maria Martins & Ernestina Carrilho (eds.), *Manual de linguística portuguesa*, 401–430. Berlin/Boston: De Gruyter.
- Martins, Ana Maria. 2016b. O sistema responsivo: Padrões de resposta a interrogativas polares e a asserções. In Ana Maria Martins & Ernestina Carrilho (eds.), *Manual de linguística portuguesa*, 581–609. Berlin/Boston: De Gruyter.

- Martins, Ana Maria. 2021. Microvariação na sintaxe dos clíticos: Os dialetos portugueses dos Açores e Madeira. *Estudos de Linguística Galega* 13. 67–105. DOI: 10.15304/elg.13.6997.
- Martins, Ana Maria & Jairo Nunes. 2021. Brazilian and European Portuguese and Holmberg's 2005 typology of null subject languages. In Sergio Baauw, Frank Drijkoningen & Luisa Meroni (eds.), *Romance languages and linguistic theory 2018: Selected papers from "Going Romance" 32, Utrecht*, 172–190. John Benjamins: Amsterdam. DOI: 10.1075/cilt.357.
- Naro, Anthony. 1976. The genesis of the reflexive impersonal in Portuguese: A study in syntactic change as a surface phenomenon. *Language* 52. 779–810.
- Paredes Silva, Vera Lucia. 1993. Subject omission and functional compensation: Evidence from written Brazilian Portuguese. *Language Variation and Change* 5(1). 35–49. DOI: 10.1017/S0954394500001381.
- Pereira, Sandra Maria. 2003. *Gramática comparada de a gente: Variação no Português Europeu*. Universidade de Lisboa. (MA thesis). https://www.clul.ulisboa.pt/files/sandrabrito_pereira/Pereira_2003_tese.pdf.
- Posio, Pekka. 2012. Who are 'we' in spoken Peninsular Spanish and European Portuguese? Expression and reference of first person plural subject pronouns. *Language Sciences* 34(3). 339–360. DOI: 10.1016/j.langsci.2012.02.001.
- Posio, Pekka & Maria Vilkuna. 2013. Referential dimensions of human impersonals in dialectal European Portuguese and Finnish. *Linguistics* 51. 177–229. DOI: 10.1515/ling-2013-0006.
- Raposo, Eduardo & Juan Uriagereka. 1996. Indefinite *se*. *Natural Language and Linguistic Theory* 14. 749–810.
- Siewierska, Anna. 2011. Overlap and complementarity in reference impersonals: Man-constructions vs. third person plural-impersonals in the languages of Europe. In Andrej L. Malchukov & Anna Siewierska (eds.), *Impersonal constructions: A cross-linguistic perspective*, 57–89. Amsterdam: John Benjamins. DOI: 10.1075/slcs.124.03sie.
- Silva-Corvalán, Carmen. 1982. Subject expression and placement in Mexican-American Spanish. In Jon Amastae & Lucia Elías-Olivares (eds.), *Spanish in the United States: Sociolinguistic aspects*, 93–120. Cambridge: Cambridge University Press.
- Taylor, Michael. 2009. On the pronominal status of Brazilian Portuguese *a gente*. In Patricia Irwin & Violeta Vázquez Rojas Maldonado (eds.), *Papers in syntax* (NYU Working Papers in Linguistics 2), 1–36. New York City: New York University.

- Travis, Catherine E. & Agripino S. 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(2). 347–376. DOI: 10.1515/shll-2009-1055.
- Vianna, Juliana Barbosa de Segadas. 2011. *Semelhanças e diferenças na implementação de a gente em variedades do português*. Universidade Federal do Rio de Janeiro. (Doctoral dissertation).

Chapter 7

On pronominal uses of *geral* in Brazilian Portuguese

Juanito Ornelas de Avelar

University of Campinas

This paper analyzes the impersonal use of *geral* ‘general’ in Brazilian Portuguese, in the light of investigations dealing with impersonalization strategies in the generative literature. I will show that *geral* behaves as a φ -featureless impersonal pronoun with regard to agreement patterns and to generic/arbitrary interpretation, but as a pronoun with φ -features if we take its syntactic distribution into consideration. Despite this incongruity, I will argue that *geral* must be analyzed as an item that is devoid of φ -features, similarly to *man* in Swedish, *si* in Italian and *on* in French, according to Egerland’s (2003) proposal. The analysis provides evidence in favor of the hypothesis that the distribution of impersonal pronouns in different sentential positions is better captured in terms of case marking instead of syntactic function (Fenger 2018). I will also show that *geral* can be used as a first-person plural pronoun, which seems to depend on strictly pragmatic factors, as a result of the lack of φ -features.

1 Introduction

This study approaches the occurrence of *geral* ‘general’ in Brazilian Portuguese in cases in which its use can be analyzed as an impersonalization strategy (i.e., denoting an indefinite human referent). *Geral* is originally an adjective, but the examples in (1) below illustrate its use as a generic pronoun (corresponding to generic ‘you’ in English – see 1a–1b) or as an arbitrary pronoun (corresponding to referentially undetermined ‘they’ in English – see 1c–1d).¹

¹Please see Appendix A for links to the internet sources for all examples.



- (1) a. *Geral* sabe que esporte gera renda.
GERAL knows.PRES that sport generates money
'Everyone knows that sports generate income.'
- b. Bom mesmo era na idade média que *geral* morria de sífilis
good really was in-the age middle that GERAL died.3SG of syphilis
e ninguém tava nem aí.
and nobody was unconcerned
'Things went really well in the Middle Ages, when everyone/you died
from syphilis and nobody cared.'
- c. de noite fomos pro baile e depois *geral* foi
at night went.PAST.1PL to-the prom and after GERAL went
chegando e curtindo pra caramba
arriving and enjoying very much
'in the evening we went to the dance and many people/some people
kept on coming and enjoyed it a lot'
- d. Olha aí a galera dos comes e bebes da festa que
look there the people of-the foods and drinks of-the party that
geral ficou alucinado!
GERAL was crazy
'These are the guys who took care of the party's foods and beverages
that got many people/lots of people crazy!!!'

Based on the Minimalist version of the Principles and Parameters Theory (Chomsky 1995), this study aims at analyzing the behavior of *geral* in the light of investigations dealing with impersonal pronouns in the generative literature. I approach in greater detail the works of Egerland (2003) and Fenger (2018), who pursue the hypothesis that impersonal pronouns fall into two groups with regard to the presence or absence of the so-called φ -features (which codify information related to grammatical categories such as gender, number and person): those that exhibit φ -features and those that are devoid of φ -features. The presence or absence of φ -features has syntactic and pragmatic implications, since they determine the syntactic positions in which impersonal pronouns may occur and, at least in part, also condition their readings in a given context. Although *geral* behaves in a way that is apparently inconsistent with what would be expected in terms of φ -features, I will argue that it must be analyzed as an item that is devoid of such features. I shall further argue that, under specific pragmatic circumstances, *geral* may gain a referentially definite reading, equivalent to the personal pronouns *nós* and *a gente* 'we', which refer to the first-person plural in Brazilian Portuguese.

This study will be carried out from an exclusively qualitative perspective. In addition to resorting to my own intuition (a common expedient in generative investigations), the analysis was based on data collected from webpages, informal writing on blogs, social networks, forums and commercial advertisement. The database currently consists of around 150 occurrences of *geral* as an impersonalization strategy and has been gathered since 2018. In most cases, it is not possible to identify the author of each utterance nor their regional provenance, which at this point prevents me from engaging in more detailed sociolinguistic considerations regarding the relevant use of *geral*.

This chapter is structured as follows: in §2, I present other uses of *geral*, i.e., as an adjective, a noun and an adverb, in which this item also conveys the meanings of indetermination or intensification; in §3, I present the proposals of Egerland (2003) and Fenger (2018) for dividing impersonal pronouns into two large groups – those that exhibit φ -features and those that are devoid of φ -features; in §4, I analyze the behavior of *geral* with regard to the expected properties of each of those types of pronouns, with the purpose of establishing in which group it belongs; in §5, I present some occurrences of *geral* in which it refers to the first-person plural; in §6, I summarize the conclusions of this study.

2 Adjectival, nominal and adverbial uses of *geral*

The use of *geral* as an impersonalization strategy has been associated with the speech of younger individuals (below 30 years of age) living in urban areas in different regions of Brazil. Nevertheless, a systematic and in-depth sociolinguistic study on the distribution of *geral* in Brazilian Portuguese remains to be conducted so as to enable a precise mapping of occurrences according to geographic, social and age criteria. Although it is commonly associated with spontaneous utterances of younger individuals, *geral* may be observed in different age groups, including the author of this paper, who is currently in his forties.

As an adjective, the *Dicionário Houaiss da Língua Portuguesa* gives the following definitions of *geral*: ‘that which applies to an array of cases or individuals’ (2a), ‘that which embraces the totality or the majority of a group of persons or things’ (2b) and ‘universal, widespread’ (2c), among others. The examples are presented in the online version of *Dicionário Houaiss*.

- | | | | |
|-----|----|--------------------------|--------------------------------|
| (2) | a. | lei <i>geral</i> | ‘general statute’ |
| | | assembleia <i>geral</i> | ‘general assembly’ |
| | | busca o bem <i>geral</i> | ‘(s)he seeks the general good’ |

In the latter case, *geral* works as an intensifier, as *muito* ‘very, much’, *demais* ‘too much’, *tudo* ‘everything’, *todo lugar* ‘everywhere’, among other possibilities. The adverbial use of *geral* is quite common even among speakers that find the impersonal use of *geral* odd and/or do not resort to it.

- (5) a. Bebeu *geral* e o que serviam ele tomava
 drank.3SG general and the that served.3PL he took.3SG
 ‘He drank a lot and took whatever was being served.’
- b. Choveu *geral* nos últimos sete dias, mas em volumes
 rained general in-the last seven days but in volumes
 diferenciados
 different
 ‘It has rained heavily in the past seven days, but in different volumes.’
- c. A coisa não funcionou legal e [ele] se machucou *geral*
 the thing not worked well and he injured general
 ‘It didn’t work out well and he was badly injured.’

The cases this study is more directly concerned with are those presented in (1), in which *geral* is used as an impersonalization strategy. It is not clear how this item came to be used as a pronoun, but it is possible that this usage represents an advanced stage of its grammaticalization in certain varieties of Brazilian Portuguese. However, I will not concern myself with this question in the present study.

3 Impersonal pronouns and ϕ -features

3.1 Agreement patterns, interpretation and syntactic distribution

In this section, I approach the studies of Egerland (2003) and Fenger (2018), who explore the idea that impersonal pronouns may be described with regard to the presence or absence of ϕ -features. Other generative studies adopt similar perspectives or introduce different approaches (Cinque 1988, D’Alessandro & Alexiadou 2003, Hoekstra 2010, among others), but I restrict myself to these two contributions because the analyses they put forward deal with aspects that are more directly pertinent to a formal comparison with properties observed in the use of *geral*.

Based on these authors’ proposals, the structures of impersonal pronouns with and without ϕ -features may be represented as in (6a) and (6b) respectively: in (a),

we see what I will term IMP pronouns here, i.e., impersonal pronouns devoid of φ -features; in (b), we see what I will term φ -IMP pronouns, i.e., impersonal pronouns with φ -features.

- (6) a. IMP: [NP N]
 b. φ -IMP: [φ P φ [NP N]]

In short, IMP pronouns are bare, in the sense that there is no functional projection associated with N, whereas φ -IMP pronouns have at least one projection (φ P), a phrase headed by specified φ -features. Egerland (2003) analyzes impersonal pronouns in Romance and Scandinavian languages to show that *man* in Swedish, *on* in French and *si* in Italian are of the IMP type, whereas *maður* in Icelandic and *du* in Swedish (as *you* in English) are of the φ -IMP type. The author seeks to derive some grammatical and pragmatic properties from the opposition in (6) above. I will address three such properties here, which shall be relevant for an analysis of the behavior observed in *geral*: (i) agreement patterns, (ii) generic and/or arbitrary readings, and (iii) its syntactic position.

3.1.1 Agreement patterns

According to Egerland, φ -IMP pronouns display previously specified φ -features and therefore always trigger the same agreement mark. The Icelandic pronoun *maður*, exemplified in (7), is just such a pronoun: adjectives related with it must be marked as singular, as *stoltur* ‘proud’ in (7a) and *sannfærður* ‘convinced’ in (7b); the use of the plural form of these adjectives (*stoltir* and *sannfærðir* respectively) is ungrammatical.

- (7) Icelandic (Egerland 2003: 78)
- a. Í hernum er maður stoltur / *stoltir af henni.
 in the army is MAÐUR proud.SG / proud.PL of her
 ‘People in the army/they are proud of her.’
- b. Þrátt fyrir sannanirnar var maður ekki alveg
 in spite of the evidence was MAÐUR not completely
 sannfærður / *sannfærðir um sekt hans.
 convinced.SG/convinced.PL about guilt his
 ‘in spite of the evidence, people/they were not convinced.’

In contrast with φ -IMP, IMP may be associated with different agreement markings. According to Egerland, this is due precisely to the fact that this kind of

pronoun has no φ -feature, which makes it possible for it to occur along with items bearing different agreement markings. The Swedish cases in (8) exemplify this property: in (8a) the pronoun *man* occurs with the singular and plural forms of the adjective corresponding to *proud* in English (*stolt* and *stolta* respectively); the same occurs with the form corresponding to *convinced* (*övertygad* and *övertygade* respectively) in (8b).

(8) Swedish (Egerland 2003: 78)

- a. Inom armén är man stolt/stolta över henne.
 within the army is MAN proud.SG/proud.PL of her
- b. Trots bevisföringen var man inte helt
 in spite of the evidence was MAN not completely
 övertygad / övertygade om hans skuld.
 convinced.SG / convinced.PL about his guilt

Egerland shows that the lack of uniformity in the agreement patterns may also be observed in the case of gender markings. In Italian, for example, the pronoun *si* in copular constructions usually requires the third-person singular to be marked in the verb, with the adjective in the masculine plural, as in (9a). However, the adjective may be used in the feminine if *si* refers to a group of women, as in (9b).

(9) Italian (Egerland 2003: 79)

- a. Quando si è giovani,...
 when SI is.SG young.PL.MASC
 ‘When people are young...’
- b. Quando si è donne, si è disposte a rinunciare a molte cose
 when SI is women SI is ready.PL.FEM to renounce to many things
 per i propri figli.
 for the children

In short, the relevant distinctions between Icelandic *maður*, on the one hand, and *man* in Swedish and *si* in Italian, on the other, are a result of the presence or absence of φ -features: whereas *man* and *si* are IMP pronouns (allowing them to occur with items bearing different φ -feature specifications), *maður* is φ -IMP (limiting their occurrence to items agreeing with their φ -features).

3.1.2 Generic and arbitrary readings

Another relevant distinction between IMP and φ -IMP is their compatibility with generic and/or arbitrary interpretations. Egerland takes pronouns that have a generic reading to refer to “a quasi-universal set of individuals” whereas those that

have an arbitrary reading describe “a non-specific group of individuals”. The author argues that IMP can have generic and arbitrary readings, as in the examples of Swedish *man* in (10a) and (10b) respectively. In contrast, the φ -IMP Icelandic pronoun *maður* only licenses the generic reading, as can be seen in the grammaticality contrast in (11): in (11a) the meaning is generic, but the only possible reading in (11b) is that it is arbitrary because of the choice of verbal tense (a form of past simple that is usually not compatible with the generic interpretation).

(10) Swedish (Egerland 2003: 76; 80)

- a. Man måste arbeta till 65.
MAN must work until 65
'People have to work until the age of 65.'
- b. Man arbetade i två månader för att lösa problemet.
MAN worked for two months to solve the problem
'Some people/they worked for two months to solve.'

(11) Icelandic (Egerland 2003: 81)

- a. Maður vinur til 65 ára aldurs.
MAÐUR works until 65 years age
- b. *Maður hefur unnið að því í tvo mánuði að leysa vandamálið.
MAÐUR has worked for two months to solve the problem

Taking into consideration this contrast between Swedish *man* (as well as Italian *si* and French *on*) and Icelandic *maður*, Egerland comes to the generalization that φ -IMP pronouns can only be generic, whereas IMP pronouns can have both generic and arbitrary readings. In order to account for this distinction, the author assumes that the generic reading is defined by the presence of a generic operator (see Krifka et al. 1995 and Chierchia 1995) that can have either a φ -IMP or an IMP pronoun under its scope. In contrast, the arbitrary reading is only triggered when a pronoun is devoid of lexical content (which the author sees as equivalent to not having φ -features) beyond the trait [+human] and is not under the scope of a generic operator. According to Egerland (2003: 89),

By and large, the meaning of such an element amounts to nothing but a [+human] entity in an episodic context. Whether the subject is understood as a single individual or a group of people is entirely determined by the discourse context and is not restrained by any syntactic restrictions. Essentially, this amounts to saying that the notion “arbitrary” [...] has no theoretical status and that there is no natural class of “arbitrary pronouns”. Also, there is no “arbitrary” feature to be appealed to in syntactic derivations.

Therefore, according to Egerland's proposal, a pronoun's ability to bear both readings (generic and arbitrary) or only one of them depends on whether it has or does not have φ -features: those that are devoid of φ -features (IMP) are naturally interpreted as arbitrary in the absence of a generic operator; those that have φ -features (φ -IMP) will only be interpreted as impersonal if they are under the scope of this operator and, since they have lexical content, they will never be interpreted as arbitrary.

3.1.3 Syntactic function

Egerland (2003) also notes that a distinction between IMP and φ -IMP lies in their syntactic functions: IMP can only appear as syntactic subjects, whereas φ -IMP can appear syntactically as both subjects and objects. This distinction may be observed in comparing (12) and (13): in (12), the syntactic object is an IMP pronoun (*man*, *si* and *on* in Swedish, Italian and French, respectively) and the resulting sentences are ungrammatical; in (13), the object pronouns are the non-nominative versions of *maður* and *du* (which can work as a φ -IMP pronoun in Swedish) and, in this case, sentences are grammatical.

(12) (Egerland 2003: 91)

a. Swedish

* Det har sett man.
they have seen MAN

b. French

* Ils ont vu on. / *Ils on ont vu.
they have seen ON

c. Italian

* Loro si hanno visto.
they have seen SI.

(13) (Egerland 2003: 91)

a. Icelandic

Svona tölur segja manni að eitthvað sé í ólagi.
such figures tell MANNI that something is wrong

b. Swedish

Om de litar på dig får du inte göra dem besvikna.
if they rely on YOU must YOU not make them disappointed

According to Egerland, IMP cannot be syntactic objects because, in the absence of φ -features, their semantic role as an internal argument cannot be properly evaluated. In favor of this analysis, Egerland (2003) draws attention to the distinction

between Swedish nominative *man* and accusative/oblique *en*, in sentences such as (14): the relation between these two pronouns is the same that can be observed between English *he* and *him*, and the morphological distinction is due to case-marking.

- (14) Om de litar på en_i får man_i inte göra dem beskvina.
 if they rely on one must MAN not make them disappointed

Because *en* is a numeral, it bears inherent number marking, which makes it possible to classify it as φ -IMP and thus allows it to occur in a non-nominative position. From an interlinguistic perspective, Egerland notes that *man*-impersonals may not occur in an object position, whereas *one*-impersonals suffer no such restriction.

Egerland's analysis thus establishes a relation between the presence or absence of φ -features and a set of grammatical and pragmatic properties in the use of impersonal pronouns: IMP pronouns, precisely because they lack φ -features, have different agreement patterns and are only possible in nominative position, where they license both a generic and an arbitrary reading; φ -IMP pronouns, in turn, have a fixed agreement pattern and may present different syntactic functions, but only license a generic reading.

3.2 Case-marking, φ -features and syntactic distribution

Fenger (2018) analyzes the behavior of the so-called “dedicated impersonal pronouns” (i.e., those that are exclusively used as impersonal pronouns) in eight Germanic languages, and argues that, differently from what Egerland (2003) had stated, it is not the syntactic function that restricts the distribution of these pronouns, but rather the presence of a case projection in their internal configuration. The author assumes that only pronouns with φ -features may have case projection (KP), as represented in (15b), whereas pronouns devoid of φ -features cannot have such projection (15a). As a result, the distinction between IMP and φ -IMP presented in the previous section can be translated as a distinction between NP (noun phrase) and KP (Kase phrase), the latter of which has φ -features in its internal constitution.

- (15) a. IMP: [NP N]
 b. φ -IMP: [KP K [φ P φ [NP N]]]

From a theoretical perspective, in the light of the Minimalist Program (Chomsky 1995), Fenger's proposal (2018) is easily motivated by the assumption that case-marking results from the agreement between the (interpretable) φ -features of a noun constituent and the (non-interpretable) φ -features of a given functional head: the agreement with φ -features present, for example, in T(ense), V(erb) and P(reposition) results in nominative, accusative and oblique markings, respectively, on a noun constituent.

In line with other studies on case-marking (for instance, Bittner & Hale 1996, Neeleman & Weerman 1999, among others), Fenger assumes nominative to be a non-case, which implies that nominative constituents bear no KP, differently from, e.g., accusatives and obliques. Fenger derives the syntactic distribution of IMP and φ -IMP pronouns from this property: since they bear no KP, IMP pronouns may only occur in nominative positions; φ -IMP pronouns may in turn occur in positions associated with other cases, precisely because they bear a KP.

One advantage of Fenger's (2018) over Egerland's (2003) proposal is related to the distribution of IMP and φ -IMP in ECM-constructions: the latter can be a subject of such constructions (provided its reading is generic), whereas the former cannot, irrespective of their generic or arbitrary reading. The distinction is illustrated in (16) and (17) below, with Icelandic and Swedish examples presented by Fenger (2018: 299–300). If the distribution of impersonal pronouns depended exclusively on syntactic function, the Swedish sentences in (16b) and (17b) should be grammatical, since *man*, an IMP pronoun, is the subject in both cases; because the sentence is an ECM-construction, the position of *man* within the embedded clause is marked with the accusative, not the nominative case, which is easily explained if we assume that this pronoun has no KP, a mandatory projection for enabling the occurrence of noun constituents in positions other than the nominative. In Icelandic, as expected, *mann*, the accusative form of *maður*, can appear in this position because it is a φ -IMP pronoun, as can be seen in (16a). The ungrammaticality of (17a) results from the fact that *mann* cannot be read arbitrarily, irrespective of the position in which it occurs.

(16) Context: He is a station master.

Intended: 'Therefore he always sees *people* leave for the holidays.'

a. Icelandic (φ -IMP, generic)

Þess vegna sér hann mann alltaf fara í frí.
that because see he IMPERSONAL always leave in holiday

b. Swedish (IMP, generic)

* Därför see han man alltid åka på semester.
therefore see he IMPERSONAL always go on holidays

- (17) Context: I lay awake all night.
Intended: ‘I heard *someone* work on the road.’
- a. Icelandic (φ -IMP, arbitrary)
* Ég heyrði mann vinna vegavinnu.
I heard IMPERSONAL work road.construction
- b. Swedish (IMP, arbitrary)
* Jag hörde man arbetade ute på gatan.
I heard IMPERSONAL work out in the.street

Fenger believes that the generic reading is achieved through the presence of a generic operator [GEN] that can bind both IMP and φ -IMP pronouns. As for φ -IMP, the author states that “its feature specification includes the speaker and the addressee, and this is not contradictory to the requirements of [GEN]. It does not mean that [GEN] always needs to have an element which necessarily includes [speaker] and [addressee], but the element cannot have features that are contradictory with [GEN]” (Fenger 2018: 310).

As for the arbitrary reading, Fenger departs from Egerland regarding the idea that the arbitrary reading stems naturally from the absence of both lexical content (φ -features in this case) and a generic operator. For Fenger, at least two possibilities can be entertained: (i) the presence of an existential operator or (ii) the local relation between the pronoun and the Asp(ect) head. The author does not commit herself to either proposal, but regards both as superior to Egerland’s hypothesis because they predict the occurrence of φ -IMP pronouns that may be read arbitrarily, such as German *wer* and English *they*, as in (18) below.

- (18) a. Ich habe die ganze Zeit *wen* auf/an der Strasse arbeiten hören.
I have the whole time INDEF on the road work hear
‘I heard someone work on the road.’
- b. *They have called for you, but I don’t know what is about.*

I shall here assume Fenger’s proposal regarding the presence of KP in φ -IMP and the absence of this projection in IMP. As I will argue, this proposal is quite advantageous in the attempt to locate the occurrences of *geral* within the set of properties of the Brazilian Portuguese pronominal system. As for the conditions for the generic and/or arbitrary reading, I shall not commit myself to any approach here, since, from a purely formal perspective, there are no relevant consequences for the properties of *geral* that I will be considering.

3.3 Comparative perspective

The table below summarizes the set of properties that have been addressed in this section according to the distinction of impersonal pronouns as IMP and φ -IMP. In the next section, I will analyze *geral* according to these properties in an attempt to determine the best characterization of this item when used as an impersonal pronoun.

Table 1: Properties of IMP and φ -IMP pronouns

Properties	IMP	φ -IMP
Is associated with a single form of agreement.	no	yes
May have either a generic or an arbitrary reading.	yes	no
Only occurs in nominative positions.	yes	no

4 Properties of *geral* in Brazilian Portuguese

4.1 Agreement patterns

As far as agreement is concerned, *geral* behaves as an IMP pronoun, since it does not trigger fixed agreement marking in verbs and adjectives. Particularly in the case of verbal agreement, *geral* can occur with verbs both in the third-person singular, as in (19), and in the third-person plural, as in (20), although the former is more frequent.

- (19) a. eu não consigo entender pq *geral* não gosta de mim
 I not can understand why GERAL not like.PRES.3SG of mine
 ‘I cannot understand why nobody likes me.’
- b. Alguém postou no Facebook uma lista com curiosidades sobre
 somebody posted in-the Facebook a list with curiosities about
 a série Vaga-Lume, aquela que *geral* conhece
 the series Vaga-Lume that one that GERAL know.PRES.3SG
 ‘Someone has posted on Facebook a list of curious facts about the
 book series Vaga-Lume, the one everybody knows.’

- (20) a. nao gosto de la, me sinto como se *geral*
 not like.PRES.1SG of there me feel.1SG as if GERAL
 tivessem me observando
 had.PAST.SNJ.3PL me observing
 ‘I don’t like that place, I feel like everyone was observing me.’
- b. sabemos que *geral* curtem os bonés da nossa
 know.PRES.1PL that GERAL like.PRES.3PL the caps of-the our
 coleção
 collection
 ‘we know everyone likes the caps in our collection’

As for adjectives, variation can be observed both in gender, with masculine (21a) and feminine (21b) forms, and in number, with singular (21) and plural (22).

- (21) a. que delícia *geral* deixando de ser otário e
 what delight GERAL leaving of to-be douchebag.SG.MASC and
 respeitando a opinião alheia
 respecting the opinion of-other
 ‘what a delight to see everyone quitting being a douchebag and
 respecting each other’s opinions’
- b. mas o que *geral* ficou interessada mesmo foi na
 but the what GERAL was interested.SG.FEM actually was in-the
 receita do meu bolo de morango
 recipe of-the my cake of strawberry
 ‘[...] but what everyone was actually interested in was my strawberry
 cake recipe’
- (22) a. CAPCOM sabe que *geral* ficaram no mínimo
 C. knows that GERAL were in-the minimum
 frustrados
 frustrate.MASC.PL
 ‘CAPCOM knows that everyone was at least frustrated’
- b. [estou] chocada que *geral* tao passados pq
 be.1.SG shocked that GERAL are astonished.MASC.PL because
 a médica [...] combinou as perguntas com os governistas
 the doctor combined-3SG the questions with the governmentists
 ‘[I’m] shocked that everyone is astonished that the doctor [...] previously agreed on the questions with government supporters’

Therefore, with regard to agreement patterns, *geral* behaves as an IMP pronoun, with no ϕ -features demanding a fixed agreement marking.

4.2 Generic and arbitrary readings

As already noted in the introduction, *geral* may occur with both the generic and the arbitrary readings. Occurrences in (23) exemplify the generic reading. More particularly, in (23a) *geral* occurs with other pronominal forms (*todo mundo* ‘everyone’, *todos* ‘all’, *você* ‘you’) that are also interpreted as generic in Brazilian Portuguese. In (24), we find some cases of *geral* taking on arbitrary interpretation.

- (23) a. *Todo mundo tem aquele autor ou autora que geral*
 everyone has that author.MASC OR author.FEM that GERAL
conhece bem e todos falam bem, mas você nunca
 knows well and all speak.PRES.3PL well but you never
chegou a pegar qualquer uma de suas obras pra ler
 arrived.3SG to take any one of their books for to-read
 ‘Everyone has a male or female author that everyone knows well and
 praises, but you never actually got to reading one of their works.’
- b. *Quem vai perder o mercado muito em breve é a Samsung, que*
 who goes to-read the market very soon is the Samsung that
acha que geral não acompanha a evolução
 thinks that GERAL not follows the evolution
 ‘Samsung will soon be out of the market, for they think people do not
 keep up with innovation.’
- (24) a. *Nem preciso dizer que geral ficou boquiaberto ao ver nós*
 not need.1.SG to-say that GERAL was agape when to-see we
dois juntos
 two together
 ‘I don’t even have to say that everyone was agape when they saw us
 together.’
- b. *rolou um pipoco [e] geral correu achando que era o bope*
 happened an uproar and GERAL ran thinking that was the bope
 ‘there was an uproar, everyone ran thinking it as the BOPE [a
 division of the police]’

Therefore, as regards generic and arbitrary readings, *geral* also behaves as an IMP pronoun, exhibiting patterns that are similar to Swedish *man*, French *on* and Italian *si*, in line with the properties presented in Egerland (2003).

4.3 Syntactic distribution

As far as its distribution within the sentence is concerned, *geral* is compatible with different syntactic functions and may occur in positions associated with the nominative, accusative and oblique cases. In addition to the cases hitherto presented, in which *geral* occurs in a nominative position, it may also appear in an accusative position, as in (25), and in oblique positions, as in (26–28).

- (25) a. falaram que ele é uma simpatia e atendeu *geral* com o
said.3PL that he is a nice person and received.3SG GERAL with the
maior carinho
biggest gentleness
'they said he is really nice and received everyone with the utmost
gentleness'
- b. Quando aquela pessoa que elogia *geral*, vem e te elogia, não
when that person that praises GERAL comes and you praises not
rola emoção
happen emotion
'When someone who praises everybody comes and praises you, you
can't feel touched.'
- c. ensinei *geral* a como jogar R6
taught.1SG GERAL to how play R6
'I have taught everyone how to play R6.'
- (26) ele já tirou print e já enviou pra *geral*
he already took print and already sent to GERAL
'he has already taken a screenshot and sent it to everyone'
- (27) Hoje não me interessa a aprovação de *geral* apenas a minha
today not me interest the approval of GERAL only the my
felicidade
happiness
'Currently I don't care about being approved by everyone, but only about
my happiness.'

- (28) *Geral* sabe que ela fica cm *geral* pega o bonde todoooooo
 GERAL knows that she stays with GERAL takes the tram entire
 ‘Everyone knows she picks up everybody, she fools around with
 everybody!!!’

As for its syntactic distribution, *geral* therefore behaves as a φ -IMP pronoun, therefore contradicting what has been established concerning its agreement pattern and its reading, criteria that would rather align *geral* with IMP pronouns.

Table 2 illustrates the behavior of *geral* in comparison with that of IMP and φ -IMP pronouns.

Table 2: Properties of IMP, φ -IMP and *geral*

Properties	IMP	φ -IMP	<i>geral</i>
Is associated with a single form of agreement.	no	yes	no
May have either a generic or an arbitrary reading.	yes	no	yes
Only occurs in nominative positions.	yes	no	no

At first sight, we are thus facing a problem for the precise characterization of *geral*, since, if it were an IMP (i.e., devoid of φ -features and of Case projection), it ought to be licensed only in nominative positions. In the next section, however, I shall argue that this apparently contradictory behavior of *geral* regarding case marking is to be expected in view of the properties of the Brazilian Portuguese pronominal system.

5 Placing *geral* within the Brazilian Portuguese pronominal system

We have seen that *geral* behaves as IMP with regard to agreement patterns and to interpretation, but as φ -IMP with regard to the syntactic positions in which it may occur. As I will argue for, this inconsistent behavior is expected if we take into consideration that the impersonal version of *geral* is integrated into the Brazilian Portuguese pronominal system.

The paradigm of Brazilian Portuguese personal pronouns, especially in its vernacular varieties, licenses the occurrence of typically nominative pronouns in accusative and oblique positions (cf. Carvalho 2008, Galves et al. 2016, among others). This property is widely observed for the third-person nominative pronouns *ele/ela* (‘he/she’), which are frequently used in the object position, instead

of *o/a* ('him/her'), both in the singular and in the plural (29–30). The forms *você* 'you' (31) and *a gente* 'we' (32), which are typically nominative position pronouns, are also frequent in the accusative position, instead of *te* ('you') and *nos* ('us') – cf. Lopes 2003, Vianna & Lopes 2012, Lopes et al. 2013, Lopes & Rumeu 2015, among others.²

- (29) *testemunhas confirmam que viram ele no local do crime*
witnesses confirm.3PL that saw.3PL he in-the place of-the crime
'witnesses confirm that they saw him in the crime scene'
- (30) *eles levaram ela pro veterinário*
they took.3PL she for-the vet
'they took her to the vet'
- (31) *eu conheço você desde os seis anos de idade*
I know.1SG you since the six years of age
'I have known you since you were six years old'
- (32) *todos cumprimentaram a gente*
all.PL greeted.3PL we
'everyone greeted us'

Although less frequently, the first-person singular (*eu*, 'I') and first-person plural (*nós*, 'we') nominative forms also occur in typically accusative positions in-

²One of the reviewers of this chapter made the following remarks: "The fact that *a gente* (and other "pronouns" that derived from NPs, such as Colloquial Brazilian Portuguese *o pessoal*, *as pessoas* and *o povo* but also Standard European and Brazilian Portuguese *o senhor*) may appear in object position could be due to them still being felt as NPs (or NP-like). *Geral* as a pronoun can be related to a noun as well; this would be an alternative explanation". Even if this alternative explanation is correct, it does not exclude the need for an additional explanation, given that, unlike *o pessoal*, *as pessoas* and *o povo*, the items *geral* and *a gente* do not trigger any specific mark of agreement with verbs and adjectives (cf. §4). In Brazilian Portuguese (at least in my variety, spoken in the metropolitan region of Rio de Janeiro), *o pessoal* and *o povo* always trigger the masculine singular mark, whereas *as pessoas* triggers the feminine plural. These items have previously specified ϕ -features and, if analyzed as impersonal pronouns, should be treated as ϕ -IMP. It is not surprising that they can occur as direct objects, regardless of whether they are nouns or pronouns. What is surprising in this picture is that *geral* and *a gente* occur as direct objects (even though they derive historically from nouns), as they behave like items without ϕ -features and, as such, should occur only as subjects. The explanation I propose in this chapter is that the Brazilian Portuguese pronominal system, in contrast to the system of the languages exemplified in Egerland (2003) and Fenger (2018), allows typically nominative pronouns (including ϕ -featureless impersonal pronouns) in any syntactic function/position, regardless of such pronouns still being felt as NPs (or NP-like). Therefore, my proposal is not incompatible with the alternative explanation suggested by the reviewer.

stead of *me* ('me') and *nos* ('us'), either as subjects of ECM-constructions (see 33) or as direct objects (see 34).

- (33) muitas que não viram eu jogar falam como se me
 many_{FM} that not saw._{3PL} I to-play talk._{3PL} as if me
 acompanhassem
 follow._{PAST.SBJ.3PL}
 'many [fem.] who didn't see me play talk as if they had been following
 my career'
- (34) os funcionários [...] atenderam nós com muita gentileza
 the employees helped._{3PL} we with much attention
 'the employees helped us with the utmost attention'

These data make it clear that any typically nominative pronoun in Brazilian Portuguese may also occur in non-nominative positions, in contrast with European Portuguese. It is not yet clear how to account for the licensing of nominative pronouns in non-nominative positions. Avelar & Galves (2016), for example, argue that noun constituents in Brazilian Portuguese, including personal pronouns, may or may not exhibit case marking; in the specific situation of pronouns, one of the consequences of this variation would be precisely the use of the morphological nominative when the pronoun is not marked for case and, as such, is licensed for any syntactic position.

Irrespective of the formal explanation that may eventually account for these observations, the fact is that, in Brazilian Portuguese, typical nominative pronouns are licensed in positions associated with different cases. This must be the exact behavior of *geral* if it has already been integrated to the pronominal system of the language, at least in the grammar of those speakers that resort to it as an impersonalization strategy. If this analysis is on the right track, then *geral* must be characterized as an IMP pronoun (that is to say, a typically nominative pronoun devoid of φ -features), but that, differently from *man*, *on*, *si* and other IMP pronouns, occurs in different syntactic positions as a result of the peculiarities of the personal pronoun system in Brazilian Portuguese. The properties of *geral* therefore provides indirect evidence in favor of Fenger's (2018) perspective, for whom the distribution of personal pronouns in different positions in the sentence is better captured in terms of case marking, and not as a result of the syntactic function of the pronoun.

6 Occurrences of *geral* with specific and inclusive readings

Egerland (2003: 82–83) mentions situations in which Swedish *man* may refer to the first-person singular, as in (35a). French *on* and Italian *si* may also occur with a specific reading, but referring to the first-person plural, as in (35b) and (35c) respectively. Egerland terms these occurrences in Swedish, on the one hand, and in French and Italian, on the other, *specific reading* and *inclusive reading*, respectively.

- (35) a. Swedish (Arbitrary/Specific)
Man arbetade i två månader för att lösa problemet.
MAN worked for two months to solve the problem
- b. Italian (Arbitrary/Inclusive)
Si è lavorato per due mesi per risolvere il problema.
si has worked for two months to solve the problem
- c. French (Arbitrary/Inclusive)
On a travaillé pour deux mois pour résoudre le problème.
ON has worked for two months to solve the problem

Geral may also occur in a reading that refers to a group of individuals that includes the speaker, as in the case of pronouns *nós* and *a gente* ‘we’. In (36a), for example, *geral* refers to a specific group of students in which the speaker is included; in (36b), *geral* also refers to a specific group of people (the speaker’s family) that may or may not include the speaker.

- (36) a. Ter aulas no sábado é horrível! Eu preferia
to-have classes in-the Saturday is horrible I prefer.PAST.1SG
ter aulas até dezembro! Minha escola ainda não
to-have classes until December My school yet not
se pronunciou quanto a isso só falou que *geral* vai voltar
pronounce.3SG regarding to that only said.3SG that GERAL go return
dia 10.
day 10
‘Having classes on Saturday is awful! I would rather have classes into
December! My school has still not made an announcement about this
and only said that we students are back on the 10th.’

- b. Minha mãe ta internada, ae ontem tive que fazer a
 my mother is hospitalized so yesterday had.1SG that to-do the
 janta, fiz um macarrão que *geral* comeu horrores, só não
 dinner did.1SG a pasta that GERAL ate.3SG horrors only not
 falo como fazer porque to com pressa
 say.1SG how to-do because am with hurry
 ‘My mother is hospitalized, so I had to cook dinner yesterday. I made
 pasta that everyone ate until they were full. I only won’t teach you
 how to make it because I’m in a hurry.’

These occurrences raise the question of determining what licenses the specific and inclusive readings, since IMP pronouns do not exhibit ϕ -features. In other words, in the absence of ϕ -features, these pronouns ought to exhibit either the generic or the arbitrary readings, since they lack the necessary ingredient for establishing a reference to the first-person. Discussing this kind of data, Egerland argues that there are reasons for believing that inclusiveness may not be predictable:

[...] whereas there are principled reasons behind the restrictions on generic and “arbitrary” readings, it is more doubtful whether there are principles deciding whether impersonal pronouns are interpreted as including or excluding the speaker. Depending on the function of the message and the communicative strategies of the speaker, the impersonal construction is pragmatically open to a variety of uses. Some such uses will be inclusive, others exclusive, not for syntactic reasons, but due to contextual (extra-linguistic) factors. (Egerland 2003: 96–97)

In other words, the factors at stake to determine the specific and inclusive readings of impersonal pronouns are not of a grammatical, but rather of a pragmatic nature, and it is altogether impossible to predict one or another reading without taking extralinguistic factors into account.

Egerland acknowledges, however, that there is interlinguistic variation in the conditions for one or another reading to be triggered. The author exemplifies by comparing the Icelandic *maður* and the Swedish *man* impersonal pronouns, as in (37) and (38). In Icelandic, *maður* and *eg* (‘I’) may be coreferential in (37a), but not in (37b); in contrast, *man* and *jag* (‘I’) may be coreferential as much in (38a) as in (38b).

- (37) Icelandic (Jónsson 1992 *apud* Egerland 2003: 98)
- a. Eg vona að maður verðði ekki of seinn.
I hope that MAÐUR will-be not too late
'I hope I won't be late.'
 - b. Maður vona að eg verðði ekki of seinn.
MAÐUR hopes that I will-be not too late
'People hope I won't be late.'
- (38) a. Swedish (Egerland 2003: 98–99)
- Jag får hopas att man inte kommer för sent.
I may hope that MAN not comes too late
'I hope I won't be late.' / 'I hope they won't be late.'
 - b. Man får hoppas att jag inte kommer för sent.
MAN may hope that I not come too late
'I hope I won't be late'
'They hope I won't be late.'
'Let's hope I won't be late.'

In his discussion of this difference, Egerland (2003: 98–99) points to a suggestion by Jónsson (1992), according to which “a lower ranked feature cannot bind a higher ranked one”. If we consider that the first-person is ranked higher than the third-person, it becomes possible to explain why *eg* may bind *maður* in (37a), producing the specific reading of the impersonal pronoun; the opposite, as in (37b), preserves the generic reading of *maður* (which is a third-person), since, by the same rule, the impersonal pronoun cannot bind a first-person pronoun. In Swedish, however, *man* has no ϕ -features and therefore escapes the condition set forth by Jónsson, thus enabling it to take on a specific reading referring to the first-person, whether it is or not bound by *jag*.

What has been noted for *man* is precisely what happens with *geral*, as can be seen in (39), with the exception of the type of readings (specific in Swedish and inclusive in Brazilian Portuguese). *Geral* may be coreferential with the pronoun *a gente* ('we', which would also apply to the form *nós*, 'we'), both in (39a) and in (39b).

- (39) a. A gente tá esperando que geral não chegue atrasado
we are hoping that GERAL not arrive late
'We hope we won't be late.'
'We hope they won't be late.'

- b. *Geral* tá esperando que a gente não chegue atrasado.
 GERAL are hoping that we not arrive late
 ‘We hope we won’t be late.’
 ‘They hope we won’t be late.’

If Egerland’s proposal is on the right track, the possibility that *geral* presents an inclusive reading in some contexts is precisely due to the fact that it bears no φ -features (that is to say, it is an IMP pronoun, and not a φ -IMP pronoun). The factors that will determine such a reading are, however, of a pragmatic nature, since they depend on contextual (extralinguistic) factors that interfere in the reading of the sentence.

7 Conclusion

Although the grammatical and/or pragmatic factors that have triggered (or have been triggering) the use of *geral* as an impersonal pronoun demand further study, there is evidence that the process of its impersonalization resulted in the emergence of an IMP pronoun, with a behavior similar to that of *man*, *on* and *si* as described by Egerland (2003). The apparent inconsistency with regard to the syntactic positions it can fill, approximating the pattern of a φ -IMP pronoun, may be easily explained if we take into account the case-marking properties of the Brazilian Portuguese pronominal system, in which nominative pronouns are licensed in non-nominative positions. In other words, *geral* has been gaining ground as an inherently nominative IMP pronoun that is also licensed in accusative and oblique positions, as other nominative pronominal forms in Brazilian Portuguese. Besides the generic and arbitrary readings, *geral* can also have an inclusive reading (when it is the referential equivalent of a first-person plural, just as pronouns *nós* and *a gente*). The inclusive reading seems to depend on strictly pragmatic factors, as a result of the lack of φ -features.

Appendix A Source documents used in the examples

- (1) a. <https://votolegal.com.br/em/brunoramos> (accessed on 8 Sep 2018)
- b. <https://twitter.com/bbcbrasil/status/1020326263335079936> (accessed on 8 Sep 2018)
- c. https://vk.com/topic-73988417_37659003?offset=1380 (accessed on 4 Jul 2019)
- d. <https://www.facebook.com/watch/?v=1722931431109561> (accessed on 1 Jul 2019)

- (5) a. http://armazendoseubrasil.blogspot.com/2011_07_10_archive.html (accessed on 30 Sep 2021)
- b. <http://diariogaucha.clicrbs.com.br/rs/noticia/2009/01/parana-tem-mais-da-metade-das-lavouras-com-qualidade-media-e-ruim-2357367.html> (accessed on 30 Sep 2021)
- c. <https://www.diariodecuiaba.com.br/ilustrado/sobras/416243> (accessed on 30 Sep 2021)
- (19) a. <https://twitter.com/isasalviattii/status/928063135046660097> (accessed on 1 Jul 2019)
- b. <https://sonhandocomdarcy.wixsite.com/sonhandocomdarcy/single-post/2015/12/10/Top-5-Autores-Que-Eu-Nunca-Li> (accessed on 1 Jul 2019)
- (20) a. <https://answers.yahoo.com/question/index?qid=20170211204407AA4anTs> (accessed on 30 Sep 2021)
- b. <https://gramho.com/explore-hashtag/bonesjs> (accessed on 30 Sep 2021)
- (21) a. https://twitter.com/Mandy_Baessa (accessed on 1 Jul 2019)
- b. <https://blogqueideia.wordpress.com/2017/03/20/bolo-de-morango-a-receita/> (accessado em 1 Jul 2019)
- (22) a. <https://beta2.gamevicio.com/noticias/2021/09/resident-evil-3-deve-receber-atualizacao-em-breve/> (accessed on 30 Sep 2021)
- b. <https://twitter.com/Neni66576183> (accessed on 30 Sep 2021)
- (23) a. <https://sonhandocomdarcy.wixsite.com/sonhandocomdarcy/single-post/2015/12/10/Top-5-Autores-Que-Eu-Nunca-Li> (accessed on 1 Jul 2019)
- b. <https://www.tudocelular.com/samsung/noticias/n142917/analise-samsung-galaxy-a10-review.html> (accessed on 4 Jul 2019)
- (24) a. <https://www.wattpad.com/590285161-visão-de-cria-cap%C3%ADtulo-35/page/2> (accessed on 1 Jul 2019)
- b. <https://twitter.com/boombapx/status/772478713963372544> (accessed on 4 Jul 2019)
- (25) a. <https://www.facebook.com/PaparazzoRN/photos/a.517262558616376/773477566328206/?type=3&theater> (accessed on 3 Aug 2020)

- Chierchia, Gennaro. 1995. The variability of impersonal subjects. In Emmon Bach, Eloise Jelinek, Angelika Kratzer & Barbara H. Partee (eds.), *Quantification in natural languages*, 107–143. Dordrecht: Kluwer.
- Chomsky, Noam. 1995. *The Minimalist program*. Cambridge, MA: MIT Press.
- Cinque, Guglielmo. 1988. On *si* constructions and the theory of Arb. *Linguistic Inquiry* 19. 521–581.
- D’Alessandro, Roberta & Artemis Alexiadou. 2003. Inclusive and exclusive impersonal pronouns: A feature geometrical analysis. *Rivista di Grammatica Generativa* 27. 31–44.
- Egerland, Verner. 2003. Impersonal pronouns in Scandinavian and Romance. *Working Papers in Scandinavian Syntax* 71. 75–102.
- Fenger, Paula. 2018. How impersonal does *one* get? A study of *man*-pronouns in Germanic. *The Journal of Comparative Germanic Linguistics* 21. 291–325.
- Galves, Charlotte, Juanito Avelar, Dorothy Brito, Danniell Carvalho, Célia Lopes & Leonardo Marcotulio. 2016. Morfossintaxe e uso dos pronomes pessoais na sincronia e na diacronia do português brasileiro. In Lucrécio Araujo de Sá Junior & Marco Antonio Martins (eds.), *Rumos da linguística brasileira no século XXI: Historiografia, gramática e ensino*, 123–153. São Paulo: Blucher.
- Hoekstra, Jarich. 2010. On the impersonal pronoun *men* in modern West Frisian. *Journal of Comparative Germanic Linguistics* 13. 31–51.
- Jónsson, Jóhannes Gíslí. 1992. *The pronoun maður in Icelandic*. University of Massachusetts, Amherst. (Doctoral dissertation).
- Krifka, Manfred, Francis J. Pelletier, Gregory N. Carlson, Alice ter Meulen, Godehard Link & Gennaro Chierchia. 1995. Genericity: An introduction. In Gregory N. Carlson & Francis Pelletier (eds.), *The generic book*, 1–24. Chicago: The University of Chicago Press.
- Lopes, Célia Regina dos Santos. 2003. *A inserção de a gente no quadro pronominal do português: Percorso histórico*. Frankfurt am Main/Madrid: Vervuert/Iberoamericana. DOI: 10.31819/9783865278494.
- Lopes, Célia Regina dos Santos, Leonardo Marcotulio & Juliana Vianna. 2013. Agreement patterns with *a gente* in Portuguese. *Journal of Portuguese Linguistics* 12. 125–149.
- Lopes, Célia & Márcia Rumeu. 2015. A difusão do “você” pelas estruturas sociais carioca e mineira dos séculos XIX e XX. *LaborHistórico* 1. 12–25.
- Neeleman, Ad & Fred Weerman. 1999. *Flexible syntax: A theory of case and arguments*. Dordrecht: Springer.
- Vianna, Juliana & Célia Lopes. 2012. A variação entre nós e a gente: Uma comparação entre o português europeu e o brasileiro. *Revista do GELNE* 14. 95–116.

Chapter 8

Toward the prevalence of a personal use of impersonal *uno* in Colombian Spanish

Rafael Orozco 

Louisiana State University

Luz Marcela Hurtado

Central Michigan University

Marianne Dieck

Universidad de Antioquia

This variationist study explores the alternation between the pronouns *uno* ‘one’ and *yo* ‘I’ in Colombian Spanish using data from the PRESEEA Medellín corpus. We test the hypothesis that, in Colombian Spanish, *uno* is being recast to the point that it functions as a variant of the first-person singular subject pronoun *yo*. We aim to go beyond the well-established *pronombrista* line of subject pronoun research with a variationist analysis of the alternation between *uno* and *yo* that examines, among other things, the role of stance and the focus of attention in terms of predictors that include transitivity, verb semantics, coreference, type of discourse, and sentence polarity. Our findings uncover the strongest conditioning effect of tense, mood and aspect as well as robust effects of transitivity, discourse genre, polarity, and type of preceding subject. Thus, the *uno/yo* alternation constitutes a linguistic variable in its own right.

1 Introduction

Uno ‘one’ is a multifunctional Spanish subject pronoun with third person singular morphosyntax whose pragmatic domain extends to other grammatical persons, mainly the first. *Uno* has traditionally been associated with the expression of



impersonality because its association with the first person – both singular and plural – involves connotations of genericity. Studies on impersonality employing semantic, syntactic, and pragmatic approaches have shown that impersonals are versatile with respect to their referential interpretation (Casielles Suárez 1996: 376, Hernanz 1990: 160, Hurtado & Gutiérrez-Rivas 2016). Their versatility is such that impersonals can coincide and fulfill the tripartite function of (1) hiding an agent or reducing the speaker's prominence (Barrajón 2005, Gómez Torrego 1992, Haverkate 1987, Hernanz 1990, Hollænder Jensen 2002, Muñiz Cachón 1998, Ricos Vidal 2002); (2) integrating the speaker, involving the interlocutor, and alluding to a group (Fernández 2008, Fernández Ramírez 1986, Gelabert-Desnoyer 2008, Muñiz Cachón 1998); and (3) denoting all speakers; i.e., humankind (Company Company & Pozas Loyo 2009, Fernández Soriano 1999, Siewierska 2008). That is, impersonals display different degrees of specificity and inclusion of the agent. However, some scholars such as Company Company & Pozas Loyo (2009: 1206) also raise the possibility of the complete reduction of the impersonal reference when textual elements such as first-person singular subjects appear; that is, a more personal reading as in the case of example (1), from of our dataset.¹

- (1) eh si es un día laboral / pues *uno se levanta* / obviamente [Ø] *se baña* / [Ø] *se va* para el trabajo / ah no / [Ø] *despacha* a los hijos / [Ø] *llevo* la señora al trabajo / al trabajo de ella obviamente / y [Ø] *me voy* ya para el trabajo mío y [Ø] *ya regreso* después de las seis de la tarde. (MEDE_H12_3)
- 'eh if it's a weekday / well *one*² *gets up* / obviously [*one*] *bathes* / [*one*] *goes* to work / ah no / [*one*] *sends* the children off / [*I*] *take* my wife to work / to her work obviously / and [*I*] *go* to work and [*I*] *return* after six in the evening.'

In (1) we observe that the use of *uno* 'one' allows a definite reading in connection with the first person and the speaker's perspective (Haverkate 1985: 19), depending on the contextual cues (Gelabert-Desnoyer 2008).

This variationist study goes beyond the traditional analysis of the alternation between null and overt subjects in Spanish addressed in the vast *pronombrista* or *pronombrismo* literature (cf. Carvalho et al. 2015, Orozco & Hurtado 2021a, Otheguy & Zentella 2012; inter alia), i.e., the research strand devoted to the study

¹Throughout this chapter, we have indicated null subjects using [Ø] in the Spanish examples. In the corresponding English translations, null subjects are indicated within brackets as well.

²In most cases, Spanish *uno* would translate as (impersonal) *you* in colloquial contemporary English. However, given that our paper deals with the *uno/yo* alternation in Spanish, we have translated *uno* as (the less commonly used English equivalent) *one*.

of subject pronoun expression (SPE). We explore the variable alternation between the subject pronouns *uno* ‘one’ and *yo* ‘I’ in the Spanish of Medellín, Colombia. Our study is motivated by recent findings from Colombian Spanish (Hurtado 2015, Hurtado & Gutiérrez-Rivas 2016, Hurtado & Ortega-Santos 2019, Orozco & Hurtado 2021a,b) that report the impersonal pronoun *uno* being used primarily with referential interpretation in which the connection with the I-speaker as a marker of positioning and subjectivation of discourse prevails.

Given that the interpretation of *uno* as an alternate to *yo* predominates within the range of referential interpretations of *uno*, in the present investigation, we seek to explore whether this role constitutes an indication that *uno* behaves in a more definite and personal way in Colombian Spanish. As our findings will show, the *uno/yo* alternation constitutes a linguistic variable in its own right with its own internal conditioning. The remainder of this chapter is organized as follows: We discuss impersonality and the nature of *uno* in the next section. §3 is devoted to the methodological approach, including descriptions of the speech community, the corpus, the dataset, the envelope of variation, our research questions and hypothesis. §4 is dedicated to the presentation of our results. §5 and §6 respectively present the discussion of findings and the conclusion.

2 Background

The frequent occurrence of *uno* with first person singular interpretation has been attested in studies that integrate semantic-pragmatic and social predictors. For instance, Morales (1995) finds that Spanish/English bilingual speakers favor impersonal *tú* ‘you’ and *uno* ‘one’ in the narration of personal experiences. Fernández (2008) points out the influence of the type of information that is transmitted and the speaker’s access to the source of information. According to this scholar, speakers generalize by using *uno* and *tú* departing from the perspective of their own experience and by using the clitic *se* to include other generally accepted voices or opinions. Bassa Vanrell (2013) also indicates that the predominant referential interpretation for *uno* in Puerto Rico and the Dominican Republic is linked to the speaker’s situation. Additionally, with regard to the social conditioning, Guirado (2011), Rodríguez Alfano (2004), and Guantiva Acosta (2000) correlate the use of *uno* with the lower socioeconomic levels in Caracas, Monterrey, and Bogotá, respectively.

Impersonality in Colombian Spanish has been explored in the varieties spoken in the Andean and Caribbean regions as well as among diasporic speakers residing in the United States. These studies have uncovered that *uno* is the predominant form of impersonalization, as follows. In the Andean varieties, *uno* with

a frequency of 51% and *se* with 43.5% are the most frequent in Bogotá (Hurtado 2016: 184). In Medellín, *uno* with a frequency of 47.9% and *se* with 27.5% are the most used impersonals (Dieck 2016: 160). In the Caribbean city of Barranquilla, the use of *uno* prevails with a frequency of 61.6%, followed by *se* with 25.1% (Hurtado & Gutiérrez-Rivas 2016: 45). Moreover, in Westchester county and Albany, NY, U.S., among bilingual speakers originally from Bogotá, Valle del Cauca, Antioquia, and Quindío *uno* registers a frequency of 62.6% followed by *se* with 25.5%, and impersonal *tú* 'you' with 11.9% (Ramírez 2007: 152).

The predominant use of *uno* with direct reference to the speaker's situation – i.e., first person singular interpretation – has been analyzed as a positioning and subjectivization of discourse strategy (Hurtado 2015, Hurtado & Gutiérrez-Rivas 2016, Ramírez 2007). The referential interpretation of *uno* as a substitute for *yo* – referring only to the speaker – dominates in Bogotá (87%, weight 0.85) together with a high overt pronominal rate (83%). This dominance of *uno* reported by Hurtado (2015) appears to increase the attention to the subject's referent. Likewise, *uno* occurs mainly with verbs whose lexical content indicates feelings, states, and opinions; that is, verbs linked to the speaker's subjectivity (Hurtado 2015: 136-137). In Barranquilla, *uno* predominantly functions as a variant of *yo* (89.3%, weight 0.73). Moreover, *uno* appears to denote positioning by being used mainly with verbs that indicate the speaker's knowledge, evaluation, feelings, and location (Hurtado & Gutiérrez-Rivas 2016: 56). These studies indicate that semantic constraints such as the verb's referential interpretation and semantic class condition the use of *uno* in Barranquilla and Bogotá.

Notwithstanding, the variable alternation between *uno* and *yo* remains unexplored in variationist sociolinguistics. Among the few existing investigations, Flores-Ferrán (2009) analyzes this phenomenon among speakers of Colombian, Cuban, Dominican, Mexican, Puerto Rican, and Uruguayan Spanish residing in metropolitan New York City. She finds that the use of *uno* is conditioned by several predictors including semantic clause type, semantic verb type, discourse type, and speaker's age. Thus, the current investigation addresses the dearth of research on the *uno/yo* alternation and is motivated by recent findings on subject pronoun expression in Barranquilla and Medellín. Hurtado & Ortega-Santos (2019) show the effect of high transitivity verbs and the focus of attention on the object in disfavoring the use of overt *uno* in Barranquilla (probability weight of 0.47 for monotransitives and 0.37 for ditransitives). They provide evidence that when the number of participants increases, the competition for the focus of attention also increases, which disfavors the use of overt subject pronouns (previously suggested by Aijón Oliva & Serrano 2013 and Posio 2011). This result suggests its potential influence on the *uno/yo* alternation. Orozco & Hurtado (2021a:

14) report that third person singular pronouns favor overt subjects in Medellín (probability weight 0.64, 42%). A subsequent analysis (Hurtado & Orozco 2022), separating *uno* from the other third person singular pronouns (*ella* 'she' and *él* 'he'), reveals that the favorable effect of third person pronouns on overt subjects is driven by *uno* (weight 0.83, 60%) whereas the other third person singular pronouns have a neutral effect (0.49). Thus, *uno* registers the highest overt pronominal rate in Medellín (60%) while the other third person singular pronouns *ella* and *él* combined have a modest pronominal rate of 27%.

3 Methodology

This section describes the speech community constituted by the city of Medellín, the corpus, and the dataset analyzed. It also presents the research questions and hypothesis that guide this investigation, and describes the predictors explored as well as the envelope of variation.

3.1 The speech community and the dataset

Medellín, founded in 1675, has constituted one of Colombia's main industrial centers since the early 20th Century. Between 1890 and 1950, the process of textile industrialization and the production of beer, ceramics, glass, tobacco, and coffee promoted an increment of the blue-collar population as well as the urbanistic growth of the city (Botero 1996: 8-10). According to the 2018 census, Medellín has a population of 2,372,330 out of which 59% were born in the city, 37% were born elsewhere in Colombia, and 2.2% abroad. This reflects the migration and displacement to urban centers Colombians suffered in the latter years of the 20th century because of social unrest and lack of economic opportunities (DANE 2019, Castañeda 2005: 82).

Medellín Spanish belongs to what Montes Giraldo (1982) classified as Western Andean Colombian Spanish. The city is located in the department of Antioquia, where two dialectal varieties, the Andean and the Caribbean converge. Because of its geographical location, Medellín developed as an isolated city with differentiated traditions, religious orientation, and family values (Fernández Acosta 2020: 95). The Spanish of this region is characterized by the extensive use of the second person singular pronoun *vos* 'you' across ages, socioeconomic levels, and registers, which has been analyzed not only as an indicator of an egalitarian and open society (Montes Giraldo 1967: 25), but as an expression of local identity and belonging to the region (Fernández Acosta 2020: 97).

Our dataset was culled from the *Proyecto para el Estudio Sociolingüístico del Español de España y de América (PRESEEA) Medellín Corpus* collected between 2007 and 2010 (González-Rátiva 2008). The PRESEEA Medellín corpus interviews contain important cultural and sociolinguistic information about the people, customs, and life in the city. They were carried out using topics prepared to elicit conversation. The predominant form of address used by interviewers was *usted* (formal ‘you’). Questions deal with a variety of topics, from weather, the neighborhood, Medellín’s people, problems in the city, and transportation, to more personal topics including family, work, daily routines, traditions, and holidays. The interviews ended with narrations of a dream or a scary event. We used a subset of 40 of the 119 socially stratified interviews in the corpus, which correspond to 20 women and 20 men whose ages ranged from 15 to 85 years old at data collection time (See Table 1). All consultants were born in Medellín or in the surrounding region.

Table 1: The speakers.

Gender	Socioeconomic level			Total
	Low	Mid	Mid-High	
Women	8	6	6	20
Men	8	6	6	20
Total	16	12	12	40

3.2 Research questions and hypothesis

The present investigation is guided by the following research questions and a main hypothesis. We seek to answer three main research questions.

1. How are *uno* and *yo* distributed when they appear interchangeably within a single speech turn, and what predictor variables condition their alternation?
2. In what discursive contexts are *uno* and *yo* used interchangeably?
3. What functions are most commonly assumed by each pronoun, and how do *uno* and *yo* construct/create the discursive subject?

Concurrently, we aim to probe the following main hypothesis: *The predominance of uno as a substitute of yo, within the range of referential interpretations, constitutes an indication that uno behaves particularly in a more definite and personal way in Colombian Spanish.* Our hypothesis and research questions were informed by recent investigations of subject pronoun expression and impersonality in Colombian Spanish (Dieck 2016, Hurtado 2015, Hurtado & Gutiérrez-Rivas 2016, Hurtado & Ortega-Santos 2019, Olave-Arias et al. 2021, Orozco 2018, Orozco & Hurtado 2021a).

3.3 Predictor variables explored

To answer the above research questions and probe our main hypothesis, we explore the effects of six predictor variables which provide information about the focus of attention and stance. These predictors operate at different morphosyntactic and discourse levels. They can be divided into two main categories:

- Predictors related to the whole clause: *discourse genre, type of preceding subject, polarity, and attenuation procedure and genericity inducers, and*
- Predictors related to the verb phrase: *transitivity and verb tense, mood and aspect [TMA].*

As with our research questions and hypothesis, our choice of these predictor variables was guided by the findings of relevant investigations of the expression of impersonality (De Cock 2014, Flores-Ferrán 2009, Guirado 2011, González Vergara & Rojas 2012, Hernanz 1990, Hurtado & Ortega-Santos 2019, Orozco & Hurtado 2021b, Repede & Leon-Castro 2019).

To analyze the role of the focus of attention, we considered the relationship between the predictors analyzed and some of Hopper & Thompson's transitivity components (1980: 252). As done by Aijón Oliva & Serrano (2013), Hurtado & Ortega-Santos (2019), and Posio (2011), we probe the premise that transitivity relates to the subject referent's focus of attention. We also test the notion that high transitivity components (e.g., perfective aspect, realis mode, two or more participants, and affirmative clauses) increase the possibility that attention focuses on the action expressed by the verb and on the object's reference.

3.3.1 Clause-related predictors

3.3.1.1 Discourse genre

We explore the effect of discourse genre seeking to provide a detailed analysis of the role of the speaker's stance, understood as the manifestation of attitudes,

feelings, judgment or commitment to their speech (Biber & Finegan 1988). In so doing, we analyzed the impact of various discursive modes to measure the effect of realis and irrealis events – one of Hopper & Thompson's (1980) transitivity components – on the *uno/yo* alternation in terms of the following three factors:

Narrative: personal experiences, daily routine, events that happened in the past or are currently happening (as in example 1),

Opinion: argumentative discourse about the city and its people (example 2), and

Hypothetical situations: projected and contrary to fact actions (example 3).

Hurtado & Ortega-Santos (2019: 53) found a predominant use of *uno* in Barranquilla, Colombia, with factual events, according to the following distribution: narration of personal experiences (80%), general facts (13%) and hypotheses and conjectures (7%). Thus, we seek to determine whether a) this trend extends to contexts in which *uno* and *yo* alternate, or b) the use of *uno* predominates with more irrealis types of discourse (hypothetical situations, opinions) – commonly associated with generalizations or impersonal interpretations – whereas *yo* predominates with more subjective discourse (personal experience narratives), as illustrated in (2) and (3).

- (2) *Yo soy partidario de eso / yo creo en eso / que a veces la suerte influye en muchas cosas / aunque no debe depender uno / eeh / completamente de eso.* (MEDE_H23_5)

'I'm a supporter of that / I believe in that / that sometimes luck influences many things / although one shouldn't depend / eeh / completely on that.'

- (3) E: ¿cómo cree que sería vivir en otro barrio?

'What do you think it would be like to live in another neighborhood?'

I: pues / la verdad [Ø] no sé / porque nunca [Ø] he vivido en otro barrio / pero yo digo que eso depende como la zona en la que uno viva / depende de los vecinos / como de la interacción que uno tenga con ellos. (MEDE_M03_5)

'well / the truth [I] don't know / because [I]'ve never lived in another neighborhood / but I say that it depends on the area where one lives / it depends on your neighbors / like on the interaction one has with them.'

3.3.1.2 Type of preceding subject

This predictor variable explores the possibility that an immediately preceding overt or null subject triggers the occurrence of either *uno* or *yo*. The type of preceding subject predictor relates to priming, a construct based on the premise that speech is cognitively patterned according to preceding discourse (Travis 2007). Though priming was earlier considered to be of linguistic nature, it has been reclassified as a cognitive constraint (cf. Labov 2010, Tamminga et al. 2016). In *pronombrista* studies, priming explores the possibility that the occurrence of a prior overt or null subject triggers further pronoun expression or omission (Cameron & Flores-Ferrán 2004). Studies on Colombian Spanish (Torres Cacoullos & Travis 2019, Travis 2005, 2007), Peninsular and Puerto Rican Spanish (Cameron 1994, Flores-Ferrán 2002) have found significant priming effects on SPE and its intersections with coreference, distance from the preceding coreferential subject, and type of discourse.

Priming is illustrated in (4) where the first overt pronominal subject (*uno cogía* ‘one would take’) triggers three successive overt *uno* pronominal subjects. Then, the null pronominal subject in $[\emptyset]$ *tiene* ‘[one] has’ triggers another null subject in $[\emptyset]$ *he visto* ‘[I] have seen.’

- (4) ... en ese entonces *uno cogía* de un barrio al centro un bus // me recuerda mucho eso porque cuando *uno llegaba* a / a la / al río Medellín // habían unos puentes // que eran como muy inclinados / entonces el bus subía y cuando bajaba // *uno sentía* un vacío muy profundo // entonces eso le causaba a uno como susto / como miedo // y en cambio ahora ya los puentes son como diferentes / ya *uno no siente* pues esos vacíos // $[\emptyset]$ *tiene* como mejor / están mejor diseñados // y pues muy cambiado / y hay espacios también peatonales que no había // eeh lo que sí $[\emptyset]$ *he visto* mucho es que / ha habido mucho aumento de carros / de motos de // de vehículos... (MEDE H21-2)

‘... in those days *one would take* a bus from a neighborhood to the city center / I remember that a lot because when *one arrived at* / at the / at the Medellín River // there were some bridges // that were like very steep / then the bus would go up and when it went down // *one felt* a very deep vacuum // then that would give you like a fright / like fear // and instead now the bridges are like different / *one no longer feels* well those gaps // *[one] has* like better / *[they]* are better designed // and well very changed / and there are pedestrian spaces too that didn’t exist // eeh what *[I] have seen* a lot is that / there has been a great increase in cars / in motorcycles of // in vehicles.’

Given that priming is a robust SPE predictor in Medellín, we aim to explore its effect on the alternation between *uno* and *yo* in that speech community. Guided by prior *pronombrista* research (Orozco 2018), we tested the effect of perseverance using three factors to code our tokens according to their preceding subject as follows: 1) pronominal overt subject, 2) null subject, and 3) other subjects. The latter factor includes lexical subjects as well as demonstratives.

3.3.1.3 Polarity

This predictor probes the effect of affirmative and negative statements on the *uno/yo* alternation. When classifying clauses into affirmative or negative, we classified as negative those produced with utterances of negation and negative quantifiers, as in (5). This example contains three clauses where negative polarity occurs as the speaker transitions from *uno* being the subject to null cases of *yo* as the subject.

- (5) ... hay veces *no puede uno hacer* / [\emptyset] *no puedo hacer una presencia pues física* / como ir todos los días *no no no* / [\emptyset] *no puedo*. (MEDE H13-2)
'...there are times *one cannot do* / [*I*] *can't make* like a physical presence / like going every day *no no no* / [*I*] *can't*.'

Flores-Ferrán (2009: 1817) found an apparent polarity effect regarding the occurrence of *uno* and *yo* in personal experience narrations during therapeutic interviews. She found that Hispanic residents of New York City and Central New Jersey favored *uno* over *yo* in neutral information clauses (those expressing general information considered neither positive nor negative in nature). Flores-Ferrán also found that *uno* occurred more frequently (41.2%) in clauses that contain negation or were framed in negative or conflicted situations than in positive contexts (6.5%).

3.3.1.4 Attenuation procedure and genericity inducers

Besides the above linguistic predictors, we explore the effects of attenuation procedure and genericity inducers, a pragmatic predictor. Impersonalization has been studied as an enunciative mitigation device, related to evidentiality,³ that reduces the speaker's responsibility for an utterance (Caffi 2007). Thus, we sought

³By means of impersonalization mechanisms, speakers can attenuate the deictic origin of their utterances, expressing that they come from an impersonal or even a more objective source, and reducing their responsibility for the content of their words or for their actions.

8 Toward the prevalence of a personal use of impersonal *uno* in C.S.

to discover whether speakers employ the same attenuation procedures with *uno* and with *yo*. We tested this predictor using a classification based on the PRE-SEEA *Guide for the study of attenuation* (Cestero & Alfano 2021), which includes linguistic and pragmatic procedures, some with argumentative value and others with interactional value (Briz Gómez 1995). According to Cestero Mancera (2020: 367), these strategies are organized as a continuum, “from the speaker’s greater to lesser commitment, from correcting the utterance or the action to defocalization.” We analyzed the effects of a series of factors that include the following:

1. Resources that correct or reformulate: *o sea* ‘that is,’ *es decir* ‘that is to say,’ *bueno* ‘well.’
2. Resources that limit or restrict: concessivity (*sí* ‘yes,’ *pues* ‘so,’ *es verdad* ‘it’s true’ followed by *pero* ‘but’); expressions with conditional meaning (*si* ‘if,’ *siempre que* ‘as long as,’ *con tal (de) que, a menos que* ‘unless that,’ *a no ser que* ‘if not,’ *mientras* ‘whereas’).
3. Resources that downgrade: verbs, verb contractions, and modal particles that express doubt or probability (*creer* ‘believe,’ *parecer* ‘seem,’ *imaginar* ‘imagine,’ *ser posible* ‘be possible,’ *ser conveniente* ‘be convenient,’ *a lo mejor* ‘perhaps,’ *quizás, tal vez, de pronto* ‘maybe,’ *dizque* ‘supposedly,’ *probablemente* ‘probably,’ *posiblemente* ‘possibly’); expressions that feign uncertainty, incompetence or ignorance (*no sé cómo decirte* ‘I don’t know how to tell you,’ *que yo sepa* ‘as far as I know,’ *no estar seguro* ‘not to be sure,’ *seguramente* ‘surely,’ *yo qué sé* ‘what do I know,’ *no creer ser capaz* ‘not think to be able to’); modal use of verb tenses (use of the conditional and the imperfect for politeness, use of the future of probability in present tense contexts).
4. Resources that minimize or blur the quantity or the quality of what is said: diminutive suffixes; downgrading quantifiers; approximators or diffusers of meaning (*poco* ‘little,’ *algo* ‘some,’ *un tanto/un poco* ‘a bit,’ *más o menos* ‘more or less,’ *medio* ‘kind of,’ *como* ‘like,’ *hay veces / a veces* ‘sometimes,’ *hasta* ‘even/kind of,’ *algo así* ‘something like that’).
5. Resources that justify: *es que* ‘well/so,’ *porque* ‘because,’ *lo que pasa es que* ‘what happens is that,’ *por así decirlo* ‘to say so,’ *por decir algo* ‘to say something,’ *por decir* ‘just to say,’ *ni qué decir* ‘it goes without saying.’

6. Resources that involve the addressee: particles and expressions of control of interaction (*¿no? ¿eh? ¿sabe?* ‘you know,’ *¿cierto?* ‘really’); ways of addressing the interlocutor (*vea* ‘look,’ *mire* ‘see, watch’ *escuche* ‘listen,’ *hombre* ‘man,’ *venga* ‘come on,’ *hermano* ‘brother’).
7. Resources that impersonalize and defocalize: impersonal constructions (*se, uno, tú, 3PL* impersonal, and the collective use of 1PL); direct speech; objectivization using modal discourse particles (*obviamente* ‘obviously,’ *la verdad* ‘the truth,’ *verdaderamente* ‘truly,’ *realmente* ‘really,’ *normalmente* ‘normally,’ *notablemente* ‘notably,’ *legalmente* ‘legally’).

We did not include the cases of direct speech in the calculations, because there was only one case of citations with *uno* out of 13. Instead, we included other impersonal inducers such as adverbial constructions of place, time, or mood.

Within this predictor, we also explored whether *uno* and *yo* appeared in post-verbal position, as this is a non-prototypical position in Spanish that could soften the agent-patient relationship (Aijón Oliva & Serrano 2013: 310) and reduce agency, making the subject less prominent (Serrano 2012). Due to its complex nature, we tested this predictor as a random effects factor.

3.3.2 Verb-related predictors

3.3.2.1 Transitivity

We use transitivity, a central property of language use, to explore the relationship between the number of participants (agent and object) and the competition for the focus of attention and its influence on the choice between *uno* and *yo*. According to Hopper & Thompson’s theory of transitivity (1980), a smaller number of participants would correlate with a lesser degree of transitivity, and the focus of attention would remain on the subject.

Given that Posio (2011), Hurtado & Ortega-Santos (2019), and Orozco & Hurtado (2021a) have found that low transitivity verbs (when the subject does not compete with an object for the focus of attention) such as *vivir* ‘live,’ *trabajar* ‘work,’ and *ser* ‘be’ favor the occurrence of overt pronominal subjects, this study aims to determine whether this reduced competition for the focus of attention influences the *uno/yo* alternation in the same way. Thus, to test transitivity, we divide factors into two main types:

1. low transitivity clauses with one participant: intransitive verbs, reflexives, and epistemic/evidential verbs in clauses that introduce a complement clause;

2. higher transitivity clauses in which the competition for the focus of attention varies between two or three participants: transitive verbs, transitive verbs with null objects, and transitive verbs with prepositional complements.

Among transitive verbs, those with prepositional complements such as *uno pensaba en* ‘one thought of’ (example 6) and null object verbs such as *comer* ‘eat’ in *[uno] come* ‘[one] eats’ (example 7) illustrate cases in which transitivity decreases. Although the number of participants is maintained, not expressing the direct object decreases the subject’s competition for the focus of attention (Posio 2011, 2013).

- (6) *uno pensaba en sus muñecas / en sus trastecitos / en las comiditas que [∅] hacía / todo así como tan / inocentemente tan rico* (MEDE_M32_3)
 ‘one would think of one’s dolls / of one’s toy dishes / of the little meals that [one] would make / everything like so innocently nice’
- (7) *yo no desprecio a un anciano ni un pobre / porque eso es un pecado / porque otro día es uno / porque uno tiene subidas y bajadas / si uno es bien conchudo y [∅] come solo / también / [∅] se ve pidiendo.* (MEDE_M31_2)
 ‘I do not despise an elderly or a poor person / because that is a sin / because another day it’s one / because one has ups and downs / if one is very shameless and [one] eats alone / also / [one] finds himself begging.’

We classify within one-participant clauses the epistemic/evidential verbs that introduce a complement clause following Thompson & Hopper’s (2001: 31) criteria. Thus, in clauses with such verbs as *saber* ‘know,’ *pensar* ‘think,’ *ver* ‘see,’ and *recordar* ‘remember’ as well as such clauses as *piensa uno que* ‘one thinks that’ or *se da uno cuenta que* ‘one realizes that’ (example 8), the complement clause is not counted as a participant or object of the main clause.

- (8) *Otro susto / es cuando [∅] estuve viajando en avión / ¡ay hermano! / eso piensa uno que / ¡cuando hay turbulencias! / piensa uno que se va a desbaratar / ahí se da uno cuenta que el problema de uno es no tener / no saber volar.* (MEDE_H13_2)
 ‘Another scare / is when [I] was traveling by plane / oh brother! / one thinks that / when there is turbulence! / one thinks that it [the plane] is going to fall apart / that’s when one realizes that one’s problem is not having / not knowing how to fly.’

3.3.2.2 Verb tense, mood and aspect (TMA)

Informed by prior findings showing that TMA conditions impersonality, we explore the effect of this predictor on the *uno/yo* alternation in Medellín. Flores-Ferrán (2009: 1819) found that *uno* occurred more frequently with infinitives, and De Cock (2014: 203) found correlation between the generic readings of *uno* and the present tense. We divided verb paradigms into six factors as follows.

- the present indicative,
- the imperfect indicative,
- the preterit of indicative,
- subjunctive forms,
- infinitives and gerunds as one factor,
- other paradigms.

In our data, the category of infinitives includes

- preposition + subject + infinitive constructions (e.g., *Es una montaña buena para uno subirla* ‘It is a nice little mountain for one to climb,’ *La costumbre de uno estar aquí* ‘The habit of one being here’);
- infinitives in independent sentences (*Sería uno saber manejar* ‘It would be one knowing how to drive’);
- infinitives in predicative complements of perception verbs (*Yo no veo donde divertirse uno* ‘I don’t see where to enjoy (oneself)’);
- complements of adverbs (*¡Después de haber trabajado [uno] tantos años!* ‘After one having worked for so many years!’); and
- subordinated expressions of desire, influence or need (*Es mejor irse uno caminando desde el barrio* ‘it’s better for one to go walking from the neighborhood’).

Other paradigms include the conditional, the perfect tenses, and the futures. These forms were initially coded as independent factors. However, we amalgamated them due to low token counts and similar tendencies found in preliminary analyses. Although this configuration departs from the traditional configuration

practiced in *pronombrista* studies, it fits the nature of our linguistic variable. TMA contributes to test whether the use of *uno* is linked to less definite or punctual temporal reference, which could promote a generic reading (De Cock 2014: 203, Hernanz 1990: 156), whereas perfective actions do not promote generic readings (Monge 2002: 355). Moreover, analyzing the distinction between perfective and imperfective actions – one of Hopper & Thompson’s (1980) transitivity components – will also provide information about the effect of the focus of attention on the subject.

3.4 The envelope of variation and the analysis

The envelope of variation employed here adheres to the Principle of Accountability (Labov 1972: 72). We set the envelope of variation for this study in terms of the exchangeability of *uno* and *yo* when these pronouns appear in a single speech turn in clauses constituting answers to direct, personal questions explicitly asking the speaker to talk about her/himself. Example (9) below illustrates the envelope of variation and the *uno/yo* alternation. The fieldworker asked about the speaker’s family. The speaker’s answer contains nine clauses whose pronominal subject is either *yo* or *uno*. In the first of these clauses ($[\emptyset]$ *subí* ‘[I] went upstairs’), the null subject is *yo* ‘I’. It is followed by four more instances of *yo* as the subject. Then, *uno* ‘one’ appears as the subject of the sixth clause (***uno se achanta*** ‘one backs down’), and *yo* is the subject again in the last two clauses.

- (9) E.: ¿y qué le dijo su familia de eso?
 I.: no, las niñas hermano / fue que la señora es la que me vio cuando $[\emptyset]$ *subí*₁ sin zapatos // no, lo que pasó / *yo cuando subí*₂ / *yo me acuerdo*₃ que $[\emptyset]$ *subí*₄ la subida y $[\emptyset]$ *no me acuerdo*₅ quién había y quién no ... *uno* / *uno se achanta*₆ y / todas esas veces que *se toma uno*₇ unas cervecitas por ahí en el centro / *yo me alejé*₈ / pues $[\emptyset]$ *me he alejado*₉ de todo eso. (MEDE H11-2)
- ‘E.: And what did your family tell you about that?
 I.: no, the girls brother / my wife was the one who saw me when [I] *went*₁ upstairs shoeless // no, what happened / when I *went*₂ upstairs / I *remember*₃ that [I] *went*₄ upstairs and [I] *don’t remember*₅ who was there and who wasn’t ... *one* / *one backs down*₆ and / all those times when *one drinks*₇ a few beers around downtown / I *backed*₈ away / well [I] *have backed*₉ away from all that.’

Thus, to probe whether in Colombian Spanish *uno* is assuming first person singular roles, we extracted from the corpus those clauses in which *uno* and *yo*

appear in the same speech turn. Given that our dependent variable consists of the alternation between *uno* and *yo*, we analyzed each of the 1582 tokens collected to determine the interchangeability of both pronouns. During this preliminary analytical stage – upon further data scrutiny and collective consensus – we excluded 219 clauses in which *yo* and *uno* do not constitute a linguistic variable by not being interchangeable. One example of the exclusions is provided in (10), where the speaker focuses on the hearer as the direct object of the enunciative verb *digo* ‘[I] tell’ and the occurrence of *uno* would render the clause ungrammatical.

- (10) E.: ¿en qué le gustaría trabajar?
‘E.: what would you like to work on?’
I: a mí me gustaría trabajar hombre... / en qué te [∅] digo / en / ya *uno* por la edad no / nooo / no *puede hacer* trabajos pesados. (MEDE_H21_5)
‘I: I would like to work, man... / in what would [I] tell you / in / now because of one’s age / nooo / *one can* no longer *do* heavy jobs.’

On the other hand, *yo* could not replace *uno* in clauses where the speakers’ intention is to provide a generic characterization to a statement that is valid to all other persons and not exclusively to themselves, as illustrated in example (11):

- (11) E.: ¿no le gustó el estudio / y ¿por qué?
‘E.: didn’t you like to study / and why?’
I.: ¡ah! / porque *yo no sé* / no / no / *uno* muchas veces así cuando *está* pequeño por los / por los / por los amigos también que / muchas veces *uno se deja llevar*. (MEDE H31-2)
‘I.: ah! / because *I don’t know* / no / no / many times when *one is* young because of / because of / because of one’s friends too that / many times *one gets carried away*.’

Thus, we analyzed the remaining 1363 tokens from sociolinguistic and pragmatic perspectives seeking to probe whether the occurrence of either *uno* or *yo* reflects the same discourse individualization process and expresses the speaker’s positioning and experience in the same way. We coded our 1363 tokens in terms of the predictors discussed above on Excel spreadsheets saved as comma-separated-value files (.csv). We subsequently conducted a series of quantitative analyses with Rbrul and Language Variation Suite (Scrivner & Díaz-Campos 2016) as our statistical tools. The first analytical step in the quantitative exploration of our data sample was a distributional analysis (Tagliamonte 2006: 193, 2012: 121) presented in the next section. This was followed by a series of crosstabulations

intended to detect factor interactions. Then, we analyzed our dataset through a series of mixed-effects multivariate logistic regressions intended to probe hypotheses pertaining to each one of the predictor variables explored. Our multivariate analyses tested speaker as a random-effects predictor with the purpose of probing whether any outliers may skew the results. We also tested attenuation procedure and genericity inducers as a random-effects predictor. This was done because this predictor, by having 18 factors, would have skewed the multivariate regression results as well as the data distribution – with half of our tokens having no attenuating elements.

4 Results

Our presentation of the results starts with the distribution of the variation between *uno* and *yo*. Subsequently, we address the predictors that significantly condition the linguistic variable explored.

4.1 Distribution of variants

The distribution of *uno* and *yo* in interchangeable contexts in Medellín, presented in Figure 1 and Table 2, shows that *uno* (52%) occurs slightly more frequently than *yo* (48%).

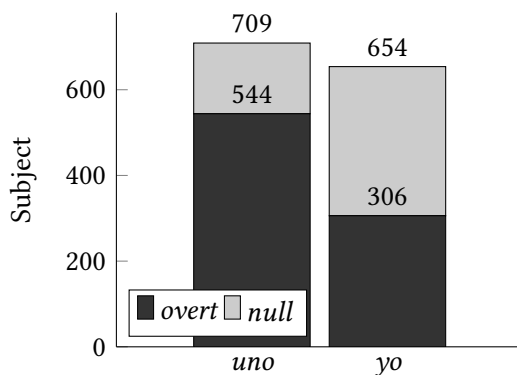


Figure 1: Distribution of *uno* and *yo* in Medellín

Despite *uno* and *yo* being similarly distributed in contexts where they are interchangeable, overt subjects are more frequent with *uno* (77%) than with *yo* (47%). That is, they register significantly different overt/null pronominal subject ratios ($\chi^2 = 128.6, p < 0.001$). The much more frequent occurrence of overt subjects

Table 2: Distribution of *uno* and *yo* in Medellín

Subject type	Variant		Total
	<i>uno</i>	<i>yo</i>	
Null subjects	165 (23%)	348 (53%)	513 (38%)
Overt subjects	544 (77%)	306 (47%)	850 (62%)
Total	709 (52%)	654 (48%)	1363 (100%)

with *uno* may stem from the fact that when *uno* occurs in speech, its first occurrence contains an overt subject whereas that is not the case with *yo*. These results, as well as those from Barranquilla, Colombia (Hurtado & Ortega-Santos 2019: 51), are congruent with the first mention of *uno* being obligatorily overt and its subsequent implications for SPE research.

4.2 Conditioning effects on the *uno/yo* alternation

The quantitative model likely to best explain the *uno/yo* alternation in our data sample is illustrated in the random forest consisting of Figure 2, which is a graphic representation of Table 3.⁴

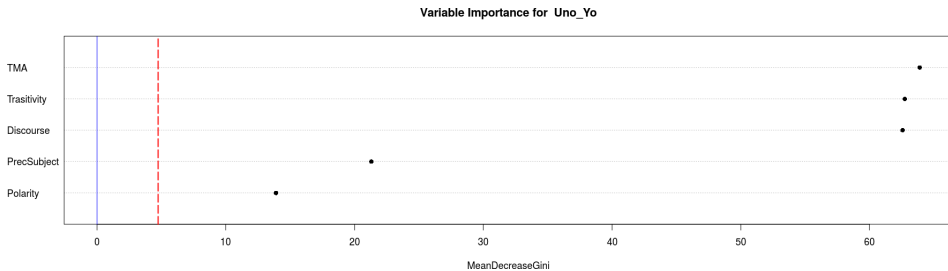


Figure 2: Random forest illustrating the analytical model on the variable *uno/yo* alternation in Medellín

⁴This random forest was established using Language Variation Suite (Scriver & Díaz-Campos 2016). A random forest is a technique embodied in R that helps determine from a set of predictors those most likely to significantly condition a dependent variable (Tagliamonte 2012: 152), in our case the *uno/yo* alternation. A random forest contributes to enhance the explanatory power of a multivariate regression analysis by providing a visual representation of the conditioning on a given linguistic variable that illustrates the relative importance of each predictor.

Table 3: Analytical model on the variable alternation between *uno* and *yo* in Medellín

Predictor	<i>p</i>	Range
Tense, mood & aspect (TMA)	< 0.001	77
Transitivity	< 0.001	51
Discourse genre	< 0.001	34
Type of preceding subject	0.003	13
Polarity	0.033	10
Log likelihood = -739.9	AIC = 1517.8	

The quantitative multivariate model of our data sample reveals that the *uno/yo* alternation is significantly conditioned by all five predictors probed in our analysis: TMA, transitivity, discourse genre, type of preceding subject, and polarity. As previously stated (§3.4), this model also includes speaker as well as attenuation procedure and genericity inducers, both tested as a random-effects factors; that is, they were not tested for statistical significance. TMA emerges as the strongest predictor variable, appearing as the farthest from the broken line in Figure 2 and the predictor with the largest range value (77) in Table 3. The variation under analysis is also strongly conditioned by transitivity and discourse genre. The difference in the relative conditioning strength of the stronger predictors (TMA, transitivity, and discourse genre) and the weaker ones (type of preceding subject and polarity) is illustrated by the gap that appears in the middle of Figure 2, which is also appreciable in Table 3 by the larger range values and smaller but more statistically significant *p*-values for the top three predictors.

Overall, the conditioning effects on the *uno/yo* alternation validate its status as a legitimate linguistic variable. The results corresponding to the conditioning effects for the different predictors are presented in the following paragraphs according to their statistical ranges and *p*-values; that is, in the same order in which they appear in Table 3.

4.3 Verb tense, mood and aspect (TMA)

TMA has the greatest conditioning effect on the *uno/yo* alternation. The results in Table 4 reflect that there are specific contexts of use for either *uno* or *yo*. Infinitives and gerunds – the non-personal verb forms – strongly favor the occurrence of *uno* (0.92). Verbs in the subjunctive mood (0.59) and those in the present indicative – the most frequent tense with 64% of the data – (0.53) also favor the

use of *uno*. At the same time, the imperfect indicative has a neutral effect (0.50) which contrasts with its strongest effect on the alternation between overt and null subjects in this speech community (Orozco & Hurtado 2021a). Conversely, the compound tenses, the conditional and the future – amalgamated as a single factor – (0.25) as well as the preterit indicative (0.15) strongly favor the use of *yo* while disfavoring *uno*.

Table 4: Logistic regression analysis of the effect of TMA on the choice of *uno* in Medellín

Factor	Prob.	Log-odds	% <i>uno</i>	<i>N</i>	% data
Infinitives & gerunds	0.92	2.42	91.2	52/57	4.2
Subjunctive	0.59	0.35	63.9	46/72	5.3
Present Indicative	0.53	0.09	55.4	486/878	64.4
Imperfect Indicative	0.50	−0.02	46.3	76/164	12.0
Other ^a	0.25	−1.12	31.6	36/114	8.4
Preterit Indicative	0.15	−1.72	16.7	13/78	5.7
Range = 77		$p < 0.001$			

^aCompound tenses, conditional & future

These tendencies differ from those registered for SPE in this community (Orozco & Hurtado 2021a). Inter alia, in this analysis we include non-personal verb forms, which are outside the envelope of variation in analyses of the alternation between null and overt pronominal subjects.

The conditional inference tree⁵ (Figure 3) isolates TMA and transitivity – the two strongest conditioners of the *uno/yo* alternation – from all other predictors. It corroborates that TMA has the greatest conditioning effect on our linguistic variable and transitivity also has a strong effect. Moreover, the conditional inference tree shows how TMA intersects with transitivity. On the left-hand side, node 2 bifurcates into node 4, which contains the TMA factors favoring the occurrence of *uno* (infinitives/gerund and subjunctives), and node 3 which contains the transitivity factors that disfavor *uno* (epistemic clauses). On the right-hand

⁵This conditional inference tree was established using Language Variation Suite (Scrivner & Diaz-Campos 2016). A conditional inference tree, like a random forest, is another technique embodied in R which also contributes to enhance the explanatory power of a multivariate regression analysis by highlighting complex interactions within a dataset. One advantage of a conditional inference tree is that it shows “the subtle interactions in the data using a hierarchical display” (Tagliamonte 2012: 153).

side, node 7 bifurcates into the two TMA factors (preterit indicative and others) that favor *yo* by disfavoring the occurrence of *uno*; that is, nodes 8 and 9, respectively.

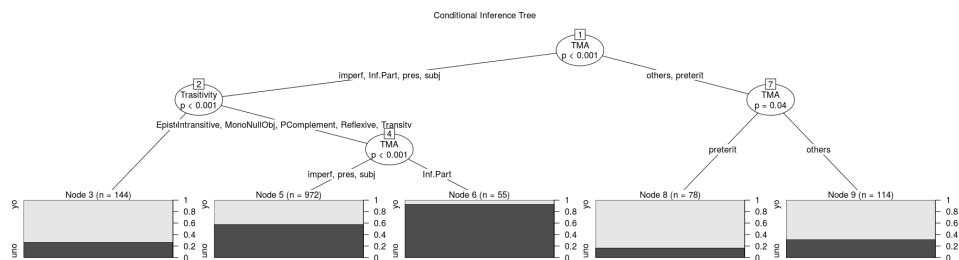


Figure 3: Conditional inference tree for TMA and transitivity

4.4 Transitivity

The results for transitivity (Table 5, page 224) uncover that verbs with different degrees of transitivity favor the occurrence of *uno*, regardless of whether *uno* competes with the object for the focus of attention. Instead, epistemic/evidential clauses with one participant promote the occurrence of *yo*. This is an important finding, as it corroborates what previous studies had already indicated: cognitive verbs behave differently with regard to the use of *yo*. In *pronombrista* studies, cognitive verbs have been found to favor first person singular overt pronominal subjects, possibly because the structure of *yo* + epistemic/evidential verbs + clause is so frequent, that it can be considered a “prefabricated unit” (Travis & Torres Cacoullos 2012: 739).

4.5 Discourse genre

As Table 6 (page 224) shows, opinion statements strongly promote *uno* with a statistical weight of 0.71. At the same time, hypothetical situations exert a neutral effect (0.47) whereas narrations favor the selection of *yo*. These findings clearly indicate that the link between *yo* and the most personal discursive types, on the one hand, and *uno* with argumentative discourse, on the other, is maintained. Because narrations mainly consisted of reporting events that happened in the past or were currently happening at data collection time, those actions are considered realis forms, high in transitivity (Hopper & Thompson 1980). As irrealis forms are low in transitivity, attention is focused on the subject, in this case, expressed by *uno*. This use could be explained as a mitigation strategy, as the speaker expresses opinions.

Table 5: Logistic regression analysis. Effect of transitivity and competition for the focus of attention on the use of *uno* in Medellín. “#P”: Number of participants

Factor	#P	Prob.	Logodds	% <i>uno</i>	<i>N</i>	% data
Verb with prepositional complement	2	0.67	0.71	64.4	29/45	3.3
Intransitive	1	0.59	0.35	56.7	216/381	28.0
Transitive	2/3	0.59	0.38	56.0	255/455	33.4
Reflexive	1	0.56	0.25	55.0	116/211	15.5
Transitive with null object	2	0.50	0.01	44.9	53/118	8.7
Epistemic + complement clause	1	0.16	-1.70	26.1	40/153	11.2
	Range = 51		$p < 0.001$			

Table 6: Logistic regression analysis. Effect of discourse genre on the choice of *uno* in Medellín

Factor	Prob.	Logodds	% <i>uno</i>	<i>N</i>	% data
Opinion	0.71	0.89	68.9	354/514	37.7
Hypothetical situations	0.47	-0.12	49.7	99/199	14.6
Narrative	0.32	-0.77	39.4	256/650	47.7
	Range = 37		$p < 0.001$		

Table 7: Logistic regression analysis. Effect of type of preceding subject on the choice of *uno* in Medellín

Factor	Prob.	Logodds	% <i>uno</i>	<i>N</i>	%data
Overt subjects	0.56	0.22	57.6	285/495	36.3
Others	0.52	0.07	53.9	173/321	23.6
Null subjects	0.43	-0.29	45.9	251/547	40.1
	Range = 13		$p = 0.003$		

4.6 Type of preceding subject

As Table 7 (page 224) shows, preceding overt pronominal subjects favor the selection of *uno* with a statistical weight of 0.56, confirming that an overt pronoun precedes a coreferent *uno*. Preceding null subjects, by disfavoring *uno*, promote the occurrence of *yo* (0.43). Concurrently, all other subjects (lexical subjects and demonstratives) have a neutral effect (0.52).

The conditioning effect of type of preceding subject is similar to what happens across the board with SPE (Cameron & Flores-Ferrán 2004), which has been attested in several Colombian speech communities including Medellín (Orozco & Hurtado 2021a), Cali (Torres Cacoullós & Travis 2019), Barranquilla, and the New York City Colombian enclave (Orozco 2018: 104), respectively. Moreover, these tendencies uncover that the conditioning effect of priming on variation in Spanish extends to the *uno/yo* alternation.

4.7 Polarity

Results for polarity (Table 8) reveal that affirmative statements favor the occurrence of *uno* with a probability weight of 0.55. Nevertheless, negative statements favor the occurrence of *yo* with a weight of 0.45, a tendency contrary to findings by Flores-Ferrán (2009).

Table 8: Logistic regression analysis. Effect of polarity on the choice of *uno* in Medellín

Factor	Prob.	Log-odds	% <i>uno</i>	<i>N</i>	% data
Affirmative	0.55	0.22	54.1	629/1162	85.3
Negative	0.45	-0.22	39.8	80/201	14.7
Range = 10			$p = 0.033$		

4.8 Attenuation procedure and genericity inducers

The effects of the attenuating procedure on the *uno/yo* alternation reveal the pragmatic dynamics involved in the expression of impersonality. Due to the nature of this predictor,⁶ as indicated above (§3.4), we tested its effect as a random

⁶Attenuation procedure and genericity inducers comprise what for a multivariate regression constitutes a large number of factors – initially 18; later reduced to 14 – which would render skewed results if included in a multivariate analysis.

effects factor to measure the pragmatic conditioning on the alternation between *uno* and *yo*. In general (see Table 9), we found that *uno* favors the use of attenuating elements more than *yo*, which suggests that the greater the number of resources available to the speaker, the higher the degree of mitigation (Cestero Mancera 2020: 368). The speaker uses *uno* as follows:

- to reduce their agentivity by placing the subject in a postverbal position (0.75);
- with two or three more attenuating elements (0.65);
- in justifying or apologizing constructions (0.64);
- with adverbial inductors of genericity – especially adverbs of time – (0.57);
- with resources that involve the addressee to draw the interlocutor’s attention (0.54); and
- with diminutives (0.53).

Conversely, *yo* is used mainly without attenuating elements, with strategies expressing doubt, probability and reformulation, with resources that impersonalize, and especially with epistemic markers that indicate uncertainty, lack of knowledge or competence (34 of 39 cases), such as *no sé* ‘I don’t know’ to establish their personal deixis (Caffi 2007).

In previous SPE investigations (cf. Flores-Ferrán 2009), *uno* has been analyzed in terms of its morphosyntactic properties. This investigation extends the analytical scope to its pragmatic properties.

5 Discussion

This variationist investigation has explored the alternation between the Spanish subject pronouns *uno* and *yo* – an underexplored pronominal expression phenomenon – in terms of the extension of *uno* to other morphosyntactic and semantic domains, mainly that of first person singular. Our study has addressed three research questions and a main hypothesis. Interestingly, *uno* and *yo* are interchangeable when they occur within a single speech turn even if they do not indicate impersonality. In fact, in our data sample, they are interchangeable in 1363 or 82% of the 1582 clauses with either pronoun as the subject. The answer to our first research question (*How are uno and yo distributed when they appear*

Table 9: Effects of attenuation procedure and genericity inducers on *uno* in Medellín

Factors	Prob.	Intercept	% <i>uno</i>	N	% data
Postverbal subject	0.75	1.10	80.4	82/102	7.5
2 or 3 attenuating elements	0.65	0.59	62.8	27/43	3.2
Resources that justify	0.64	0.56	66.1	72/109	8.0
Adverbs of time, place and mood	0.57	0.27	66.0	33/50	3.7
Resources that involve the addressee	0.54	0.14	50.0	13/26	1.9
Diminutive suffixes	0.53	0.12	58.3	14/24	1.8
Approximators or diffusers of meaning	0.49	-0.06	57.0	61/107	7.9
Objectivization using modal discourse particles	0.49	-0.05	50.0	9/18	1.3
Resources that limit or restrict ^a	0.47	-0.11	48.1	37/77	5.6
Verbs and particles that express doubt or probability	0.45	-0.19	40.6	13/32	2.3
Resources that correct or reformulate	0.44	-0.26	44.7	21/47	3.4
Absence of attenuating elements	0.42	-0.31	46.7	314/673	49.4
Resources that impersonalize	0.38	-0.50	50.0	8/16	1.2
Verbs and particles that feign uncertainty, incompetence or ignorance	0.20	-1.37	12.8	5/39	2.9

^a(concessivity, expressions with conditional meaning, modal use of verb tenses)

interchangeably within a single speech turn, and what predictor variables condition their alternation?) reveals that both pronouns are similarly distributed although *uno* (52%) is slightly more frequent than *yo* (48%). The internal predictors that most strongly condition the *uno/yo* alternation are TMA, transitivity, and discourse genre. Our analysis uncovered that *uno* is favored by low transitivity contexts such as impersonal verb forms (infinitives and gerunds), imperfective actions (present indicative) and irrealis forms (subjunctive mood and opinion), which promote the focus of attention to remain on the subject. Concurrently, *yo* is favored in high transitivity contexts where the focus is more on the action and the object, as happens in perfective actions (preterit) and realis discursive types such as narration.

The findings for TMA are especially important, not only because this predictor exerted the strongest conditioning influence, but also because they partially support the premise that actions involving impersonal pronouns are framed preferentially in the habitual present and the imperfect (Muñiz Cachón 1998), given that the imperfect indicative has a neutral effect, as example (12) illustrates.

(12) E.: ¿y qué le dijo su familia de eso?

‘E: And what did your family tell you about that?’

I.: no las niñas hermano / fue que la señora es la que me vio cuando [Ø] *subí* sin zapatos // no lo que pasó / *yo cuando subí* / *yo me acuerdo* que [Ø] *subí* la subida y [Ø] *no me acuerdo* quién había y quién no ... *uno* / *uno se achanta* y / todas esas veces que *se toma uno* unas cervecitas por ahí en el centro / *yo me alejé* / pues [Ø] *me he alejado* de todo eso. (MEDE H11-2)

‘I.: no, the girls brother / my wife was the one who saw me when [I] *went*₁ upstairs shoeless // no, what happened / when I *went*₂ upstairs / I *remember*₃ that [I] *went*₄ upstairs and [I] *don’t remember*₅ who was there and who wasn’t ... *one* / *one backs down*₆ and / all those times when *one drinks*₇ a few beers around downtown / I *backed*₈ away/ well [I] *have backed*₉ away from all that.’

The strongest favoring effect of non-finite forms (infinitives and gerunds amalgamated as a single factor [0.92]) on *uno* can be related to the premise that, as these forms do not provide morphological information about tense, aspect and person, they are often used as infinitives of generic interpretation (Real Academia Española 2009: 2002). Our results also concur with those of Flores-Ferrán’s (2009), given that *uno* was also used more frequently in preposition +

subject + infinitive constructions (39%). There were also cases in which the infinitive appears after subordinate clauses (14.6%), with the occurrence of *uno* + infinitive subordinated to a subjunctive clause, as example (13) illustrates. This also occurs frequently in our data in the coreference between subjects with infinitives and the subject of the subordinating clause whose verb occurs in the subjunctive (14):

- (13) ¿Qué [Ø] *pienso*? pues / ojalá que cambie / ojalá que cambie / porque *uno ver uno* por donde pasa y // [Ø] *ve gente muerta* y // eso es muy duro *uno ver gente* ahí tirada. (MEDE_ M11_4)

‘What *do I think*? well [I hope] it changes/ hopefully [it] will change because *one to see* where *one passes* and [*one*] *sees* dead people and // that is very hard *one to see* people lying there.’

- (14) E.: ¿qué es lo que más le gusta? / algún espacio / ¿o le gusta porque se siente cómodo? / ¿qué es lo que más le gusta de su casa?

‘E.: ‘What do you like best? / some space / or do you like it because you feel comfortable? / what do you like most about your house?’

I.: Pues que *uno se sienta* cómodo / *estar bien* esto y esto. (MEDE_H11_2)

‘I.: ‘Well that *one feels* comfortable / *to be* well this and this.’

In both cases, the infinitive obtains temporal information from the predicates to which it is subordinated (Real Academia Española 2009: 1976).

However, when we analyzed the role of the number of participants, we found no effect of the competition for the focus of attention on the selection of *uno*. Instead, *yo* was considerably favored by epistemic/evidential verbs that introduce complement clauses, in which the speaker expresses a stance toward the content of that clause. This is an important finding because Posio (2011) and Hurtado & Ortega-Santos (2019) have found a correlation between low transitivity verbs with one participant and the overt expression of *yo* and *uno*, respectively. Our findings regarding the *uno/yo* alternation corroborate that besides the influence of low transitivity verbs and the focus of attention on the subject, speakers also favor *yo* to express a stance with epistemic verbs, as suggested by Posio (2011). The use of *yo* with low transitivity verbs is also evident in the favoring effect of the attenuation procedure of expressing uncertainty, incompetence or ignorance. *No sé* ‘I don’t know’ was the recurrent form in this category, a negated verb which indicates the degree of commitment of the speaker toward what is said. It is also interesting that, in Medellín, the use of *yo* happens not only with this downgrading strategy that involves the speaker, but also with impersonalization mechanisms that indicate distancing from the speaker (Cestero Mancera 2020).

The answer to our second research question (*In what discursive contexts are uno and yo used interchangeably?*) reveals that, in certain interchangeable contexts, *uno* and *yo* share the same tendencies: with present indicative and hypothetical situations; with attenuation procedures that limit or restrict (concessivity, expressions with conditional meaning, modal use of verb tenses), modal discourse particles for objectivization, and approximators. These findings contribute to validate the premise that *uno* has extended semantically and pragmatically beyond the third person singular – as it is morphosyntactically inflected – to contexts now shared with the first-person singular pronoun.

The tendencies for discourse genre contribute to answer our third research question (*What functions are most commonly assumed by each pronoun, and how do uno and yo construct/create the discursive subject?*), as they suggest a clear differentiation of discursive functions. *Uno* assumes the function of expressing opinions about the city and its people, whereas *yo* assumes narrative functions (personal experiences). Concomitantly, both pronouns equally facilitate the formulation of hypothetical discourse. Despite the tendencies for the other predictors that condition the *uno/yo* alternation not reflecting a clear differentiation of functions, *uno* appears to facilitate the expression of the subjunctive and non-personal verb forms, overt pronominal subjects, and postverbal subjects, respectively. On the other hand, *yo* appears to assume the expression of the preterit, epistemicity, null subjects, as well as that of verbs and particles that feign uncertainty, incompetence, or ignorance.

With regard to TMA, the strongest favoring effect of impersonal forms on *uno* (statistical weight 0.92) highlights the effect of a factor that is not measured in classic *pronombrista* studies because non-finite, impersonal forms are outside the SPE envelope of variation given that such forms do not facilitate the identification of the grammatical person of a null subject (Otheguy & Zentella 2012, Orozco 2018). Concurrently, a priming effect is evident, as the tendencies for type of preceding subject uncover the favorable effect of a preceding overt subject on *uno* whereas *yo* is promoted by preceding null subjects. This priming effect, being similar to the priming effect on SPE, constitutes a structural commonality between the *uno/yo* alternation and the alternation between null and overt pronominal subjects.

6 Conclusion

The present variationist study corroborates the strong effect of *uno* promoting the occurrence of overt pronominal subjects previously reported in Medellín

(Orozco & Hurtado 2021b: 717). Here we have shown how *uno* has entered the morphosyntactic and semantic domain of the first person singular. Our results uncover that there is not a complete reduction of the impersonal reference in all cases where *uno* and *yo* alternate as sentential subjects within a single speech turn. Among other things, our analysis has demonstrated that the *uno/yo* alternation constitutes a linguistic variable in its own right. Along with the goals of this volume, this chapter contributes to augment our collective knowledge of pronominal expression and related linguistic phenomena. Our investigation also contributes to show that the analysis of pronominal expression still has much to contribute to the study of language variation and change.

References

- Aijón Oliva, Miguel Á. & María José Serrano. 2013. El sujeto posverbal: Función pragmática y cognición en las cláusulas declarativas. *Neuphilologische Mitteilungen* 114(3). 309–331.
- Barrajón, Elisa. 2005. Un caso de impersonalidad semántica: El uso de los llamados “singulares arbitrarios” en corpora orales. *Estudios de lingüística* 19. 47–64.
- Bassa Vanrell, Maria del Mar. 2013. Personal *uno* in Puerto Rican and Dominican Spanish. *Studies in the Linguistic Sciences: Illinois Working Papers* 2013. 1–20.
- Biber, Douglas & Edward Finegan. 1988. Adverbial stance types in English. *Discourse Processes* 11. 1–34.
- Botero, Fernando. 1996. *Medellín 1890–1950: Historia urbana y juego de intereses*. Medellín: Universidad de Antioquia.
- Briz Gómez, Antonio. 1995. La atenuación en la conversación coloquial: Una categoría pragmática. In Luis Cortés (ed.), *El español coloquial: Actas del I simposio sobre análisis del discurso oral*, 103–122. Almería: Universidad de Almería, Servicio de Publicaciones.
- Caffi, Claudia. 2007. *Mitigation*. Oxford: Elsevier.
- Cameron, Richard. 1994. Switch reference, verb class, and priming in a variable syntax. In Katharine Beals et al. (eds.), *CLS 30: Papers from the 30th regional meeting of the Chicago Linguistic Society*, vol. 2: The parasession on variation and linguistic theory, 27–45. Chicago Linguistic Society.
- Cameron, Richard & Nydia Flores-Ferrán. 2004. Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context* 1(1). 41–65. DOI: 10.1075/sic.1.1.05cam.

- Carvalho, Ana M., Rafael Orozco & Naomi Shin (eds.). 2015. *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Washington, DC: Georgetown University Press.
- Casielles Suárez, Eugenia. 1996. ¿Es la interpretación arbitraria realmente arbitraria? *Revista Española de Lingüística* 26(2). 359–377.
- Castañeda, Luz Estella. 2005. El parlache: Resultados de una investigación lexicográfica. *Forma y Función* 18. 71–101.
- Cestero, Ana María & Lidia Rodríguez Alfano. 2021. *Guía PRESEEA de estudio de la atenuación*. Alcalá de Henares, Spain: Universidad de Alcalá.
- Cestero Mancera, Ana María. 2020. Uses and resources of mitigation in contrast. *Spanish in Context* 17(2). 362–383.
- Company Company, Concepción & Julia Pozas Loyo. 2009. Los indefinidos compuestos y los pronombres genéricos-impersonales *omme* y *uno*. In Concepción Company Company (ed.), *Sintaxis histórica de la lengua española, Segunda parte: La frase nominal*, 1073–1222. Ciudad de México: Fondo de Cultura Económica y Universidad Autónoma de México.
- De Cock, Barbara. 2014. *Profiling discourse participants: Forms and functions in Spanish conversation and debates* (Pragmatics & Beyond New Series 246). Amsterdam: John Benjamins.
- Departamento Administrativo Nacional de Estadística. 2019. *Resultados del censo nacional de población y vivienda de Colombia*. https://sitios.dane.gov.co/cnpv/#!/vihope_clase.
- Dieck, Marianne. 2016. ¿Y qué más hace uno, pues? La expresión de la impersonalidad en el español de Medellín. *Revista Lingüística y Literatura* 69. 145–175. DOI: 10.17533/udea.lyl.n69a06.
- Fernández, Susana. 2008. Generalizaciones y evidencialidad en español. *Revue Romane* 43(2). 217–234.
- Fernández Acosta, Diana. 2020. El voseo en Medellín, Colombia: Un rasgo dialectal distintivo de la identidad paisa. *Dialectología* 24. 91–109.
- Fernández Ramírez, Salvador. 1986. Las personas gramaticales. Comprensión y transposición. Las personas generales. In *Gramática española*, vol. 4: El verbo y la oración, 21–57. Madrid: Arco Libros.
- Fernández Soriano, Olga. 1999. El pronombre personal. formas y distribuciones. pronombres átonos y tónicos. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, 1209–1273. Madrid: Espasa.
- Flores-Ferrán, Nydia. 2002. *A sociolinguistic perspective on the use of subject personal pronouns in Spanish narratives of Puerto Ricans in New York City*. Munich: Lincom-Europa.

- Flores-Ferrán, Nydia. 2009. Are you referring to me? The variable use of *uno* and *yo* in the oral discourse. *Journal of Pragmatics* 41. 1810–1824.
- Gelabert-Desnoyer, Jaime. 2008. Not so impersonal: Intentionality in the use of pronoun *uno* in contemporary Spanish political discourse. *Pragmatics* 18(3). 407–424.
- Gómez Torrego, Leonardo. 1992. *La impersonalidad gramatical: Descripción y norma*. Madrid: Arco Libros.
- González Vergara, Carlos & Evelin Hugo Rojas. 2012. “Cuando te lo piden uno no siempre sabe qué decir.” Uno y tú como estrategias evidenciales en el español de Chile. In *Actas del IV congreso internacional de letras*, 647–654. Buenos Aires: Universidad de Buenos Aires.
- González-Rátiva, María Claudia. 2008. *Corpus sociolingüístico de Medellín*. Medellín: Facultad de Comunicaciones, Universidad de Antioquia, Colombia. <http://comunicaciones.udea.edu.co/corpuslinguistico/>.
- Guantiva Acosta, Ricardo. 2000. Distribución del uso de las formas de tratamiento pronominal tónica *yo*, *mí*, *conmigo*, e indefnida *uno* en el español hablado en Bogotá. *Forma y Función* 13. 253–262.
- Guirado, Kristel. 2011. La alternancia tú-uno impersonal en el habla de Caracas. *Lingüística* 26. 26–54.
- Haverkate, Henk. 1985. La desfocalización referencial en el español moderno. *Hispanic Linguistics* 2(1). 1–21.
- Haverkate, Henk. 1987. La cortesía como estrategia conversacional. *Diálogos Hispánicos de Ámsterdam* 6. 27–63.
- Hernanz, María Luisa. 1990. En torno a los sujetos arbitrarios: La 2ª persona del singular. In Violeta Demonte & Beatriz Garza Cuarón (eds.), *Estudios de lingüística de España y México*, 151–178. México: Universidad Nacional Autónoma de México & El Colegio de México.
- Hollænder Jensen, Mikkel. 2002. La referencia en algunas expresiones impersonales. *Diferentes lecturas de uno y la segunda persona del singular*. *Romansk Forum* 16(2). 127–138.
- Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299.
- Hurtado, Luz Marcela. 2015. El predominio de *uno* en Bogotá, Colombia: ¿Estrategia para desfocalizar o para subjetivizar? *Anuario de Letras* 3(2). 105–151. DOI: 10.19130/iifl.adel.3.2.2015.1333.
- Hurtado, Luz Marcela. 2016. Dinámica social en la expresión de impersonalidad en Bogotá: Un cambio lingüístico evidente en tres periodos. *Lingüística y Literatura* 69. 177–191.

- Hurtado, Luz Marcela & Carolina Gutiérrez-Rivas. 2016. La versatilidad del pronombre *uno* para expresar posicionamiento frente a lo enunciado en el español de Barranquilla (Colombia). *Forma y Función* 29(1). 37–60.
- Hurtado, Luz Marcela & Rafael Orozco. 2022. La variable expresión de sujetos pronominales al estilo paisa. In Nestor F. Ruiz Vásquez (ed.), *Perspectivas actuales de la investigación en lingüística: Entre tradición y modernidad*, 207–233. Bogotá, Colombia: Instituto Caro y Cuervo.
- Hurtado, Luz Marcela & Iván Ortega-Santos. 2019. On the use of *uno*: The role of transitivity in the Spanish of Barranquilla. *Studies in Hispanic and Lusophone Linguistics* 12(1). 35–64.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William. 2010. *Principles of linguistic change*, vol. 3: Cognitive and cultural factors. Malden, MA: Wiley-Blackwell.
- Monge, Félix. 2002. Las frases pronominales de sentido impersonal en español. In Cristina Sánchez López (ed.), *Las construcciones con SE*, 341–391. Madrid: Visor.
- Montes Giraldo, José Joaquín. 1967. Sobre el voseo en Colombia. *Thesaurus* 22. 21–44.
- Montes Giraldo, José Joaquín. 1982. El español de Colombia. *Thesaurus* 37. 23–92.
- Morales, Amparo. 1995. The loss of the Spanish impersonal particle *se* among bilinguals: A descriptive profile. In Carmen Silva-Corvalán (ed.), *Spanish in four continents*, 148–162. Washington, DC: Georgetown University Press.
- Muñiz Cachón, Carmen. 1998. *Impersonalidad y despersonalización: Estudio contrastivo*. Oviedo: Universidad de Oviedo.
- Olave-Arias, Giohanny, Natalia Gómez-Arcila & Mireya Cisneros-Estupiñán. 2021. Pronominalización personal en el discurso de docentes víctimas del conflicto armado colombiano. *Educación y Humanismo* 23(40). 1–18.
- Orozco, Rafael. 2018. *Spanish in Colombia and New York city: Language contact meets dialectal convergence*. Amsterdam/Philadelphia: John Benjamins.
- Orozco, Rafael & Luz Marcela Hurtado. 2021a. A variationist study of subject pronoun expression in Medellín, Colombia. *Languages* 6(1). Article 5. DOI: 10.3390/languages6010005.
- Orozco, Rafael & Luz Marcela Hurtado. 2021b. Variable subject pronoun expression revisited: This is what the *paisas* do. *Proceedings of the Linguistic Society of America* 6(1). 713–727. DOI: 10.3765/plsa.v6i1.5006.
- Otheguy, Ricardo & Ana Celia Zentella. 2012. *Spanish in New York: Language contact, dialect leveling, and structural continuity*. Oxford: Oxford University Press.

- Posio, Pekka. 2011. Spanish subject pronoun usage and verb semantics revisited: First and second person singular subject pronouns and focusing of attention in spoken Peninsular Spanish. *Journal of Pragmatics* 43(3). 777–798. DOI: 10.1016/j.pragma.2010.10.012.
- Posio, Pekka. 2013. The expression of first-person-singular subjects in spoken Peninsular Spanish and European Portuguese: Semantic roles and formulaic sequences. *Folia Linguistica* 47(1). 253–291. DOI: 10.1515/flin.2013.010.
- Ramírez, Dora. 2007. Language attrition and language maintenance: The case of Colombian immigrants in New York State: State University of New York, Albany. (Doctoral dissertation).
- Real Academia Española, Asociación de Academias de la Lengua Española. 2009. *Nueva gramática de la lengua española*. Madrid: Espasa Calpe España.
- Repede, Doina & Marta Leon-Castro. 2019. Procedimientos de impersonalización en la ciudad de Sevilla: El caso del pronombre *uno*. *Cuadernos de Investigación Filológica* 46. 77–102.
- Ricos Vidal, Amparo. 2002. Construcciones impersonales en el español medieval y clásico: estructuras con *omne*, *se* y *uno*. In María Teresa Echenique Elizondo & Juan Pedro Sánchez Méndez (eds.), *Actas del V congreso internacional de historia de la lengua española*, 945–958. Madrid: Gredos.
- Rodríguez Alfano, Lidia. 2004. Deixis personal manifestación discursiva del sujeto y su ideología. In *¿Qué opinas con verbos y pronombres? Análisis del discurso de dos grupos sociales de Monterrey*, 151–214. Monterrey: U Autónoma de Nuevo León.
- Scriver, Olga & Manuel Díaz-Campos. 2016. Language variation suite: A theoretical and methodological contribution for linguistic analysis. *Proceedings of the Linguistic Society of America* 1. Article 29. DOI: 10.3765/plsa.v1i0.3734.
- Serrano, María José. 2012. El sujeto pronominal *usted/ustedes* y su posición: Variación y creación de estilos comunicativos. *Spanish in Context* 9(1). 109–131.
- Siewierska, Anna. 2008. Impersonalization from a subject-centred vs. agent-centred perspective. *Transactions of the Philological Society* 106(2). 115–137. DOI: 10.1111/j.1467-968X.2008.00211.x.
- Tagliamonte, Sali A. 2006. *Analysing sociolinguistic variation*. New York: Cambridge University Press. DOI: 10.1017/CBO9780511801624.
- Tagliamonte, Sali A. 2012. *Variationist sociolinguistics: Change, observation, interpretation*. Malden, MA/Oxford: Wiley-Blackwell.
- Tamminga, Meredith, Laurel MacKenzie & David Embick. 2016. The dynamics of variation in individuals. *Linguistic Variation* 16. 300–336. DOI: 10.1075/lv.16.2.06tam.

- Thompson, Sandra A. & Paul J. Hopper. 2001. Transitivity, clause structure, and argument structure: Evidence from conversation. In Joan L. Bybee & Paul J. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 27–60. Amsterdam & Philadelphia: John Benjamins.
- Torres Cacoullos, Rena & Catherine E. Travis. 2019. Variationist typology: Shared probabilistic constraints across (non-)null subject languages. *Linguistics* 57(3). 653–692.
- Travis, Catherine E. 2005. The yo-yo effect: Priming in subject expression in Colombian Spanish. In Randall Gess & Edward J. Rubin (eds.), *Theoretical and experimental approaches to Romance linguistics: Selected papers from the 34th linguistic symposium on Romance languages, 2004*, 329–349. Amsterdam & Philadelphia: John Benjamins.
- Travis, Catherine E. 2007. Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change* 19(2). 101–135. DOI: 10.1017/S0954394507070081.
- Travis, Catherine E. & Rena Torres Cacoullos. 2012. What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics* 23(4). 711–748. DOI: 10.1515/cog-2012-0022.

Chapter 9

Discourse participants in impersonal constructions: The case of first and second person object pronouns with Spanish non-anaphoric third person plural subjects

Émeline Pierre 

Université catholique de Louvain

Barbara De Cock 

Université catholique de Louvain

This paper offers a quantitative and qualitative analysis of object pronouns referring to discourse participants used with Spanish non-anaphoric third person plural subjects, e.g., *me han criticado por largas respuestas* ('they have criticized me for long answers'). Non-anaphoric third person plurals being agent-defocusing mechanisms which are sometimes considered to be impersonal, the discourse participant objects have a higher referentiality than the subject. We will show how this impacts the construction and conceptualization as a whole. By combining data from oral and written mode and informal and formal register, we examine the syntactic and semantic roles of discourse participants as well as the verb types they are used with. Our fine-grained study highlights that the use of discourse participant object pronouns and their roles are closely associated with the presence of relational verbs and communication verbs. Moreover, we show how the higher referentiality of the discourse participant object with respect to the subject is reflected through its being the main anchoring point for topic continuity. Finally, results indicate that the combination of a discourse participant object pronoun with a non-anaphoric third person plural is highly associated with the informal register and, whereas within oral conversations discourse participants tend to be oriented towards the speaker, in written interactions from a digital forum they rather involve the interlocutor.



1 Introduction

In this chapter, we want to focus on non-anaphoric third person plural subjects. Non-anaphoric third person plurals consist of a verb agreeing with a third person null subject which, contrary to their anaphoric counterparts, lack an antecedent in the linguistic context. This is why we resort to the term “non-anaphoric third person plurals”, following Cabredo Hofherr (2006). More specifically, we will look into the impact of the presence of a discourse participant object, that is, an object in first or second person (be it direct or indirect), on the use of this form. We will do this through a comparison with non-discourse participant object pronouns (third person forms), focusing particularly on the ones referring to an animate entity, since those are most likely to have similar functions to the (by definition animate) discourse participants.

Example (1) combines a non-anaphoric third person plural (*han criticado* ‘have criticized’) with a first person pronoun object pronoun (*me* ‘me’) and illustrates the phenomenon we are interested in.

- (1) Los sistemas eléctricos son una de las características que los diferencian y también otros importantes, pero alargaría mucho esto y ya *me han criticado* por largas respuestas, pero si quieres aprender tienes que leer.
Saludos (YAHOO)

‘Electrical systems are one of the features that differentiate them and also other important ones, but I would lengthen this a lot and *they* already *have criticized me* for long answers, but if you want to learn you have to read. Regards’

Non-anaphoric third person plural forms have mainly been studied as impersonal constructions and have been considered to have a “non-referential human subject which excludes the speaker and the addressee” (Siewierska & Papastathi 2011: 577). Given the above, non-anaphoric uses of the third person plural are considered “to be impersonal under the functional, agent defocusing view of impersonality which associates defocusing of an agent with loss of subject status and/or lack of full referentiality” (Siewierska 2010: 74). As such, this use can be considered functionally similar to passive structures. Such agent-defocusing structures, among which we can find constructions such as the periphrastic passive (auxiliary with a past participle), the *se*-construction or the numeral-based *uno*, defocus or demote the agent of a verbal clause through under-elaboration, non-elaboration or both processes (see Siewierska 2008). As Spanish is a pro-drop language, the presence of a subject pronoun is not obligatory, the subject

being retrievable from the verbal morphology. Traditionally, it has been considered that the presence of a subject pronoun even cancels the non-anaphoric reading of the Spanish third person plural form (Fernández Soriano & Táboas Baylín 1999: 1739). However, Shin & Otheguy (2005) have shown that in some varieties of Spanish, a non-anaphoric reading is possible with overt *ellos* ‘they’. Such cases were not found in the data studied for the present analysis.

Focusing on colloquial English, Weiner & Labov (1983: 35–36) show that the non-anaphoric third person plural can be considered a major alternative for the agentless passive. De Cock (2014: 194) has shown previously that it is also much more frequent in Spanish informal spoken interaction than in more formal spoken interactions, such as TV debates or parliamentary debates. In the more formal interactions, other agent-defocusing structures such as the *se*-passive and the periphrastic passive are more frequent, suggesting that in some cases the non-anaphoric third person plural can function as an alternative for Spanish agentless passive constructions. While there is a wider variety of agent-defocusing constructions in Spanish, we focus particularly on the alternation between a periphrastic passive and a non-anaphoric third person plural form, which has been described more amply for English. We will occasionally refer to the alternation between the non-anaphoric third person plural form and other agent-defocusing constructions as well, but refer the reader to Pierre (2021) for a more elaborate analysis of the interaction between all agent-defocusing constructions.

Different uses of the non-anaphoric third person plural form have been identified among others by Cabredo Hofherr (2006) and brought together in a semantic map by Siewierska & Papastathi (2011), based on criteria such as the delimitation through a locative, predicative, time or inference. Posio (2015) looks into the applicability of this proposal for Spanish. Through a study of a translation corpus of some Harry Potter chapters, Siewierska & Papastathi (2011) point out the importance of analyzing the role of speech act verbs with non-anaphoric third person plural forms (2011: 606). In order to investigate the importance of speech act verbs (and other verbs), we will analyze with which verb types the non-anaphoric third person plurals occur in various genres. We then expect these speech act verbs to be frequently used with non-anaphoric third person plurals. We will furthermore look into the impact of the presence of an object pronoun, particularly discourse participant object pronouns, in order to offer a more detailed account of the impact of the verb type and the presence of a discourse participant on the use of non-anaphoric third person plural subjects.

Indeed, we wish to focus on the presence of discourse participant object pronouns with these constructions since they are often the most referential element of the utterance, given that the human subject is non-referential (Siewierska &

Papastathi 2011: 577). We thus expect this highly referential element to have an impact on the utterance as a whole and to fulfill a more crucial role for the discursive progression than the non-specific third person plural subjects.

The article is structured as follows. Data and methodology are presented in §2. Using comparable corpora from different situations, we start by offering a quantitative overview of the data (§3). Then, in §4, we analyze with which semantic verb types and with which semantic roles these (non)-discourse participant pronouns occur (§4.1). The analysis then focuses on the pragmatic-discursive effects of the construction choice on the presentation of the first and second person (§4.2). Finally, we examine the presence of the phenomenon in function of genre and register variation (§4.3). §5 synthesizes our main observations and presents the general conclusions drawn from this corpus-based study.

2 Data and methodology

In order to obtain a fine-grained description of the presence of direct object pronouns with non-anaphoric third person plurals, the present investigation looks into the presence of this phenomenon in both formal and informal registers and oral and written modes. Formal data consist of excerpts from the Wikipedia corpus (Reese et al. 2010) for the written part and from European parliamentary debates (PROCEP) for the oral part. Only two occurrences of a non-anaphoric third person plural with a discourse participant were found in formal data. We will thus not provide further information on these corpora (for more details see Pierre 2021: 74–78) but the low frequency of the phenomenon in formal data is in itself a relevant finding. The informal written part consists of extracts from the Spanish version of the Yahoo-based Contrastive Corpus Questions & Answers (see De Smet 2009) compiled from the digital forum Yahoo Questions and Answers. A total of 46,603 words from daily life topics such as means of transportation and food habits has been analyzed. The informal oral part of the data comes from the corpus Español Lengua Oral (ESLORAv.2) (see Barcala et al. 2018 or Vázquez Rozas et al. 2020 for a detailed description of the corpus) and consists of spontaneous conversations between two or more interlocutors, usually friends or relatives, recorded between 2007 and 2015. The metadata available with the corpus allowed us to select only the participants whose main language is Castilian Spanish and to ensure that they use it on a daily basis. Participants whose native and main language was Galician or another official language of Spain were discarded. Similarly to the data selected from the digital forum, conversations mainly address topics from day-to-day life. The oral corpus totals 76,749 words.

This greater size of the oral corpus compared to the written one is explained by the fact that the data used in the present work are part of a larger investigation on structures that are functionally similar to the passive, which are, overall, less frequent in oral data (Pierre 2021: 80) and required a larger corpus to collect sufficient material. All the examples used in this Chapter are presented in their original form and, thus, may contain non-standard language.

The degree of formality of language production can be measured following various methods such as the heuristic one suggested by Heylighen & Dewaele (1999) or the continuum offered by Briz (2010) which considers features such as the relation between the interlocutors, the degree of shared knowledge, the setting and the topic of the interaction and the planning and the tone of the discourse (Briz 2010). We rely on this latter methodology to classify the data used in the present study and thus consider the interactions collected from the Yahoo and ESLORA corpora as informal language, whereas the fragments extracted from Wikipedia articles and parliamentary debates exhibit a higher degree of formality and are classified as samples of formal language.

All occurrences of non-anaphoric third person plurals were extracted by a combination of automatized searches. This process was carried out in the corpus processing system Unitex,¹ a text searching program that allows the automatic retrieval of linguistic phenomena. This process was followed by a manual revision to ensure the agent-defocusing character of each example. A manual annotation of the data was then performed on all the non-anaphoric third person plurals. Two parameters have been annotated: (i) the verb types, following the taxonomy developed in the *Alternancias de Diátesis y Esquemas Sintáctico-Semánticos del Español* (ADESSE) project (García-Miguel & Albertuz 2005), designed by the University of Vigo (Spain) and (ii) the syntactic and semantic role of the (non)-discourse participants.

3 Quantitative overview

Following the extraction methodology presented above, a total of 498 occurrences of non-anaphoric third person plurals was extracted, which gives a normalized frequency of 23.85 structures per ten thousand words. Out of this set, a manual revision and annotation made it possible to identify non-anaphoric third person plurals used with a discourse participant (first and second person object pronouns) and/or a non-discourse participant object pronoun (third person object pronoun). No cases of *usted/ustedes*, the polite form of address of the

¹Unitex/GramLab retrieved from <https://unitexgramlab.org/fr>.

pronoun *you*, were found in our data. This amounts to 274 occurrences² of (non)-discourse participant object pronouns, which represents more than half of all the occurrences of the non-anaphoric third person plurals. These occurrences constitute the final data set of the present study. The next two examples illustrate cases of non-anaphoric third person plurals with a discourse participant (the second person singular *te* ‘you’ in Example (2)) and with a non-discourse participant, illustrated in (3) with a third person plural pronoun *les*.³

- (2) H2 no se me ocurrió[a tampoco pero es que] si hacía esas cosas mi padre me daba un / so[papo]
Speaker 2 ‘I didn’t thin[k of it either but] if I did those things my father would give me a / s[lap]’
H1 [ni se te ocurría]
Speaker 1 ‘[you wouldn’t even think of it]’
H1 [sí oh] <inint> <Pausa>
Speaker 1 ‘[yes oh] <inint> <Pause>’
H2 y me decía / <cita>[no vuelves] a salir en [tu vida]</cita>
Speaker 2 ‘and he would tell me / <quote>[you’ll never again] go out in [your life] </quote>’
H1 [lo que pasa que]
Speaker 1 ‘[the thing is]’
H1 [ahora no se les] puede pegar que que= es maltrato infantil / y *te pueden denunciar* <Pausa> ¿sabes?
Speaker 1 ‘[now you can’t] hit them that= it’s child abuse / and *they can report you* <Pause> you know?’
H2 pues te voy a decir una [cosa]
Speaker 2 ‘Well, I’m going to tell you [something]’
H1 [fff] <Pausa>
Speaker 1 [fff] <Pause>
H2 unas hostias a tiempo solucionan muchas cosas
Speaker 2 ‘A few punches in time will solve a lot of things.’ (ESLORA)

²Nineteen occurrences of non-anaphoric third person plurals combine two discourse participants, one in direct object position and one an indirect object position. This phenomenon is illustrated and discussed in §3.

³We have slightly modified the transcriptions of the ESLORA corpus in order to distinguish the pauses more clearly from the remainder of the production.

- (3) H2 que una vez se quejaron los peregrinos que= que *les habían cobrado* por= / meterlos en un pabellón porque normalmente <Pausa>

Speaker 1 ‘that pilgrims once complained that= *they had been charged* for= / putting them in a pavilion because they were normally <Pause>’

H1 es gratis

Speaker 1 ‘it’s free’ (ESLORA)

Third person pronouns represent participants who are not directly involved in the interaction but rather participants about whom certain facts or events are reported.

Table 1: Distribution of (non-discourse) participants according to their syntactic role and animate nature (raw frequencies).

Type of participant	Raw number direct object			Raw number indirect object		Total
	Ani-mate	Inani-mate	Indetermi-nate	Animate	Inanimate	
1SG	16	0	0	34	0	50
1PL	1	0	1	4	0	6
2SG	12	0	0	54	0	66
2PL	0	0	0	3	0	3
Sub-total	29	0	1	95	0	125
3SG	14	65	2	23	3	107
3PL	17	13	1	9	2	42
Sub-total	31	78	3	32	5	149
Total						274

As indicated in Table 1, 125 non-anaphoric third person plurals appear with a discourse participant object pronoun and 149 with a third person object pronoun. When used with a discourse participant, results show that it is preferably a singular pronoun since, out of the 125 discourse participants, first and second person singular pronouns total 50 and 66 occurrences, respectively. The overall dominance of singular pronouns can partly be due to first and second person singular pronouns being generally more frequent than their plural counterparts in our data. These two object pronouns are illustrated in (4) and (5), respectively.

- (4) H2 el ciclo este que quiero hacer después de= de lo de Forestales que es de= / también de Forestales de lo de= <Pausa>
Speaker 2 'the cycle that I want to do after= the Forestry course which is from= / also from Forestry from= <Pause>'
H1 sí <Pausa>
Speaker 1 'yes <Pause>'
H2 <#> ¿cómo se llama? / gestor del= // de recursos forestales o algo así <Pausa> al tener la carrera *me convalidan* en primero / dos asignaturas <Pausa>
Speaker 2 '<#> what is it called? / manager of the= // of forestry resources or something like that Pause as I have the degree *they validate* me in the first / two subjects <Pause>' (ESLORA)
- (5) H2 pero no la llamaron de ningún lado y [empez]ó ahí porque necesitaba dinero
Speaker 2 'but they didn't call her from anywhere and she [started] there because she needed money'
H1 [claro] claro <Pausa>
Speaker 1 '[of course] of course <Pause>'
H2 pues es lo que hay que hacer <Pausa> tú has estudiado para <Pausa> pues para trabajar [si]
Speaker 2 'so that's what you have to do <Pause> you have studied for <Pause> so to work [if]'
H1 [<inint>] pues ya *te llamarán*= [<inint>]
Speaker 1 '[<inint>] then *they will call you*= [<inint>]'
H2 [si hay suerte] <Pausa>
Speaker 2 '[if there is luck] <Pause>' (ESLORA)

Though in substantially more limited proportions, it is interesting to note that plural discourse participants are not absent from non-anaphoric third person plurals.

- (6) Me gustaría hacer un par de puntualizaciones en este sentido. Está bien que hablemos de limitaciones de la producción, pero estas limitaciones de la producción tienen que venir asociadas al concepto de soberanía alimentaria, porque lo que no puede ser es que *nos hagan* lo mismo que con la leche: *no nos digan* — a España — que tenemos que producir menos leche para que se invada nuestro mercado de leche extranjera, porque eso no nos hace más sostenibles, nos hace más pobres. (PROCEP)

‘I would like to make a couple of clarifications in this regard. It is right that we talk about production limitations, but these production limitations must be associated with the concept of food sovereignty, because what cannot be is that *they do* the same to us as with milk: *they do not tell us* – Spain – that we have to produce less milk so that our market is invaded by foreign milk, because that does not make us more sustainable, it makes us poorer.’

Example (6), an excerpt from parliamentary debates, shows an instance with the first person plural object pronoun *nos* (‘us’). This ties in with a higher overall presence of first person plural forms in parliamentary debates, where speakers typically are spokespersons for a whole group, typically their political party (Gelabert-Desnoyer 2006a,b, De Cock 2014: 35). In this excerpt from the European Parliament, the speaker specifies that she speaks for her whole national group (by referring to *a España*), thus resolving a possible ambiguity between reference to the political or national identity.

While first and second person pronouns, that is, pronouns encoding participants directly involved in the interaction, are rarely used in the plural in our data, third person plural pronouns, illustrated in (7) with the direct object pronoun *las* ‘them’, tend to be more commonly employed.

- (7) H1 [parece ser que es bas]tante timo el bufé<@@> <Pausa>
 Speaker 1 ‘[it seems to be quite a fraud the buffet<@@> <Pause>’
 H2 ¿sí? <Pausa>
 Speaker 2 ‘Yes? <Pause>’
 H1 creo que son= para entrar diez euros <Pausa> pero tienes que tomar una consumición <Pausa> y las consumiciones dicen que *las clavan*= bien clavadas <Pausa>
 Speaker 1 ‘I think it’s= ten euros to get in <Pause> but you have to have a drink <Pause> and they say that the drinks *they ripped* them off = well ripped off <Pause>’
 H2 ¡bua! pues yo cuando fui a Barcelona Pausa pagué diez euros / con= bebida incluida
 Speaker 2 ‘Wow! well when I went to Barcelona <Pause> I paid ten euros / with= drink included’ (ESLORA)

We also analyzed the syntactic role of the discourse participant, that is, direct or indirect object and identified its nature (animate versus inanimate), as it interacts with the verb type used in the construction.

A look at the syntactic role of the object pronouns reveals that discourse participants are preferably used as indirect objects, that is, as the receiver or the beneficiary of the action (95 occurrences out of 125), and especially the second person singular pronoun (54 occurrences). However, a larger set of data would be necessary to confirm the tendency. A comparison with non-discourse participants shows that, while discourse participants are preferably used as indirect objects, non-discourse participants are rather direct objects, reflecting more general tendencies of indirect objects often having animate referents. The relatively high proportion of inanimate third person singular pronouns (65 occurrences out of 107 third person singular pronouns) largely contributes to this result.

The analysis also highlights 19 cases where both a direct and indirect object pronoun co-occur with the non-anaphoric third person plural. In our data, these cases always combine a discourse with a non-discourse participant, as illustrated in (8), which includes a third person direct object *lo* 'it' and a first person indirect object *me* 'me'.

- (8) H2 que sí que= había visto algún- disfraces muy chulos y todo / lo
[que pasa que] yo no entiendo nada de lo de= ya te digo de
Comadres y todas estas historias porque nunca *me lo*
[*explicaron* tampoco]

Speaker 2 'that yes that= I had seen some very cool costumes and everything / what [happens that] I don't understand anything about= I'm telling you about Comadres and all these stories because *they* never [*explained it to me* either].'

- H1 [claro][pues eso / Comadres] y Compadres es así creo que es un día que se disfrazan todos de ho- mujeres y otro día también pero es más de postureo [¿sabes?]

Speaker 1 '[of course] [so that / Comadres] and Compadres is like that I think it's a day when they all dress up as me-women and another day too but it's more like pretending [you know?]' (ESLORA)

As will be discussed in §4.1, the syntactic role of the object pronouns is closely linked to the semantics of the verb.

Regarding the nature of the object pronouns, that is, animate versus inanimate,⁴ results indicate that animate pronouns dominate, whether the object discourse participant acts as a direct complement or an indirect one. The tendency

⁴Three occurrences were categorized as *indeterminate*. In these cases it was not possible to determine the nature of the pronoun due to a lack of context.

is not observed when the object pronoun does not act as a discourse participant, in other words, when it is a third person pronoun (9).

- (9) H1 bueno yo mañana sin falta lo pago sí porque yo creo que estoy fuera de plazo total= <Pausa>
 Speaker 1 ‘well tomorrow without fail I will pay it yes because I think I am out of schedule total= <Pause>’
 H1 y si *la cortan* por lo menos que no sea de noche <Pausa>
 Speaker 1 ‘and if *they cut it*, at least not at night <Pause>’
 H2 sí / toma to[ma] H1 vete a saber si mañana nos levantamos sin luz ;eh!
 Speaker 2 ‘yes / take ta[ke] H1 you never know if tomorrow we’ll wake up without electricity, eh!’ (ESLORA)

Third person pronouns display a clear trend towards the substitution of direct object nouns referring to inanimate entities (9) or part of the discourse (10). The alternative for (9) in a periphrastic passive, namely [*la electricidad*] *es cortada* ‘the electricity is cut off’ is imaginable but unlikely in colloquial language, underlining again the importance of the non-anaphoric third person plural as an agent-defocusing strategy in this register.

- (10) H1 y si faltas a la práctica / estás suspenso [/ tú eliges o ha]cer huelga o ir a la prácti[ca]
 Speaker 1 ‘and if you miss practice / you are suspended [/ you choose whether to] strike or go to prac[tice]’
 H2 [ho]y lo dije[ron *lo dijeron*] en la manifestación rollo que había profesores
 Speaker 2 ‘[to]day [*they said it said it*]’ in the strike that they were teachers’
 H1 [<dud>hay dos</dud>] <Pausa>
 Speaker 1 ‘[<doubt>there are two</doubt>] <Pause>’
 H2 incluso ponían exámenes a propósito estos tres días para que la gente no faltara
 Speaker 2 ‘they even put exams on purpose these three days so that people wouldn’t be absent’ (ESLORA)

It is interesting to note that when third person direct object pronouns are used to substitute an animate entity, they do not stand for humans but also frequently

refer to an animal (18 occurrences out of 31). In our data, this result does not apply to indirect object pronouns since animals account for 2 occurrences out of 32.

In addition to the variation in number and person, in syntactic role and nature of the object, the analysis has revealed that discourse participants included in non-anaphoric third person plurals fulfil different semantic roles, as will be discussed in more detail in the qualitative analysis.

4 Exploring the data: A qualitative analysis

4.1 Types of processes

The analysis of the *type of process* involved in a non-anaphoric third person plural with and without an object pronoun provides valuable insights into the understanding of the use of the agent-defocusing mechanism and the subsequent semantic roles of discourse participants.

Experience “consists of a flow of events, or ‘goings-on’” (Halliday & Matthiessen 2014: 170). In Halliday’s (1994) terminology, this flow of events is divided into different *process types* which reflect experiences of happening, doing, sensing, saying, being or having (see Halliday & Matthiessen 1999). They unfold through time and have participants being directly involved in this process in some way; and, in addition, there may be circumstances of time, space, cause, manner or one of a few other types (Halliday & Matthiessen 2014: 170). Building on this systemic functional approach and the adapted classification offered in the ADESSE database (*Alternancias de Diátesis y Esquemas Sintáctico-Semánticos del Español*), designed by the University of Vigo (Spain) (García-Miguel & Albertuz 2005), verbs will be classified according to the type of process they reflect: *mental, relational, material, verbal, existential* and *modulation*⁵ (labelled according to García-Miguel & Albertuz 2005’s work).

Let us first compare the behaviour of non-anaphoric third person plurals with and without object pronouns. Tables 2 and 3 reveal that when used with object pronouns, verbs of *possession* largely predominate (31.9%, 81 occurrences), followed by verbs of *communication* and verbs denoting *spaces*. These findings will be deepened in what follows. A different picture emerges when non-anaphoric

⁵Modulation processes include verbs expressing causation (e.g., *to help* or *to allow*), acceptance (e.g., *to accept* or *to reject*), disposition (e.g., *to dare* or *to try*) and verbs of support. This latter category gathers verbs that combine with nominal clauses to form semantically complex phrases which confer another meaning on them than the one conveyed by the base verb.

Table 2: Verb types in non-anaphoric third person plurals with direct and indirect object pronouns vs. without object pronouns (raw frequencies (%)).

Main category of process	Subtype of process	Non-anaphoric 3PL with object pronouns	Non-anaphoric 3PL without object pronouns	Examples
Mental	Sensation	2 (0.8)	3 (3.5)	<i>querer</i> ('to love, to want')
	Perception	12 (4.7)	0	<i>ver</i> ('to see')
	Cognition	3 (1.2)	5 (5.8)	<i>enseñar</i> ('to teach')
Relational	Attribution	9 (3.5)	2 (2.3)	<i>asignar</i> ('to assign')
	Possession	81 (31.9)	7 (8.1)	<i>cobrar</i> ('to charge'), <i>pagar</i> ('to pay')
Material	Space	31 (12.2)	23 (26.7)	<i>guardar</i> ('to keep'), <i>aislar</i> , ('to isolate')
	Change	23 (9.1)	5 (5.8)	<i>reconstruir</i> ('to rebuild')
	Other facts	16 (6.3)	3 (3.5)	<i>utilizar</i> ('to use')
	Behaviour	5 (2)	1 (1.2)	<i>violar</i> ('to rape')

Table 3: Verb types in non-anaphoric third person plurals with direct and indirect object pronouns vs. without object pronouns (raw frequencies (%)).

Main category of process	Subtype of process	Non-anaphoric 3PL with object pronouns	Non-anaphoric 3PL without object pronouns	Examples
Verbal	Communication	50 (19.7)	30 (34.9)	<i>explicar</i> ('to explain'), <i>decir</i> ('to say')
	Assessment	2 (0.8)	0	<i>criticar</i> ('to criticize')
Modulation	Causation	4 (1.6)	0	<i>ayudar</i> ('to help')
	Acceptation	3 (1.2)	2 (2.3)	<i>aceptar</i> ('to accept')
	Verbs of support	9 (3.5)	4 (4.7)	<i>darse cuenta</i> ('to realize')
Existential	Phase-Time	2 (0.8)	1 (1.2)	<i>desencadenar</i> ('to trigger')
	Life	2 (0.8)	0	<i>embarazar</i> ('to get [sb] pregnant')
Not codifiable due to truncated utterance		1 (0.4)	0	
Total		255 (100)	86 (100)	

third person plural subjects do not co-occur with object pronouns, since it appears that *possession* verbs are hardly used (8.1%) while *communication* verbs (34.9%) and, to a smaller extent, *space* verbs (26.7%) are the only two prevalent categories. These results point towards a different use of the non-anaphoric third person plurals when accompanied by object pronouns.

Interestingly, when non-anaphoric third person plurals are used with object pronouns, the study indicates that the results can be influenced by the frequent use of specific verbs. It is the case for the semantic category *communication* where the verb *decir* ('to say') predominates, followed by the verb *llamar* ('to call'). A similar pattern occurs with *possession* verbs, where the verb *dar* ('to give') largely dominates, followed by *pagar* ('to pay') and *vender* ('to sell'). The tendency for one or a restricted number of verbs to appear frequently is not noticed in the semantic category that includes *space* verbs. These observations lead us to assume that, depending on the semantic category, it is either a specific verb within the category or the semantics of the category itself which plays a decisive role in explaining the behaviour of non-anaphoric third person plurals. A brief review of the structure without object pronouns reveals a comparable phenomenon, the only difference being that *tener* ('to have') is the most frequent verb in the category *possession* verbs. As figures are lower when non-anaphoric third persons do not occur with an object pronoun, this result should be considered with caution.

In what follows, we will examine the potential association of object pronouns with the different types of verbs.

We now focus on discourse participants. It can be seen from Table 4⁶ that two categories largely predominate: *possession* verbs, which pertain to the head category *relational* verbs, and *communication*, which is one of the two subtypes of *verbal* types of processes. The dominance of these types ties in with the dominance of indirect object forms among discourse participant object pronouns shown in Table 1, since these verbs privilege an indirect object. This high frequency of communication verbs further underpins the suggestion of Siewierska & Papastathi (2011) to create a specific category of these verbs in their non-episodic use in the analysis of non-anaphoric third person plural forms.

Verbs of *possession* are the most common verbs used with discourse participants (37.6%, 47 occurrences). This type of verb traditionally receives two semantic labels: belonging, where an entity owns (part of) another entity, and transfer, which implies a change of owner of an entity, where a transfer from an agent (initial owner) to a recipient (final owner) takes place (García-Miguel & Albertuz

⁶The total of Table 4 amounts to 274 as it includes the 19 cases that combine both a direct and an indirect object.

Table 4: Verb types in non-anaphoric third person plurals occurring with object pronouns referring to discourse participants vs. not referring to discourse participants (raw frequencies (%)).

Main category of process	Subtype of process	Discourse participant Raw number (%)	Non-discourse participant Raw number (%)
Mental	Sensation	1 (0.8)	1 (0.7)
	Perception	7 (5.6)	6 (4)
	Cognition	2 (1.6)	1 (0.7)
Relational	Attribution	4 (3.2)	6 (4)
	Possession	47 (37.6)	43 (28.9)
Material	Space	17 (13.6)	18 (12.1)
	Change	2 (1.6)	23 (15.4)
	Other facts	2 (1.6)	15 (10.1)
	Behaviour	2 (1.6)	3 (2)
Verbal	Communication	29 (23.2)	22 (14.8)
	Assessment	2 (1.6)	0
Modulation	Causation	3 (2.4)	1 (0.7)
	Acceptation	2 (1.6)	1 (0.7)
	Verbs of support	4 (3.2)	5 (3.4)
Existential	Phase-Time	0	2 (1.3)
	Life	0	2 (1.3)
Not codifiable		1 (0.8)	0
Total		125 (100)	149 (100)

2005). Example (11) illustrates this category with the verb *pagan* ('they pay') accompanied by the discourse participant *te* 'you', in a discussion between pilots and trainee pilots.

- (11) el descanso es depende del trabajo que tengas si vuelas una aerolinea o si vuelas un privado si es de una aerolinea es por horas son ciertas horas al mes si vuelas mas *te pagan* mas y si vuelas menos te pagan lo mismo (...). (YAHOO)

'the rest depends on the job you have if you fly an airline or if you fly private if it is an airline it is by hours it is a certain number of hours per month if you fly more *they pay you* more and if you fly less they pay you the same (...).'

There is a transfer, of money in this case, from the agent (the person in charge of salaries in the airline) to the recipient (the pronoun *te* 'you'), which would refer to a person working as a pilot or a flight attendant. Based on this deictic use of *te*, this second person singular form can be interpreted in the context of the forum as entailing a broader reference possibly including other forum members and the speaker's own experience. In her scalar interpretation of reference of second person singular forms, Kluge (2012: 91) qualifies this use as "anyone but addressee as a typical representative". The discourse participant object occupies a recipient role, becoming the ultimate possessor. *Relational* verbs serve to characterize and to identify (Halliday & Matthiessen 2004: 210). This type of verb thus helps to establish the relationship among different entities (Ammara et al. 2019). *Relational* processes, and more particularly verbs of *possession*, are perfect candidates for the presence of an object since these verbs usually require reference to more than one entity, the main one being the subject. As explained by Halliday & Matthiessen (2004: 213), while clauses including a *material* process can appear with only one participant, *relational* verbs imply at least two participants. Our results indicate that in the examined construction these participants are mainly indirect objects.

Verbal types of processes express acts of saying, within which participants function either as sayer, verbiage, receiver or addressee (Halliday & Matthiessen 2004). *Communication* verbs are the second most frequent type of verb process with 23.2% (29 occurrences). This type of verb is a typical resource to transfer information from one participant to another. As explained by Pierre (2021), verbal types of processes "engage the speaker on the cognitive and communication level" (2021: 146). Example (12) illustrates a case of a non-anaphoric third person

plural used with a verbal process *dicen* 'tell' and including a discourse participant, which is, in this case, the second person singular *te* 'you'. The discourse participant pronoun has then a recipient role.

- (12) entre otras cosas, en el curso *te dicen* a ke velocidad debes despegar y aterrizar el avion ke te asignaron, a cuantos grados debes girar los flaps segun el clima y la pista, o ke velocidad de crucero debes mantener. en un determinado viaje.. osea la informacion basica suerte! (YAHOO)
'among other things, in the course *they tell* you at what speed you should take off and land the plane you have been assigned, how many degrees you should turn the flaps depending on the weather and the runway, or what cruising speed you should maintain on a given trip... that is the basic information, good luck!'

Thus, the two most frequent verb types place the discourse participant object in the recipient role, configured as an indirect object. This shows that the use of these object pronouns with an agent-defocusing strategy, leaving the subject underdetermined, typically configures roles in a context of transfer to a recipient, and not in an inversion of the agent-patient scheme as may be the case with some other agent-defocusing structures, such as the periphrastic passive.

In addition to processes of *possession* and *communication*, verbs denoting *space* also occur with discourse participants (17 occurrences out of 125, which represent 13.6% of our data).

- (13) [speaking about cleaning service]
H1 porque= / yo // en plan / estaba en= // en la habitación
<Pausa>
Speaker 1 'because I was in the room <Pause>'
H2 hm <Pausa>
Speaker 2 'hm <Pause>'
H1 y= aún no ac- / no aún no había acabado ¿sabes? // de= [de] //
estaba tomando el desayuno / y *me echaron* básicamente
¿sabes? me dijeron
Speaker 1 'and not yet / no I wasn't done yet you know? // of= [of] // I
was having breakfast / and *they* basically *kicked me out* you
know? they said to me'
H2 [hm]
Speaker 2 '[hm] <Pause>'

- H1 <cita>bueno vamos a limpiar la otra habitación <Pausa>
 Speaker 1 ‘<quote>well we’re going to clean the other room <Pausa>’
 H1 y venimos</cita> // ¿sabes? <Pausa>
 Speaker 1 ‘and we’re coming</quote> // you know // <Pausa>’
 H1 de esto que <Pausa> no habían pasado ni cinco minutos
 ¿sabes? // y ya habían venido entonces me fui para vues- /
 para= vuestra habitación
 Speaker 1 ‘from this that <Pausa> it wasn’t even five minutes later you
 know // and they had already come so I went to your- / to=
 your room.’ (ESLORA)

In Example (13), the object discourse participant, here the first person singular expressed through the pronoun *me* ‘me’, undergoes a change of location: an unspecified agent causes Speaker 1 to leave the room, thus representing the speaker as undergoing a (hostile) action by a group of agents that remains underdetermined. It then underlines the power relation between the patient and agent.

It has to be noted that the other types of processes present a low frequency of use. *Existential* verbs, which express processes linked to life, existence and phase-time relations, are even absent from non-anaphoric third person plural with discourse participants. This may seem logical since *existential* verbs typically have only one argument. Moreover, a low frequency of use of *existential* verbs was already found in non-anaphoric third person plurals, regardless of the presence of discourse participants (Pierre 2021: 204) (see also Tables 2 and 3).

A closer look at the type of verb processes engaged in non-anaphoric third person plurals co-occurring with a third person object pronoun reveals a relatively different picture. While *possession* verbs remain the dominant types of verb processes, they occur in a smaller proportion (28.9%) in comparison with non-anaphoric third person plurals with discourse participant objects (37.6%). In addition, Table 4 indicates that *communication* verbs are slightly less used with non-discourse participants. This category of verb presents a similar frequency of use as verbs expressing *spaces* (14), *changes* (15) or *other facts* (16). The latter two categories appear as rather typical of verbs occurring with non-discourse participants.

- (14) viaje por Air Canada, cuando llegue a CYYZ me no encuentre mi maleta, informe en el aeropuerto y me dieron un numero, llame y a los dos días *la vinieron a dejar* a la casa... creo que podrias hacer lo mismo o denunciar a la empresa... (YAHOO)

'I travelled by Air Canada, when I arrived at CYYZ I couldn't find my suitcase, I informed the airport and they gave me a number, I called and two days later *they came to drop it off* at the house... I think you could do the same or report the company...'

- (15) estube viendo un especial del A380, en NatGeo Channel de como lo hicieron desde el diseño hasta su construcción etc, está super interesante ese programa. hablaron de todo, lo de los wingets, como acortaron sus alas acortaron las medidas, donde *lo construyen*, todo lo que recorre para ir a la linea de ensamblaje final etc. (YAHOO)

'I was watching a special of the A380, on NatGeo Channel about how they made it from design to construction etc, it was very interesting this program. they talked about everything, the winglets, how they shortened the wings, where *they build it*, everything it goes through to go to the final assembly line, etc.'

- (16) Sin embargo en otras paginas encuentre que era un sistema que se diseño en los 60 para los modulos Apollo de la NASA y de ahí *lo utilizaron* en algunos tipos de misiles a principios de los 70 Honeywell lo ofrecio a Douglas y fue ahí que surgio primeramente montarlo en simuladores, IBM fabricaba los chips y tenian gran variedad de fallas comenzando con sobrecalentamiento. (YAHOO)

'However on other pages I found that it was a system that was designed in the 60's for Apollo modules of the NASA and from there *they used it* in some types of missiles in the early 70's Honeywell offered it to Douglas and it was there that it was first mounted in simulators, IBM manufactured the chips and they had a variety of failures starting with overheating.'

The verb phrase *vinieron a dejar* ('came to drop off') in (14) illustrates the category *space*, where the pronoun *la* which substitutes the suitcase (*maleta*) is moved from one place to another. In (15), the verb *construyen* 'build' implies a change of state, from non-existence to creation, including the necessary steps to be created. Example (16) illustrates a verb classified as 'others'. This rather heterogeneous category includes verbs referring to a physical type of action that does not meet the criteria to be related to changes, space or behaviour. In (16), the verb *utilizaron* ('used') denotes a physical manipulation but the patient, here the pronoun *lo*, does not suffer any modification.

Verbs expressing *spaces*, *changes* and *other facts*, pertain to the head category *material verbs*. Our analysis points, thus, towards the use of *material verbs*, which

refer to physical actions, as a key characteristic of occurrences with non-discourse participants. This ties in with the high presence of inanimate non-discourse participant object pronouns, which are more likely to be the object of a material verb.

As a preliminary conclusion, this leads us to assume that the type of verb and the type of participant involved (or not) in the discourse are closely associated. The results have shown that the use of discourse participant objects is favoured by the presence of *relational* verbs, followed by *communication* verbs. Non-discourse participant objects also appear with these categories, though in smaller proportions. Our analysis has indicated that this type of object clearly links to *material* verbs (*changes, spaces and other facts*), a characteristic not observed in the behaviour of discourse participant objects. Finally, non-anaphoric third person plurals used without any object pronouns exhibit a marked tendency for *communication* verbs and, to a lesser extent, for *space* verbs. The analysis has, thus, helped us highlight the importance of the type of verb in the variation of the presence and the specificities of object pronouns, as well as the types of verbs that are most used with the non-specific third person plural form.

4.2 Impact of the construction on the representation: A pragmatic-discursive approach

In this section, we will adopt a more pragmatic-discursive analysis and focus on the impact of the construction of a non-anaphoric third person plural with a discourse participant object on the conceptualization of the event. We will also discuss contrasts with non-discourse participant objects and the construction without an object.

Let us first focus on the utterances with a communication verb, one of the most frequent verb types with this construction. Siewierska & Papastathi consider that the non-episodic uses with *say* should be considered a separate type, rather than being considered as falling under the vague type proposed by Cabredo Hofherr (2003, 2006), which is linked to a specific moment in time (Siewierska & Papastathi 2011: 585). Other cases with communication verbs are, however, episodic. In many cases, the discourse participant object pronoun continues the reference of a deictic form in the preceding utterances (or is continued in what follows), showing the central position of the deictic forms for the development of topic continuity. This analysis of the broader discursive context shows indeed that, although the non-anaphoric third person plural form is the subject, the object pronoun with reference to a discourse participant is actually the form that anchors the utterance in the interaction, by referring to the speaker or addressee,

and that ensures the topic continuity with regard to the preceding and following parts of the interaction. Thus, from a discursive perspective, focusing on the discourse participant object is key to analyzing the utterance in the broader context. Example (17), an episodic use, illustrates this with a conversation concerning surgery on the broken leg of Speaker 1 (H1).

- (17) HI (...) me operé en enero mes y medio *me llamaron* [con la historia d]el seguro
Speaker 1 (...) 'I got surgery in January a month and a half later *they called me* [about] the insurance'
H3 claro.
Speaker 3 'of course.'
HI ¿qué pasa? / que como me la había roto antes,
Speaker 1 'What happens? That, since I had broken it previously,'
H3 [sí]
Speaker 3 'Yes'
HI era un seguro de fractura
Speaker 1 'it was a fracture insurance (...)' (ESLORA)

The first person singular object pronoun continues the narrative about being operated on (*me operé* 'I got surgery') and is further taken up when recounting a previous fracture (*me la había roto antes* 'I had broken it previously'). We see then a clear configuration where the object pronoun refers to one of the discourse participants and constitutes the main thread of the narrative through co-reference with previous mention of the speaker. The non-anaphoric third person plural can be interpreted through contextual information (*con la historia del seguro* 'about the matter of the insurance') as referring to the insurance company and its actions towards the discourse participant, showing the importance of the agent-defocusing strategy for the discursive development.

However, not all 1st and 2nd person object pronouns combine deictic anchoring with establishing topic continuity through coreference. Indeed, in various cases, the discourse participant object does not establish a coreference with preceding or following references, as in (18). It is then the main element that anchors the utterance in the interaction.

- (18) No es una planta inteligente. Tiene formas de supervivencia, pero de ninguna manera inteligencia. O sea *te mintieron!!!* (YAHOO)
'It is not an intelligent plant. It has forms of survival but by no means intelligence. So *they lied to you!!!*'

The third person object pronouns that appear with communication verbs are mainly the direct object, that is the topic of communication. Only 11 are indirect objects, typically anaphorical, and have an antecedent that is linked to the discourse participants via a possessive pronoun, e.g., *mi amigo* in (19). Thus, while the third person object pronoun is ensuring topic continuity, the presence of a deictic form in the wider context establishes a link with the speaker, thus showing again that, while the non-anaphoric subject is underdetermined, the contribution of Speaker 1 as a whole is clearly tied into the ongoing interaction.

- (19) H1 bueno / entonces / mi amigo ~Diego y su amiga ~Nuria
Salgado decidieron presentarse
Speaker 1 ‘well / then / my friend Diego and his friend Nuria Salgado
decided to be candidates.’
H2 ¿y los cogen?
Speaker 2 ‘and do they take them?’
H1 por qué / nadie lo sabe // pero en plan= <Pausa> hizo= / o sea /
se- llamaron por teléfono *les hicieron* una entrevista rápida por
teléfono // y les dijeron que ya les avisarían y esa misma tarde
les mandaron un correo // con un cuestionario de noventa
preguntas cada u- / noventa y cuatro preguntas cada uno
Speaker 1 ‘why / no one knows // but as a way of... <Pause> he did / so /
they- they called by phone *they interviewed them* quickly by
phone // and they told them that they would inform them and
that same afternoon they sent them a mail // with a
questionnaire of ninety questions each / ninety-four
questions each’ (ESLORA)

If we contrast these uses with non-anaphoric third person plural forms with communication verbs but without an object pronoun, as in (20), we see a different picture. This is in addition a non-episodic use.

- (20) H1 [la se]rie
Speaker 1 ‘[the ser]ies’
H2 [ah] <Pausa>
Speaker 2 ‘[ah] <Pause>’
H2 no sé <Pausa>
Speaker 2 ‘I don’t know <Pause>’

H1 *dicen* que es muy buena también

Speaker 1 ‘*they say* that it’s really good as well.’ (ESLORA)

In those cases where there is no topic continuity nor deictic anchoring, the non-specific nature of the subject pronoun becomes the dominant feature. It functions then as a kind of evidential strategy, referring to hearsay but without further information. This hearsay meaning is also present in the uses with an object pronoun, but the presence of the concrete object pronoun puts the focus on the (highly specific) recipient of the message.

When investigating the non-anaphorical third person plurals overall, it should be noted that some of the second person singular object pronouns allow a reading that is not merely deictic. Indeed, as pointed out by Posio (2016: 4), in a so-called impersonal use of the second person singular “the speaker may be included or excluded and the reference may concern either a group of people or an individual”. Following Kluge (2012: 89), we refer to the generic use of the second person singular and opt for “a scalar model of referentiation of the second person singular, with five more or less well-defined focal points”. This proposal ranges from a speaker reference *I* hiding behind *you* over anyone (a generic use) to *you* as term of address, with intermediary forms where respectively *I* or *you* are representative of a larger entity. While most cases included in our data deictically refer to the hearer, some cases include other uses on the scale, where the position of the discourse participant merits further discussion. Thus, (21) illustrates a reference to ‘anyone’, clearly not anchored in the speaker or hearer’s personal experience, since the interlocutors conclude they will have to go one day to this club, revealing that they do not have a concrete experience yet. However, the link to the discourse participants remains present in that they are discussing their own options to go there.

- (21) H1 me dijo eso que que *te* [*cobran* bastan]te
Speaker 1 ‘he told me that that *they* [*charged you* quite a lot]’
H2 [cucadas] <Pausa>
Speaker 2 ‘[cute things] <Pause>’
H1 pero no sé <Pausa> habrá que ir un [día]
Speaker 1 ‘but I don’t know <Pause> we’ll have to go one [day].’
(ESLORA)

Example (22), by contrast, refers to the speaker as a representative of a larger entity, which may include the hearer. Indeed, the speaker narrates a personal

experience as advice for the hearer. The generic reading then does not at all preclude a reference to the discourse participants. Quite the contrary, it often involves both speaker and hearer.

- (22) H1 un trenecito turístico de Monforte a Orense que te costaba= //
(no sé) // veinte euros / creo que era
Speaker 1 ‘A tourist train from Monforte to Orense that costed you I don’t know twenty euros I think it was’
H2 [hm / hm / hm / hm] hm / hm
Speaker 2 ‘hm hm hm hm hm hm’
HI te subías al tren / *te llevaban* de Monforte a Orense / antes de llegar a Orense te hacían un recorr- / hacían un recorrido por to da= la = / ciudad
Speaker 1 ‘you got on the train *they took you* from Monforte to Orense before reaching Orense they took you for a tour through the town.’ (ESLORA)

In Example (23) Speaker 1 fears electricity will be cut off due to late payment. Speaker 2 explains that advance warning is given, addressing this to Speaker 1 but also representing more general information concerning how electricity companies work.

- (23) H1 [no si / me puedes dar para la fac]tura de la luz [que no sé cuánto será] y que mañana a primera hora [tengo que pagar
Speaker 1 ‘[not if you / can give me for the light bill, since I don’t know how much it will be] and tomorrow first thing [I have to p]ay’
H2 [<inint>] H2 [<inint>] <Pausa>
Speaker 2 ‘(not understandable) <Pausa>’
H1 ¿qué hago? llamo= / y digo que no me llegó el recibo / que me den algo pa- / un código para [pagar o]= en /en internet ¿no?
Speaker 1 ‘What do I do? I call and say that the receipt didn’t arrive that they give me a code to [pay or] on internet, didn’t they?’
H2 [claro] <Pausa> H2 <inint> <Pausa>
Speaker 2 ‘Indeed <Pausa> (not understandable) <Pausa>’
H1 antes de que salgan a cortarla porque no te digo yo que no vengán mañana [a cortarla ¿eh? // de] hecho no te aseguro yo que no hayan venido ya y que no hayan encontrado el por[tal] </dud>

Speaker 1 ‘before they go out to cut it off because I don’t tell you that they won’t come tomorrow [to cut it off] actually I can’t assure you that they haven’t come already and that they haven’t found the por[tal]’

H2 [<inint>] H2 [no] no / te tienen / *te tienen* que dar un aviso / te dan un aviso

Speaker 2 ‘(not understandable) no no *they have to give you a notification they give you a notification*’ (ESLORA)

Again, though the reference is larger than a strictly deictic one, it does involve one of the discourse participants concretely and thus maintains a deictic anchoring. It falls under the use described by Kluge (2012: 89) as “anyone, but addressee as a typical representative”, since the addressee’s concrete situation is the starting point for a reference that can cover more people but in which the addressee remains included.

Thus, these cases where a second person with a not exclusively deictic use appears still are to be considered as references to discourse participants and by no means make the whole construction impersonal. Overall, the agent-defocusing effect of the non-anaphoric third person plural forms then entails a more prominent position for the (discourse participant) object pronoun as compared to the less prominent non-anaphoric third person plural form, rather than a low referentiality for the utterance as a whole. Indeed, these discourse participant pronouns are then the main reference in the ongoing interaction, relating to one of the interaction participants, and as such occupy a crucial position.

4.3 Mode and register variation

The literature shows that the use of non-anaphoric third person plurals, regardless of the presence of a discourse participant object, is specific to spontaneous interactions (Siewierska & Papastathi 2011: 585). Pierre (2021: 118) indicates that this non-referential mechanism appears more typically in informal oral mode, though the mechanism is still relatively frequent in written informal productions. A closer look at the mechanism used with object (non)-discourse participants confirms the tendency to appear in discursive situations considered informal (see Table 5). However, it seems that the preference for appearing in oral rather than written types of data is less marked when the non-anaphoric third person plurals co-occur with object pronouns. Indeed, the results reveal a normalized frequency of 21.9 occurrences per ten thousand words in the oral data used for this study and 18.2 in the written data used in this study.

Table 5: Distribution of (non)-discourse participants in formal and informal data (raw frequencies (normalized per ten thousand words)). Frequencies (“freq.”) are given twice, as both raw and normalized.

	Discourse participant				Non discourse participant (only)		Total	
	1 st person		2 nd person		3 rd person			
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
<i>Informal data</i>								
Oral (ESLORA)	43 (5.6)	25.6	38 (4.9)	22.6	87 (11.3)	51.8	168(21.9)	100
Written (Yahoo Q&A)	11 (2.4)	12.9	31 (6.6)	36.5	43 (9.2)	50.7	85 (18.2)	100
<i>Formal data</i>								
Oral (PROCEP)	2 (0.5)	100	0	0	0	0	2 (0.5)	100
Written (Wikipedia)	0	0	0	0	0	0	0	0

Table 5 provides more details on the distribution of object pronouns across the four types of language production used to collect the data of the present study (two different modes and two different registers). Since, as previously explained, nineteen occurrences combine two participants, where at least one of them is a non-discourse participant, Table 5 only includes non-discourse participants occurring alone, that is, without the simultaneous presence of another participant. This prevents counting twice those utterances that contain a combination of participants and allows us to focus on how the register and the mode impact the presence of discourse versus non-discourse participants.

As shown in Table 5, it is especially the first person singular form that appears in the informal spoken mode (25.6% of all object pronouns occurring in informal oral data compared to 12.9% in informal written data), whereas the second person singular form appears more in the informal written mode (36.5% and 22.6%, respectively). This finding thus points towards a more interlocutor-oriented approach in written productions than what is observed in conversational situations. This finding can be related to the specific advice-giving function of the Yahoo Q&A forum, where addressing the interlocutor is a crucial feature. As shown in the previous section, this second person form is not always to be interpreted in a purely deictic way, though. The dominance of the first person object pronoun

in informal oral mode ties in with overall research results concerning the use of deictics in informal oral interactions (see De Cock 2014: 35) and with the literature concerning person reference in different modes and registers. They thus highlight the impact of the nature of the language production.

The higher presence of first person pronouns in informal oral conversations suggests that this type of conversation engages the speaker much more than communicative exchanges achieved through the written mode. When looking at the research body on the use of reference to discourse participants in English, deictic forms have been associated with spoken mode, and non-deictic and canonical passive constructions with written mode (Biber 1988). The presence of deictics in spoken mode has also been pointed out by Chafe (1982), who formulates this tendency in terms of involvement, as opposed to the detachment reflected e.g. in the use of passives in the written mode. Ochs pointed out the higher use of passives in planned discourse vs. the preference for active constructions in unplanned discourse (Ochs 1979: 76). The constructions we focus on in this paper can be explained partly through these findings. Indeed, the use of a non-anaphoric third person plural form as an alternative to a passive form in informal conversation can be explained in part through the preference for active constructions in unplanned discourse, leading to the Spanish periphrastic passive being even less frequent in unplanned spoken discourse than it already is in other genres. When adapting Chafe's (1982) and Biber's (1988) ideas to Spanish, we have to take into account, however, that the periphrastic passive is much less used in Spanish than in English, since Spanish also has impersonal and passive constructions formed with the third-person reflexive clitic *se*, the latter being much more frequent than the periphrastic passive (see e.g., Laslop & Díaz 2010 or Pierre 2021), though in some, mainly informal, spoken genres, the non-anaphoric third person plural form is more frequent than the *se*-passive (De Cock 2014: 194, Posio 2015, Pierre 2021: 117–118). The non-anaphoric third person plural form also competes to some extent with these *se*-constructions (Siewierska 2011: 86). With regard to our informal written data, they seem to behave differently from the written mode commented upon by Biber (1988) and Chafe (1982), who looked into English formal written data. The written data from the Yahoo Q&A forum fall into Ochs' description of unplanned discourse and are informal (as can be seen also by the lexical choices and spelling), though, which explains the presence of deictics and the use of active constructions, rather than canonical passive ones or *se*-constructions. Indeed, research has shown (Pierre 2021) that periphrastic passives and *se*-constructions occur more frequently in formal written texts than in informal written texts, whereas the latter contain non-anaphoric third person plural forms, which the former lack. In addition, Posio (2015: 384) shows that there is no link between the degree of formality of the discourse and the pres-

ence of the *se*-constructions, whereas a high degree of contextuality favours the non-anaphoric third person plurals (Posio contrasts formality with contextuality, following Heylighen & Dewaele 2002).

Finally, Table 5 reveals that the use of non-discourse participants remains stable as they total 51.8% of object pronouns in the informal oral corpus and 50.7% of object pronouns in the informal written corpus. These results seem to indicate that the use of pronouns referring to discourse participants is more influenced by the type of language production than the use of pronouns referring to non-discourse participants.

Summarizing the behavior of non-anaphoric third person plurals occurring with vs. without pronouns referring to discourse participants in two registers and modes of productions, the following results can be put forward. Typically, the non-anaphoric third person plural is associated with informal situations, which confirms what is reported in the literature. Within oral data, discourse participants tend to be oriented towards the speaker whereas in written data they rather involve the interlocutor. It can thus be suggested that the distribution of participants is considerably impacted by the degree of formality and the mode of production of the language, but also by the specificities of the genre (informal oral conversations and written exchanges on a digital forum). However, as the figures are low, these tendencies need to be confirmed.

5 Conclusions

In this study, we have looked into the Spanish non-anaphoric third person plural form. Following earlier research, we focus on the verb types with which these forms are used. Given the non-referential nature of the subject, we have paid particular attention to the cases in which they appear with a referential object pronoun, be it a discourse participant or non-discourse participant.

Through an analysis of corpora representing informal and formal oral and written genres, we have shown that the non-anaphoric third person plural form is virtually absent in formal genres, which is in line with previous findings by, e.g., De Cock (2014), Pierre (2021). Siewierska & Papastathi (2011: 606) also argue that the structure is particularly related to spontaneous conversations. A more detailed analysis of the occurrences found in our datasets has shown that references to discourse participants tend to occur as indirect object in the roles of receiver or beneficiary, whereas the non-discourse participant objects are rather used as direct objects, in line with more general tendencies of indirect objects being typically animate. The non-discourse participant objects refer mainly to inanimate entities (altogether another type of referent than the necessarily an-

imate discourse participants). Note that the animate non-discourse participant objects mainly refer to animals, rather than humans.

The analysis of the verb types has shown that the use of discourse participant objects in the examined construction is favoured by the presence of relational verbs, followed by communication verbs. This furthermore ties in with the use of discourse participant object pronouns as indirect objects in a receiver or beneficiary role. As such, both the verb semantics and the thematic role of the object pronoun play a role. Also, non-anaphoric third person plurals without any object pronoun are frequently used with communication verbs, regularly in a non-episodic use. These results support Siewierska & Papastathi's (2011) suggestion to consider non-episodic uses of speech act verbs as a separate category in the study of non-anaphoric third person plural forms.

Our pragmatic-discursive analysis sheds light on the impact of using a referential discourse participant object pronoun with a non-referential subject pronoun. In the absence of a referential subject, it is above all these deictic object forms that ensure the anchoring in the ongoing interaction and frequently also ensure topic continuity. The non-discourse participants, third person objects, on the other hand, tend to be anaphoric but about one third of the occurrences in our corpus include a deictic reference by means of a possessive pronoun in the object or in a coreferential object to which it refers. Some pronouns referring to discourse participants are not used with a merely deictic reference, but also allow for a generic reading. However, even in such cases the link with one or more discourse participants remains present and such utterances are then not to be considered entirely impersonal.

The specificities of the genres analyzed explain the preference for first person object pronouns in informal conversation and second person singular object pronouns in the Yahoo Q&A data, where participants answer questions. These results also show that it is the informal nature of the data, rather than their being written or spoken, that influences the presence of non-anaphoric third person plural forms, since the written and spoken informal datasets present similar frequencies of non-anaphoric third person plural forms.

Through this study of non-anaphoric third person plural forms with particular attention to their use with discourse participant object pronouns, we have aimed to contribute to the literature, which has hitherto focused mainly on the non-referential subject. By examining referential objects and the verb types with which the non-anaphoric third person plural forms appear, we hope to have contributed to a more complete image of how discourse participant objects are used with these forms, as well as to the place they hold in the development of interaction.

Corpora

Corpus para el estudio del español oral: <http://eslora.usc.es>, versión 2.0 de septiembre de 2020, ISSN: 2444-1430.

Proceedings from European Parliamentary debates (PROCEP): <https://www.europarl.europa.eu/plenary/en/debates-video.html>

Wikicorpus V.1.0: Catalan, Spanish and English portion of Wikipedia, <https://www.cs.upc.edu/~nlp/wikicorpus/>

Yahoo Contrastive Corpus of Questions and Answers: Compiled by Hendrik De Smet at the Department of Linguistics, University of Leuven, 2009

References

- Ammara, Um-e, Rehana Yasmin Anjum & Maryiam Javed. 2019. A corpus-based Halliday's transitivity analysis of *To the Lighthouse*. *Linguistics and Literature Review* 5(2). 139–162. DOI: 10.32350/llr.v5i2.
- Barcala, Mario, Eva Domínguez, Alba Fernández, Raquel Rivas, Paula Santalla, Victoria Vázquez Rozas & Rebeca Villapol. 2018. El corpus ESLORA de español oral: Diseño, desarrollo y explotación. *Chimera: Romance Corpora and Linguistic Studies* 5(2). 217–237.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Briz, Antonio. 2010. El registro como centro de la variedad situacional: Esbozo de la propuesta del grupo Val.Es.Co. sobre las variedades diafásicas. In Irene Fonte Zarabozo & Lidia Rodríguez Alfano (eds.), *Perspectivas dialógicas en estudios del lenguaje*, 21–56. México: Universidad Autónoma de Nuevo León, Facultad de Filosofía y Letras.
- Cabredo Hofherr, Patricia. 2003. In Matthias Weisgerber (ed.), *Proceedings of the conference Sinn und Bedeutung 7 (SuB7)* (Arbeitspapiere des Fachbereichs Sprachwissenschaften 114), 81–94. <http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-11039>.
- Cabredo Hofherr, Patricia. 2006. “Arbitrary” pro and the theory of pro-drop. In Peter Ackema, Patrick Brandt, Maaïke Schoorlemmer & Fred Weermann (eds.), *Agreement and arguments*, 230–258. Oxford: Oxford University Press.
- Chafe, Wallace. 1982. Spoken and written language: Exploring orality and literacy. In Deborah Tannen (ed.) (*Advances in discourse processes* 9), 35–53. Norwood, NJ: Ablex.

- De Cock, Barbara. 2014. *Profiling discourse participants: Forms and functions in Spanish conversation and debates* (Pragmatics & Beyond New Series 246). Amsterdam: John Benjamins.
- De Smet, Hendrik. 2009. *Contrastive corpus of questions and answers*.
- Fernández Soriano, Olga & Susana Táboas Baylín. 1999. Construcciones impersonales no reflejas. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, vol. 2, 1723–1778. Madrid: Espasa.
- García-Miguel, José M. & Francisco J. Albertuz. 2005. Verbs, semantic classes and semantic roles in the ADESSE project. In Katrin Erk, Alissa Melinger & Sabine Schulte im Walde (eds.), *Proceedings of the interdisciplinary workshop on the identification and representation of verb features and verb classes*. <https://web.archive.org/web/20140514083908/http://adesse.uvigo.es/textos/saarb05.pdf>.
- Gelabert-Desnoyer, Jaime. 2006a. La deixis espacio-temporal en el lenguaje parlamentario español contemporáneo. *Círculo de Lingüística Aplicada a la Comunicación* 26. 17–52.
- Gelabert-Desnoyer, Jaime. 2006b. Registro y funciones de *nosotros* en el discurso parlamentario español. *Lingüística en la Red* 4. 1–21.
- Halliday, Michael. 1994. *An introduction to functional grammar*. 2nd edn. London: Hodder Arnold.
- Halliday, Michael & Christian Matthiessen. 1999. *Construing experience through meaning: A language-based approach to cognition*. London: Cassell.
- Halliday, Michael & Christian Matthiessen. 2004. *An introduction to functional grammar*. 3rd edn. London: Routledge.
- Halliday, Michael & Christian Matthiessen. 2014. *Halliday's introduction to Functional Grammar*. 4th edn. Oxford: Routledge.
- Heylighen, Francis & Jean-Marc Dewaele. 1999. Internal report, Center “Leo Apostel”, Free University of Brussels. Formality of language: Definition, measurement & behavioral determinants.
- Heylighen, Francis & Jean-Marc Dewaele. 2002. *Variation in the contextuality of language: An empirical measure*, vol. 7. 293–340.
- Kluge, Bettina. 2012. *Referential ambiguity in interaction: Establishing generic reference with second person pronouns in the Romance languages*. Universität Bielefeld. (Habilitation thesis).
- Laslop, María Eugenia & Axel Hernández Díaz. 2010. Impersonalidad y pasividad en las normas sobre derechos humanos: El uso de *se* y de pasivas perifrásticas en las constituciones mexicanas (1824–2001). *Miscelánea de lingüística iberoamericana* 8(2). 203–231.
- Ochs, Elinor. 1979. *Planned and unplanned discourse*. Talmy Givon (ed.) (Syntax and Semantics 12). New York City: Academic Press. 51–80.

- Pierre, Émeline. 2021. *Agent-defocussing structures in Spanish, English and French: A contrastive corpusbased account of the interaction between different passive-like structures*. Université catholique de Louvain. (Doctoral dissertation).
- Posio, Pekka. 2015. Alcance referencial y variabilidad de las construcciones impersonales con referencia humana en español peninsular hablado: Se y la tercera persona del plural. *Spanish in Context* 12(3). 373–395. DOI: 10.1075/sic.12.3.03pos.
- Posio, Pekka. 2016. You and we: Impersonal second person singular and other referential devices in Spanish sociolinguistic interviews. *Journal of Pragmatics* 99. 1–16.
- Reese, Samuel, Gemma Boleda, Montse Cuadros, Lluís Padró & German Rigau. 2010. *Wikicorpus: A word-sense disambiguated multilingual Wikipedia corpus*. Valletta, Malta: European Language Resources Association (ELRA). <https://aclanthology.org/L10-1152/>.
- Shin, Naomi & Ricardo Otheguy. 2005. Overt non-specific *ellos* in Spanish in New York. *Spanish in Context* 2(2). 157–174.
- Siewierska, Anna. 2008. Impersonalization from a subject-centred vs. agent-centred perspective. *Transactions of the Philological Society* 106(2). 115–137. DOI: 10.1111/j.1467-968X.2008.00211.x.
- Siewierska, Anna. 2010. From third plural to passive: Incipient, emergent and established passives. *Diachronica* 27(1). 73–109. DOI: 10.1075/dia.27.1.03sie.
- Siewierska, Anna. 2011. Overlap and complementarity in reference impersonals: Man-constructions vs. third person plural-impersonals in the languages of Europe. In Andrej L. Malchukov & Anna Siewierska (eds.), *Impersonal constructions: A cross-linguistic perspective*, 57–89. Amsterdam: John Benjamins. DOI: 10.1075/slcs.124.03sie.
- Siewierska, Anna & Maria Papastathi. 2011. Towards a typology of third person plural impersonals. *Linguistics* 49(3). 575–610.
- Vázquez Rozas, Victoria, Mario Barcala, Eva Domínguez Noya, Alba Fernández Sanmartín, Guillermo Rojo & Paula Santalla. 2020. Codificación y anotación del habla en un contexto bilingüe: El corpus ESLORA de español de Galicia. In Ángel Gallego & Francesc Roca (eds.), *Dialectología digital del español: Anexo 80 de verba*, 189–224. Santiago de Compostela: Universidade de Santiago de Compostela, Servizo de Publicacións e Intercambio Científico.
- Weiner, Judith & William Labov. 1983. Constraints on the agentless passive. *Journal of Linguistics* 19(1). 29–58.

Part III


Referring to the addressee

Chapter 10

Referring to discourse participants in European Portuguese: The form of address *o senhor*

Isabel Margarida Duarte 

Universidade do Porto

Maria Aldina Marques 

Universidade do Minho

We will examine the uses of the noun phrase *o senhor* (formal ‘you’), as well as its linguistic and discourse status. As a form of address, it has acquired features that are typical of pronominal forms of address, with bleaching of semantic traits that point to an ongoing process of grammaticalization. In European Portuguese, despite being an issue that has been addressed several times, a comparison of the existing theoretical explanations has yet to be accomplished. Furthermore, its usage has not been analysed in different discourse contexts so as to attest to these changes. It is therefore necessary to revive and broaden the discussion. The data we have employed in this analysis is taken from the corpus *Perfil sociolinguístico da fala bracarense* (‘Sociolinguistic profile of Braga speech’), consisting of sociolinguistic interviews. We also built an ad hoc corpus, comprising political debates and interviews. In addition, for specific questions, some data was obtained from the CETEMPúblico corpus, and from the Davies & Ferreira corpus for diachronic data. The overall goal of this study is the analysis of the linguistic and discourse features of the address form *o senhor*. It is a qualitative approach, complemented by quantitative analysis of the occurrences recorded.

The results of our study show that *o senhor* is a hybrid form of address, revealing features from the two categories, the nominal form of address and the pronominal form of address. The confrontation of diachronic and synchronic data shows that the semantic values of the noun affect the current pragmatic values of the forms of address (FA).



1 Introduction

Nominal structures that show signs of a process of grammaticalization (cases like *a gente* ‘us’, *o senhor* ‘Mr/you’, *vossa excelência* ‘your excellence/your lordship/your grace’), changing from noun to pronoun, are a linguistic phenomenon that has been studied extensively in European Portuguese (EP) and Brazilian Portuguese (BP) (see, e.g. Cintra 1972, Nascimento 1987, Cook 1994–1995, Faraco 1996, Pereira 2003, de Paiva Sória 2013, among many others). Some of these processes certainly do not exhibit the same degree of grammaticalization in the two stabilized varieties of the Portuguese language. Although some of these structures in BP, such as *a gente* ‘us’, have been the focus of more studies than in EP, this does not mean that identical processes have not also occurred in EP, and several studies have been published on the topic (Nascimento 1987, Pereira 2003, Posio 2021, among others).

In the case of EP, although this issue has been addressed several times, a comparison of the proposed theoretical explanations of the form *o senhor* has yet to be accomplished, nor has its usage in different discourse contexts been analysed in order to identify the changes it has experienced. It is, therefore, necessary to revive and broaden the discussion, paying particular attention to the actual uses of *o senhor*.

In this chapter, we aim to gain new insights into the categorization and uses of the noun phrase (NP) *o senhor*, as well as its linguistic and discourse status.¹ As a form of address (FA), we intend to show that *o senhor* is a hybrid form, whose usage is at times closer to nominal forms of address (NFA) and at others closer to pronominal forms of address (PFA).

Our analysis is based on a qualitative approach, due to the theoretical need to consider the linguistic and extralinguistic contexts (situational, social, experiential) in the description and explanation of how the object under study works in discourse. This approach will be complemented by a quantitative examination of the occurrences recorded in the corpus and an analysis of the results.

This chapter is organised as follows. After this brief Introduction (§1), the state of the art is presented in §2, starting with an overview of the address system in Contemporary European Portuguese (§2.1), followed by a review of the main theories and previous studies on *o senhor* (§2.2). The theoretical and methodological framework on which our analysis is based is described in §3. The following §4

¹We will not analyse the morphosyntactic variants *senhor/senhora*, *senhores/senhoras* ‘Mr/Mrs/Ms, sir/madam, gentlemen/ladies’ according to gender and number categories. In fact, these variations are more complex and involve the morpho-phonetic, syntactic, semantic and pragmatic levels that would require a different approach.

is devoted to the analysis of the form of address *o senhor*, taking into account also the vocative form *senhor*. This is a synchronic approach, completed by a diachronic comparison. Historical data is employed to explain the current functionalities of these forms. In §4.1, we present the changes the form has experienced from a diachronic perspective and, in §4.2, from a synchronic perspective, its occurrence and pragmatic-discursive features in different discourse genres, namely, in sociolinguistic interviews (§4.2.1), as well as in television interviews and electoral political debates (§4.2.2). The next section focuses on *o senhor* as the only form of address (§4.3) and the process of phonetic erosion that is currently taking place, namely in the terms *sô*, *sor*, *se*, and *sotor* (§4.4). This section ends with a discussion of the results (§4.5). In §5, we present the final considerations on the issues analysed, suggesting avenues for future research.

2 The address system in European Portuguese: State of the art

2.1 Proposals to categorize the forms of address in European Portuguese

The forms of address (FA) are a pragmatic category, central to the description and analysis of discourse organization, especially in the construction of interpersonal relations. In other words, they are “relationèmes” (Kerbrat-Orecchioni 1992: 37, 2010: 8). The importance of the role they play is evident in the bibliography on the topic, which encompasses a wide range of theoretical frameworks (Cintra 1972, Medeiros 1985, Carreira 1997, 2004, Hammermüller 2004, Duarte 2010, 2011, Kerbrat-Orecchioni 2010, and Scherre et al. 2015).

The FA have been systematically organized (Kerbrat-Orecchioni 2010: 8) into two subcategories as pronominal forms of address (PFA) and nominal forms of address (NFA), taking into account the lexical or pronominal nature of their members, as well as their different semantic and pragmatic features and functions. Amongst these, there are the semantic-pragmatic functions performed by their constituent elements, namely a deictic (personal) pronominal function and a nominal definitory function concerning the identity and other features of the addressee, prototypically performed by PFA and NFA, respectively.² They are

²According to Johnen (2014), “La distinction entre «déictique» et «définitoire» a été introduite par Bühler (1982 [1934]: 114–120), faisant lui-même référence à Apollonius Dyscole, pour saisir la différence entre pronoms (dont la fonction est déictique, car ils font partie du «système d’orientation subjective ici-maintenant-moi», Bühler 1982 [1934]: 149) et substantifs (dont la

complementary functions, in that the deictic function is prototypical of the FA category as a whole and the identifying function may be present in varying degrees of prominence. However, as individualizing prototypical features of each category, the deictic function is specific to PFA, whereas the definitory function characterizes the NFA category (Carreira 1997, 2004, 2007, Kerbrat-Orecchioni 2014).

For European Portuguese, we highlight three major theoretical proposals. The first is the work of Cintra (1972), who identifies three FA categories. This is a particularity of the Portuguese address system in comparison to other Romance languages like French (Nascimento et al. 2018, Duarte & Marques forthcoming). Diverging from the dichotomous model of Brown & Gilman (1960), and taking into account the subject function of FA, Cintra analyses and divides the address system in European Portuguese into three categories, pronominal (PFA), nominal (NFA) and verbal (VFA) forms of address (see Table 1 regarding Cintra's tripartite morphosyntactic categorization). They are organized according to an individual or collective addressee and the interpersonal relations established, of greater or lesser intimacy, or of greater or lesser hierarchy and deference. The VFA, constituted only by the verbal form, marks in the 3rd person the zero degree of deference, as a strategy of avoiding the specificities that govern the FA choices for EP speakers (Carreira 1997, Hammermüller 2004, Duarte & Marques forthcoming).

Table 1, Cintra's tripartite morphosyntactic model, shows the complexity of the FAs in Portuguese, whose translation into English is reduced, in almost all cases, to the use of the form "you" (formal or informal), such as: *Quer?*/ Do you want?; *O senhor quer?*/ Do you want?; *O António quer?* Do you want?; *Queres?* Do you want? and so on.

In addition to this morphosyntactic categorization, the NFA category is broad and complex (Kerbrat-Orecchioni 2010: 7), encompassing not only the vocative uses (1) that are characteristic of English and French NFA, for instance (Formentelli 2009), but also uses with the syntactic functions of subject and complement (2), exclusive to PFA in these two languages.

- (1) Ana, trouxe o livro?
Ana PST.3SG the book?
'Ana, did you bring the book?'

fonction est définitoire, car ils caractérisent sémantiquement leurs référents)". "The distinction between 'deictic' and 'defining' was introduced by Bühler (1982 [1934]: 114–120), himself referring to Apollonius Dyscolus, to grasp the difference between pronouns (whose function is deictic, because they are part of the "subjective orientation system here-now-me", Bühler 1982 [1934]: 149) and nouns (whose function is defining, because they semantically characterize their referents)".

Table 1: Cintra's tripartite morphosyntactic categorization of EP (Cintra 1972: 11-12)

Pronominal FA	Nominal FA	Verbal FA
<i>tu (Tu queres?)</i>	<i>O senhor, a senhora, os</i>	<i>Quer? Querem?</i>
<i>vós (Vós quereis?)</i>	<i>senhores, as senhoras (o</i>	
<i>você (Você quer?)</i>	<i>senhor quer?)</i>	
<i>vocês (Vocês querem?)</i>	<i>O senhor Dr., o senhor</i>	
<i>Vossa Excelência (V. Ex.^a</i>	<i>Ministro (o senhor Dr.</i>	
<i>quer?)</i>	<i>quer?)</i>	
<i>Vossas Excelências (V.^{as}</i>	<i>O pai, a mãe, o avô (o</i>	
<i>Ex.^{as} querem?)</i>	<i>pai quer?)</i>	
	<i>O António, a Maria (A</i>	
	<i>Maria quer?)</i>	
	<i>O meu amigo, o patrão (o</i>	
	<i>meu amigo quer?)</i>	

- (2) A Ana trouxe o livro?
 ART.DEF.F.SG Ana.SBJ PST.3SG the book?
 'Did you bring the book?'

In turn, although maintaining Cintra's tripartite morphosyntactic categorization, Carreira (1997, 2004, 2007) proposes a verbal proxemic criterion that organizes the address forms into a continuum from proximity to social distance. She also develops a new definition of the address system in EP, which integrates locutive and delocutive forms, besides the traditional allocutive forms, concerning, respectively, the designative forms of the speaker and of others as objects of discourse (Carreira 1997).

Finally, Medeiros/Oliveira (1985, 1992, 2004)³ is noteworthy in her reorganization of Cintra's categories into *pure pronouns*, *pro-pronouns* and *zero forms*. The researcher brings to the discussion the sociolinguistic categories of power and solidarity established by Brown & Gilman (1960), aggregated to informal and formal forms of address (T/V, according to the Latin system), to propose a more comprehensive, psycho-sociolinguistic model of the forms of address. According to the author, Brown and Gilman's theoretical model is only applicable to "conventionalized forms of address" and, therefore, unable to explain the complexity of the phenomenon, contrary to the model she proposes (Medeiros 1992: 340).

³Medeiros and Oliveira refer to the same author.

Noting that there is a process of negotiation of address forms in verbal interactions, Medeiros (1992: 335/338) emphasizes how idiosyncratic issues determine the choices made by speakers. In fact, she proposes a model that takes into account the contextual variability based on the idiosyncrasies of the participants in the interaction. Rooted in the concept of negotiation, her theoretical proposal is a fundamental contribution to the study of the FA, which forces us to consider an experiential, idiosyncratic dimension that governs the speakers' choices. Formentelli (2009) came to the same conclusion, after investigating the forms of address in an academic environment.

2.2 Previous studies on the categorization of the form of address (*o senhor* in contemporary European Portuguese

In addition to the above-mentioned individualized categories according to their lexical, pronominal, or verbal nature, researchers have identified other features that underlie new categories, namely the syntactic distribution of the form of address. Thus, for instance, in relation to the English address system, Formentelli (2009: 182) identifies NFA with the vocative category (which has its own international profile), highlighting the high productivity rate of this open category, with a very free distribution within the utterance. However, this transposition to EP raises some theoretical questions. The NFA in EP do not present the same syntactic restrictions as in English or in French. This is one of the reasons why the form *o senhor* is subject to different categorizations.

As an address form, (*o senhor* is a general appellative (3–7), which may occur aggregated to other forms of nominal address (anthroponyms, such as first/last name) (3–4); functional appellatives (5), such as academic titles, professional titles, positions, etc., and as the NP *o senhor* (7), and as a single form or combined with the aforementioned nominal forms (6):

- (3) Bem, *senhor Vicente*, ficamos por aqui.
Well, Mr. Vicente, we can end here.
'Well, Mr. Vicente, I think we can end here.' [PSFB interview 12]⁴
- (4) Olhe, *sr. Machado*, acha que em Lisboa se fala da mesma forma
Look, Mr. Machado, PRT.3SG in Lisbon people speak in the same way
que em Braga?
as in Braga?
'Look, Mr. Machado, do you think people in Lisbon speak in the same way as in Braga?' [PSFB interview 22]

⁴These examples are taken from the corpora analysed (see §4.2).

- (5) Eu?! Ó *senhor engenheiro*...
 Me?! voc Mr Engineer...
 ‘Me?! Really, Mr Engineer...’ [debate, 2011]
- (6) - *Não foi o Partido Socialista. Foi o Eurostat.*
 ‘It wasn’t the Socialist Party. It was Eurostat.’
 - Foi o *senhor engenheiro José Sócrates*...
 PST.3SG ART-DEF-M-SG Mr Engineer José Sócrates...
 ‘It was you (the Mr Engineer José Sócrates)...’ [debate, 2005]
- (7) Se o *senhor* for eleito o que é que o *senhor*
 If ART-DEF-M-SG Mr PASS.3SG what ART-DEF-M-SG Mr
 fará?
 will do?
 ‘If you are elected, what will you do?’ [interview, 2010]

The address form (*o*) *senhor* occurs as a vocative in (3), (4) and (5), but also with an allocutive function in the syntactic position of subject in (6) and (7), a trait associated with the address pronouns in English and French (Formentelli 2009, Kerbrat-Orecchioni 2010, Johnen 2014).

It is on the basis of these classifications and the discussions generated around the status and features of the NFA and PFA that we should examine the categorizations proposed for the FA *o senhor*. First, the distinction of syntactic contexts, which characterizes the functionalities of the FA (nominal and pronominal) in English and serves to distinguish between the PFA category and the syntactic function of the pronoun, is not valid for Portuguese (see Johnen 2014: 376). In some Romance languages, including Portuguese, third person forms are used in allocation to refer to the second person, which Pountain calls a “third-as-second person form” (Pountain 2003: 149–150). Heine & Song (2011) draw attention to other languages where the same phenomenon occurs. As Pountain rightly points out, this structural possibility leads to a very open pronominal system. In EP, third-as-second person forms can occur with either a delocutive or allocutive function. In these contexts, forms that can occur both with allocutive and delocutive values intersect. The immediate context and the global context are essential for the disambiguation of the function. So, (*O*) *senhor*, as a FA, is distinct from the uses of (*el*) *señor* in Spanish, a language that otherwise shares many affinities with Portuguese. Castillo Lluch (2014: 264) is adamant when she states that the form (*el*) *señor* in its allocutive function does not occur with a definite article, contrarily to its delocutive use.

The tripartite categorization of the FA in EP is accepted by researchers, despite minor differences in the delimitation and designation of each category. However, the inclusion of the FA *o senhor* is more problematic, as we will see next (§4). In fact, due to its traits, this FA shows the porosity of the categories, the continuum of values, and can, therefore, be framed in different categories.

Cunha & Cintra (1984: 292) include the FA *o senhor* in the category of address pronouns, which also includes the forms *você*, *vossa excelência* or *vossa senhoria* ‘you, your grace, or your lordship’. The category is constituted by “... certain words and locutions that are equivalent to actual personal pronouns, such as *você*, *o senhor*, *Vossa Excelência*”. Note that Cintra, in an earlier work (1972), places *o senhor* in the nominal address forms (saying that *o senhor* and *a senhora* are the most pronominalized of these forms, Cintra 1972: 12). Medeiros (1985) considers *o senhor* a pro-pronoun, but *você* a pure pronoun, whereas Preti (2004: 184) distinguishes two pronoun subcategories for BP, pronoun forms and pronominalized forms, and includes the FA *o senhor* in the latter: “...pronominalized forms, that is, with personal pronoun value (*você*, *o senhor*, *Vossa Excelência*, *Vossa Senhoria* and its variations)”. Ilari et al. (1996: 184) also argues that the “...set of personal pronouns in Portuguese (...) includes, in the second person, *o senhor/a senhora*”. In Raposo (2013: 900), the author speaks of “pronominal locution”, as it consists of two elements (unlike pronouns).

Diverging from these categorizations, Nascimento et al. (2018: 248) consider *o senhor* a nominal form, albeit “equivalent to the 3rd person paradigm of *você* (*Você quer? O senhor quer?/Do you [-formal] want? Do you [+formal] want?*)”. Table 2 summarizes the different terms used by the authors discussed here.

Table 2: Categorization of the address form *o senhor*

	Author					
	Cintra	Medeiros	Preti	Ilari	Nascimento	Raposo
Nominal form of address	✓				✓	
Pronominalized form			✓			
Pronominal locution						✓
Pro-pronoun		✓				
Pronoun of address	✓			✓		

For all the researchers cited, on a list that is far from exhaustive, there is an identification, approximation or parallelism of the address form *o senhor* to the PFA category. Although the designations may obliterate this issue, the classification of the NP *o senhor* as a pronoun does not derive from a purely decontextualized grammatical classification, but rather from one of its heterogeneous discourse uses. The central conclusion to be drawn from this literature review is that *o senhor* is a fuzzy category.

3 Theoretical and methodological framework

Taking into account the contributions discussed in §2, we have adopted a pragmatic-discursive theoretical approach. We focus on discourse genres, as a nuclear research concept, with an emphasis on oral verbal interactions characterized by different degrees of formality, as a central factor to consider in the analysis of the variety of address forms in EP and the contexts in which they occur (Marques 2014). Besides the authors mentioned previously, the works of Kerbrat-Orecchioni (1992, 2005) constitute the basis for our approach to forms of address in the construction of interpersonal relations in discourse interactions. Assuming, therefore, an interdisciplinary perspective, and in order to explain some of the uses of the NP *o senhor* as a form of address, we have also used the grammaticalization theory by Traugott & Heine (1991), Lehmann (2015), and subsequent developments since then, such as Heine & Kuteva (2004: 17), who established four criteria of grammaticalization, as follows: “(a) desemantization (or “semantic bleaching”) – loss in meaning content, (b) extension (or context generalization) – use in new contexts, (c) decategorialization – loss in morphosyntactic properties characteristic of lexical or other less grammaticalized forms, and (d) erosion (or “phonetic reduction”) – loss in phonetic substance.”

As mentioned previously, the address forms are sensitive to local and global contexts: thus, a qualitative analysis of the data collected is required, so as to identify the pragmatic meanings they bring to discourses. However, we will combine it with a quantitative analysis, in a complementary perspective, which will serve to show the changes of their uses over time, and the predominance of certain forms of address in a certain genre.

Having examined different oral and written discourse genres that exemplify the diversity of address forms in Portuguese, we have selected the NP (*o*) *senhor* in its pragmatic function of allocution as the object of our analysis. Phonetic and morphosyntactic issues will also be considered in the analysis of the occurrences and uses of this form.

The main questions of this research are related to the linguistic and discourse status of the address form (*o*) *senhor*, the semantic and pragmatic features that stand out in its use, both diachronically and synchronically, and the contexts and frequency of occurrences of this form, according to the discourse genre. In order to answer these questions, our expected results are that: (1) *O senhor* is a frequent form of address in contemporary Portuguese; (2) limited by its lexical origin, the form (*o*) *senhor* includes a pragmatic trait of respect; (3) in contemporary Portuguese, the use of the form has become widespread, marking a relationship of social distance; (4) the nominal form of address *o senhor* is in a process of grammaticalization in EP.

Thus, the goals of this chapter are (1) to determine the linguistic and discursive features and uses of *o senhor* from both the diachronic and synchronic perspectives, in order to (2) identify and analyse its pragmatic-discursive functions; (3) to identify features of use that support its classification as an address form in a process of grammaticalization towards pronominalization. To perform the analysis, we have employed data from the corpus *Perfil Sociolinguístico da Fala Bracarense* (PSFB), ‘Sociolinguistic Profile of Braga Speech’, built from sociolinguistic interviews. From this corpus, consisting of 80 interviews of about 60 minutes each, we have selected interviews with male informants ($N = 36$). Taking into account the occurrences of the forms *tu* and *o senhor*, the interviews were grouped into two categories (those that used the FA *tu* and those in which the participants used the FA *o senhor*) from which four interviews per category were randomly selected. These are interviews 1, 3, 5, 7, and 12, 22, 31 and 43, respectively.

An ad hoc corpus was also built, composed of televised political debates and interviews. They are verbal interactions that took place during the legislative elections in the early 21st century in Portugal. The corpus has about five and half hours of recordings distributed over the following interactions: three debates, held in 2005 (c. 90 minutes), 2011 (c. 45 minutes), and 2015 (c. 60 minutes); and three interviews, held in 2005 (c. 41 minutes and c. 40 minutes), and in 2009 (c. 46 minutes). Additionally, for specific issues, some data from the CETEMPúblico⁵ corpus and from *Corpus do Português* (Davies & Ferreira 2016–) were used as sources of written and diachronic data, respectively.

The variety of data selected provides both a diachronic and synchronic perspective on the uses of (*o*) *senhor*: firstly, written texts, from the 13th century to the present day, in the *Corpus do Português* (Davies & Ferreira 2016–) provide evidence of semantic and pragmatic changes; secondly, from a synchronic perspective, sociolinguistic interviews (PSFB), which cover diverse social groups, provide

⁵To account for occurrences in the written press as a source of endorsement, we also used, for three examples, the CETEMPúblico, a journalistic corpus.

evidence of informal orality in interactions with strong features of colloquialism; and finally, a corpus of political interviews and debates provides evidence of a more formal register. The diachronic perspective adds data that serve to better capture the synchronic functions, i.e., there are dominant semantic traits from other time periods that persist today in certain allocutive uses, although their meaning has changed significantly over time, as we will show in the next section. According to Dickey (1997: 257), this has also happened with French *monsieur*, English *Mister* (from ‘master’), and German *Herr*, for example. These FA no longer mean ‘older’, ‘wiser’, ‘more respectable’, as in *senior*, *lord* or *master*, nor ‘owner’, as in *landlord* or *proprietor*, as they once did, nor are they used exclusively to address the nobility, although they conserve traces of deference.

4 Analysis: The form of address *o senhor* in European Portuguese

4.1 The (*o*) *senhor* structure from a diachronic perspective

A diachronic perspective on how *o senhor* has shifted from NP to FA serves to contextualize the current functions of the FA and is the basis for the synchronic analysis conducted in §4.2. The data were collected from the *Corpus do Português* (Davies & Ferreira 2016–). The noun *senhor* comes from the Latin, *senex/senior* > *senhor* and its distinguishing value of deference and respect comes from this sense of ‘older, wiser, more respectable’, values that different societies attribute to a generational status. The occurrences available in the *Corpus do Português* (Davies & Ferreira 2016–), from which all pre-twentieth century examples were collected, point to the diachronic meaning of *o senhor* as a lexical item endowed with a certain content, organized into two interrelated semantic dimensions. From the 13th to the 18th century, *senhor* occurred mainly as a proper noun, in reference to God (8), and as a common noun, referring to someone of the male gender in a very high social position, owner of various types of assets (9).⁶ In this case, *o senhor* is synonymous of *dono* ‘owner/lord’: *o senhor do lagar*, *o senhor da herdade*, *o senhor do preito*, *o senhor da terra* ‘the lord of the press, the lord of the estate, the lord of the servants, the lord of the land’, as in example (9), taken from the *Terceira Partida de Afonso X* (1221–1284):

⁶Biderman (1972–1973) gives a detailed account of the forms of address in Portuguese from a diachronic perspective.

- (8) E ante que chegue ao logar hu diz por que *o senhor* todalas cousas cria. am dofereçer os clerigos o Olio da hũa das enpolas que dissemos que he ûtar os enfermos. [Corpus do Português]
'And before reaching the moment when they say "because the Lord creates all things", the clerics should offer the holy oils from one of the containers and anoint the sick.'⁷
- (9) Outrossy dizemos que se *o senhor* mãdasse ao seruo conprar algũa cousa ... [Corpus do Português]
'Also, we can say that if his grace were to send the servant to buy something...'⁸

From the 16th century onwards, there are frequent examples of the NP *o senhor* where it is used in honorific titles, preceding the designation of professions, anthroponyms, noble titles (10) or the proper name of a member of the upper class (11), as in the examples:

- (10) *O senhor* Rei D. Pedro tinha um couteiro em Alcântara... [Corpus do Português]
'His Royal Highness King Dom Pedro had a gamekeeper in Alcantara...'⁹
- (11) E *o senhor* Dom Alvaro yrmão do duque, E o duque e *o senhor* Dom Jorge postos a pee cada hum de sua parte levaram a princesa. [Corpus do Português]
'And his grace Dom Alvaro brother to the duke, And the duke and his grace Dom Jorge stood up and took the princess.'

These are all delocutive uses. In fact, in the *Corpus do Português* (Davies & Ferreira 2016–), very few occurrences of allocution are attested before the 19th century. The first occurrence (12) dates from the 17th century, in a literary text by the writer Francisco Manuel de Melo:

⁷Primeira Partida de Afonso X [Corpus do Português].

⁸Example (9) is taken from the 3rd Partida de Afonso X [Corpus do Português]. As a vocative, *senhor* appears only addressed to God, as in the Crónica General de España of 1344: "Ó Senhor Jhesu Cristo, cujo he o reyno e o inperio e todos os poderios som em tuas mãos!"/"O Lord Jesus Christ, whose is the kingdom and the empire, and all the powers are in your hands!". The use in religious discourse as a form of addressing the divinity is systematic and has continued to the present day.

⁹*Vida e feitos d'el-rey Dom João Segundo*, by Garcia de Resende (1533) [Corpus do Português].

- (12) *Vá-se já o senhor* muito embora, que,
 go-SBJ.3SG.=RFL.3 right now ART.DEF.M.SG sire because,
sendo destes senhores, poucas saudades nos deixará
 being one of these gentlemen, you will not be missed at all
 ‘You should leave right now, sire, because, being one of these gentlemen,
 you will not be missed at all.’ [*Corpus do Português*]

In the 18th century, the epistolary genre seems to create room for allocutive uses. However, once again, there are very few occurrences and not without some ambiguity between allocutive or delocutive use (13), given that only short excerpts are available which do not fully contextualize the uses found:

- (13) *Aqui perguntaria o Senhor* João se os arredores de Roma serão tão lindos como os do Porto?
 ‘Here would Mr João ask/Here would you [Mr João] ask if the surroundings of Rome are as beautiful as those of Porto?’¹⁰

The occurrences of *o senhor* increased and reached their peak in the 19th century, as Table 3 shows.

Table 3: Occurrences of *o senhor*, according to Corpus do Português (Davies & Ferreira 2016–)

Cronology	Occurrences	Per million
sXIII	28	50.82
sXIV	101	78.44
sXV	794	279.12
sXVI	867	200.08
sXVII	424	129.58
sXVIII	445	203.25
sXIX	4642	476.76
sXX	3920	193.44

In the 19th century, the form started to appear in dialogues in novels (14) and plays, as an allocutive form of address, not exclusive to the nobility, marking a formal relationship of social distancing:

¹⁰Cartas do Abade Antônio da Costa, 1744 [*Corpus do Português*].

- (14) O seu amigo é um canalha!... O *senhor* é
Your friend is a scoundrel!... ART.DEF.M.SG mister be.PRS.3SG
um homem de bem.
a good man
'Your friend is a scoundrel! ... You're a good man.'¹¹

Still in the 19th century, there are many instances of *o senhor* in a complex NFA, placed before a title, as in *o senhor pároco*, *o senhor cônego* (15), *o senhor administrador*, *o senhor doutor* literally, 'Mr Parish Priest', 'Mr Canon', 'Mr Administrator', 'Mr Doctor', etc.

- (15) - E *o senhor cônego* toma um copinho de geleia, sim?
'- And you, Mr Canon, will you have a cup of jam?'¹²

The use of *o senhor* as a form of address was reinforced in the 20th century (Biderman 1972–1973), providing thus more relevant contexts of use. For the first time, we are able to analyse what is explicitly described as registered uses in orality. The form *o senhor* is much more frequent in orality than in writing, at least judging from the *Corpus do Português* (Davies & Ferreira 2016–). In these cases, it is almost always used as a form of address, and it is therefore understandable that it exists in oral dialogical interactions and in the fictional dialogues that seek to reproduce them.

In this diachronic reading of the occurrences of *o senhor*, a degree of semantic bleaching is noticeable. As from the 19th century, not only does *o senhor* reveal semantic features considered exclusive to nouns, but it is also used in allocution, as a way to address male addressees with whom the speaker does not have a close relationship. The restriction of FA use only to addressing members of the nobility disappears and the appellative becomes more common, directed at a wider range of addressees, while maintaining a dimension of respect that comes from its initial use.

Interpreting the change in address forms, Biderman (1972–1973: 370) considers that *o senhor* fills a void in the former system occupied by the pronoun *vós* (you-2SG.DEFERENTIAL). She refers to a tripartite system of pronoun forms: *tu – você*¹³ – *o senhor* (Biderman 1972–1973: 373). This position in the address system causes *o senhor* to be used as a pronominalized form, without the possibility of different pronominal choice.

¹¹ *Singularidades de uma rapariga loira* by Eça de Queirós [Corpus do Português]

¹² *O Crime do Padre Amaro* by Eça de Queirós [Corpus do Português]

¹³ See Nascimento et al. (2018).

Despite these changes, the semantic-pragmatic values of the lexeme of positive appreciation remains:

- (16) A começar pelo princípio (...) e a acabar no fim (...), Kevin é *um senhor*.
 ‘Starting at the beginning (...) and finishing at the end (...), Kevin is a gentleman.’ [CETEMPúblico]

This positive evaluation is also a part of the new meanings of *o senhor* as ‘adult male person.’

- (17) O meu pai conhece *um senhor* que deita fogo de artifício.
 ‘My father knows a gentleman who sets fireworks.’ [PSFB interview 4]

If *conhece um senhor* ‘knows a gentleman’ were substituted by *conhece um homem* ‘knows a man’, this would imply a decrease in the positive valuation of the object of discourse, even though it is perfectly acceptable to say ‘knows a man’. The FA *o senhor* shares this positive value associated with deference and respect, as we shall see.

4.2 Contexts of the occurrence of *o senhor* in contemporary European Portuguese: A synchronic perspective

The FA *o senhor* is a challenge to the classical conception of a watertight categorization. As a lexical item with semantic content, this form is also addressed to an adult male, known or unknown. As the only form (*o senhor*) or occasionally followed by other nominal forms of address (*o senhor + Presidente*), *o senhor* occurs in contexts that, according to the syntactic criterion adopted (to distinguish nominal forms of address from pronominal forms of address), are specific to the pronominal forms of address.

As this FA can occur with an allocutive or delocutive function, the ambiguity this may create is resolved by the linguistic or situational context, as in example (6). In fact, this usage is only apparently delocutive. *O senhor engenheiro José Sócrates* ‘Mr Engineer José Sócrates’ is the locutor’s addressee. The ambiguity may be reinforced by linguistic mechanisms, such as repeating the 3rd person singular pronoun, and non-verbal mechanisms, namely, by eye contact. If politicians look directly at the moderator, a reorganization of relationships among all the addressees takes place. Kerbrat-Orecchioni (2010: 109) emphasizes the importance of non-verbal mechanisms of address to identify the addressee. Goffman (1981: 133) also defines addressee as “(...) the one to whom the speaker addresses his visual attention”. José Sócrates in (6) may be shown as a secondary addressee by

this gesture but is nevertheless the main target of the illocutionary act of criticism (Goffman 1981, Maury-Rouan 2005, Rossano 2013, Constantin de Chanay & Kerbrat-Orecchioni 2017).

4.2.1 Occurrences of *o senhor* in sociolinguistic interviews in European Portuguese

In interpersonal relationships, the choice of *o senhor* underlines and simultaneously constructs a formal relationship of respect and deference, in contexts where the form *você* is assessed by the speakers as inappropriate. In the corpora consulted, the sociolinguistic interviews are particularly productive in terms of this type of occurrence. As we mentioned in the methodological framework, we have used the *Perfil Sociolinguístico da Fala Bracarense* corpus. It is a stratified sample, according to age, gender, and education. The interviewers (E) are young university students. The interviewees or informants (I) are organized into four age brackets (15–25; 26–59, 60–75 and +75). In the forms of address adopted, the intragenerational, interpersonal relationship determined the use of the informal second person address form (*tu*) (18). Furthermore, because they may also belong to different generations, an intergenerational relationship determined the use of 3rd person forms, with variations between use of the verbal form (3SG) and the use of *o senhor* (19). *O senhor* is the most frequent form, sometimes the only one, along with the occurrence of the 3rd person verbal form. The interactional relationship that is established is one of reciprocity of address forms combined with proximity (18) or social distance (19), as in the following examples:

- (18) E: *Tu se pudesses viver noutro sítio...*
E: PRON.2SG if can-SUBJ.IMP.2SG INF anywhere else
I: *Ao fim, sempre em frente, já vês a minha escola.*
I: At the end, straight ahead, see-PRS.2SG my school
'E: If you could live anywhere else...'
'I: At the end, straight ahead, you can see my school' [PSFB interview 1]
- (19) E: *O senhor gosta de viver aqui?*
E: ART.DEF.M.SG mister like-PRS.3SG PREP INF here?
I: *Desculpe, mas perdi o fio à meada, da pergunta que*
I: Sorry, I've lost my train of thought, regarding the question
a menina fez.
ART.DEF.F.SG Miss asked.
'E: Do you like living here?'
'I: I'm sorry, I've lost my train of thought, regarding the question you asked' [PSFB interview 31]

Given the characteristics of the interview genre (and specifically, of the sociolinguistic interview genre, which is aimed at getting the informants to talk about their personal lives, experiences and opinions), we find the forms of address mainly in the interviewer's interventions.

The results of the analysis of the interview data in terms of absolute occurrences, presented in Tables 4 and 5, corroborate the interpersonal relationship profile presented with regard to the FA used, and show how *tu* (pronoun and/or verb form) and *o senhor* (and/or verb form) are in complementary distribution:

Table 4: Forms of address of the 2nd person singular in sociolinguistic interviews

FA Interviews	<i>tu</i> + verb in 2SG	only verb in 2SG	total
1	67	70	137
3	65	88	153
5	36	238	274
7	15	31	46
Total occurrences	183	427	610

Table 5: Forms of address of the 3rd person singular in sociolinguistic interviews

FA Interviews	<i>o senhor</i> + verb in 3SG	(<i>o</i>) <i>senhor</i> name + verb in 3SG	only verb in 3SG	total
12	46	2	57	125
22	37	6	106	149
31	16	0	6	22
43	0	14	33	47
Total occurrences	99	22	202	333

Given the features of the discourse genre, the findings show that it is the interviewers who mostly use the forms of address as part of question acts. In some interviews, the interviewees hardly use any form of address, or indeed none at all.

In Table 4, the FA *tu* is the only FA used together with the 2SG verbal form, corresponding to 30% of all verbal forms of address used in the interviews (*tu* + verb in 2SG and only verb in 2SG). It is important to stress the Pro-Drop nature of European Portuguese in order to understand those occurrences of the verbal form. It should be also noted that interview 5 stands out for the number of occurrences of second person verbal forms. The interviewee gives short answers, which leads to the occurrence of more than three hundred question acts in the course of the 60 minutes dedicated to each interview. This points out the interviewer's need to provoke the informant to get him to talk.

In Table 5, the VFA (3SG) is still prevalent (202 occurrences), and *o senhor* is the most used FA (99 occurrences against 22 occurrences of *(o) senhor* + F-L name). The divergence of occurrences that stands out in interview 43 stems from the social prestige of the interviewee (parish priest). The interviewer prefers the structure *(o) senhor abade* 'mister parish priest' to *(o) senhor* as a more deferential FA. The higher prevalence of only verbal forms (verbal address, i.e., 3SG without an expressed subject) in this interview may also be due to the fact that the use of FA is not exclusively determined by linguistic rules, but also involves idiosyncratic features.

4.2.2 Occurrences of *o senhor* in political interviews and debates on Portuguese television

The debates and interviews in the corpus we compiled were collected according to the established criteria, namely, occurring in the 21st century during electoral campaigns with male political participants, given the objective of analysing the occurrences and characteristics of the form of address *o senhor* in these discourse genres. Three debates between leaders of the two main parties were selected, the first in 2005, between José Sócrates (JS), leader of the Socialist Party (PS), and Pedro Santana Lopes (PSL), leader of the Social Democratic Party (PSD), lasting about 90 minutes; the second in 2011, between José Sócrates, leader of PS, and Pedro Passos Coelho (PPC), leader of PSD, lasting approximately 45 minutes. The third debate was held in 2015 between Pedro Passos Coelho, leader of PSD, and António Costa (AC), leader of PS, lasting 60 minutes.

Three interviews were also selected, two in 2005, one with Pedro Santana Lopes, leader of PSD, lasting about 41 minutes, conducted by the journalists Paulo Magalhães and Manuel Carvalho, and the other with Jerónimo de Sousa, leader

of the Portuguese Communist Party (PCP), lasting approximately 40 minutes and conducted by the journalists Raquel Abecassis and Eduardo Dâmaso. The third interview was held in 2019, with Rui Rio, leader of PSD, and the journalists Carlos Daniel and António Esteves, lasting about 30 minutes.

In the debates analysed, the frequency of *o senhor* and other forms of address is related to the degree of interaction in each debate. In debates with high interactivity like these ones, the politician's question and confront each other and often disagree. Interruptions, overlapping turns, and the moderator's difficulty in controlling the course of the debate are signs of this high level of interaction. In these contexts of confrontation and combative aggressiveness, features of colloquialism are frequent. In these highly interactive debates and as we can see in Table 6, the frequency of FA occurrences with (*o*) *senhor* should be taken into account in the analysis of genre features and the interpersonal relationship built. They mark a formal relationship of respect, with varying degrees of distancing, building a polite relationship, despite the pragmatic dimension of confrontation that runs through the discourse genre.

Table 6: Occurrences of FA with *senhor* in political debates

	2005	2011	2015	Total
	JS – PSL	JS – PPC	PPC – AC	occurrences
<i>o senhor</i>	21	98	51	170
<i>o senhor</i> (Title / F-L Name)	2	31	38	71
<i>os senhores</i>	8	1	16	25
<i>senhor</i> (Title / F-L Name)	2	61	77	140
<i>meus senhores</i>	2	0	0	2
Total occurrences	35	181	182	408

The address form *o senhor* is the most frequent, with 170 occurrences of the total 408 address forms in which this term is present. Moreover, delocutive forms with allocutive value predominate (241 occurrences), although forms with a vocative function seem to be growing (from 2 occurrences in 2005 to 77 in 2015). *Meus senhores* 'my [possessive] gentlemen' is residual, because even though the term may be used by the moderators, it seems to have fallen into disuse. We believe this abandonment is related to possible changes in the forms of address used by journalists, who seem to opt more frequently for nominal forms of address, for example, first and last names. *Os senhores* ('gentlemen') is also a form of address that is rarely used. The fact that these are debates between party leaders directs

the allocution to the individual rather than to the group to which they belong. When the attack is directed at the political group, the participants in the debate choose to mention the name of the party alongside *os senhores*. The contestation of the adversary through acts of criticism and accusation are preferential contexts of occurrence of the forms *o senhor* and *senhor* + NFA. The numbers found suggest a significant relationship between verbal aggressiveness and the occurrence of these forms (example 6).¹⁴

It should be noted that we consider the genre to be of central importance for the study of these topics, and that the defining traits of electoral political debates should not by any means be confused with electoral interviews. However, the latter reveal interesting similarities with debates, considering that the journalist(s) generally ask the interviewees controversial and difficult questions whose answers are similar to those they would give a political opponent. Moreover, the journalists assume they are the spokespeople of the Portuguese people, and more than just question, they may actually confront their interviewees, coming closer to the relationship of conflict inherent to a debate.

In this corpus of interviews, there are numerous occurrences of *o senhor* (20–23). This seems to be the interviewers' preferred FA to address electoral candidates, besides the use of their first/last name (20), which may combine with *Doutor* (Doctor) (23). Together with these occurrences, the forms *sotor* (21) and *o sotor* (22), contracted forms of *senhor doutor/o senhor doutor* ((the) Mr Doctor), reveal a pattern of occurrence to whose analysis we will return:

- (20) Jerónimo de Sousa, *o senhor* disse que o euro
Jerónimo de Sousa, ART.DEF.M.SG mister say-PRT.3SG that the euro
não trouxe o crescimento prometido...
has not brought the promised growth...
'Jerónimo de Sousa, you [the Mr] said that the euro has not brought the
promised growth...' [interview, 2005b]
- (21) Boa noite *sotor*, *o senhor*
Good evening, Mr doctor [contracted form], ART.DEF.M.SG mister
desde há duas semanas que anda a dizer...
have been saying that for two weeks now...
'Good evening, sir, you have been saying that for two weeks now...'
[interview, 2005]

¹⁴In contrast, in debates with low interaction, politicians assign the moderator the role of direct addressee, marked verbally and non-verbally. The direction of their gaze is a key indicator. The debate assumes a question-answer structure, with fewer interruptions, less overlapping, and easier turn alternations. Allocution marks are infrequent.

- (22) Ou seja, *o sotor*
 In other words, ART.DEF.M.SG Mr doctor [contracted form]
 não se sente preso por este acordo com com com o PP...
 do not feel bound to this agreement with with PP...
 ‘In other words, you [the Mr doctor, contracted form] do not feel bound
 to this agreement with with PP...’ [interview, 17/02/2005]
- (23) *Doutor Rui Rio*, o jornal Expresso revelou hoje que há
 Doctor Rui Rio, the Expresso newspaper revealed today that there is
 uma conspiração do Ministério Público com envolvimento ...
 a conspiracy in the Public Prosecutor’s Ministry with the involvement ...
O senhor acha crível uma tese com estas
 ART.DEF.M.SG Mr think credible a theory with these
 características?
 characteristics?
 ‘Mr Rui Rio, the Expresso newspaper revealed today that there is a
 conspiracy in the Public Prosecutor’s Ministry with the involvement [...].
 Do you [the Mr] think a theory with these characteristics is credible?’
 [interview, 2019]

The FA *o senhor* occurs frequently as an anaphoric resumption of a NFA with a vocative function as in (20), (21) and (23). It occurs in these cases with the function of pronominal deixis, preceding the verb as syntactic subject. Given that Portuguese is a null-subject language, the speaker could have opted for *Jerónimo de Sousa, disse que o euro não trouxe o crescimento prometido* ‘Jerónimo de Sousa, [you] said that the euro didn’t bring the promised growth’. However, there is a change at the pragmatic level, which is fundamental. The occurrence of *o senhor* stresses a relationship of politeness between the speakers. There is in fact a clear difference in the degree of politeness between *o senhor + V* and the exclusive use of the verb, which is less empathetic and aloof, the “zero degree of politeness” that Carreira (1997) refers to.

These data show strong idiosyncratic variability which is typical of the discourse genre but maintaining always a minimal relationship of respect, to which the use of the forms of address and *o senhor*, in particular, contribute. The political interview genre (as well as the electoral debate) determines a formal relationship between the speakers, but with remarkable variability. Indeed, there are idiosyncratic traits that mark the speech of the journalists, as they systematically opt for certain variations of this form. Table 7 summarizes the occurrences of address forms in the three interviews considered.

Table 7: FA occurrences in political interviews

	Journalists	Interviewees	Total
<i>o senhor</i>	35	0	35
<i>sotor/o sotor</i>	16 / 15	0	31
<i>os senhores</i>	2	3	5
<i>doutor</i> + N + last name	3	0	3
first name + last name	11	0	11
Total	82	4	85

Some conclusions can be drawn from the figures obtained: it is mostly journalists who address the politicians using the FA, which is common in this journalistic genre. They have to take the initiative to ask the questions, which is why they address the interviewee using the FA. The most frequent forms are *o senhor* (35 occurrences) and *sotor/o sotor* (with 31 occurrences). The fact that *sotor/o sotor* are so frequent in these records suggests the standardization or conventionalization of this form. The nominal address form first/last name appears in third place in the number of occurrences, but far below the others.

In conclusion, we can say that in 50 of the 82 occurrences involving the journalists, *o senhor* and *sotor* are found before a verb as a syntactic subject. These findings can be related to the results highlighted in Allen (2019), about the growth in productivity of these forms by the end of the 20th century.

From the analysis of the occurrences in the different corpora, we further conclude that the NP *senhor* does not occur as a vocative, contrary to medieval uses. We are aware, however, of its use in a religious context, addressed to God, and also in children’s speech and in popular registers, to call the attention of an unknown adult. In these last two cases, it is usually accompanied by the particle *ó* ‘hey’ (as in *hey mister*).

4.3 *O senhor* as the only form of address

As a form of address, *o senhor* is used alone in the utterance, marking a systematic relationship of respect with the addressee. It is frequent in interviews, whether political or sociolinguistic interviews. In the PSFB corpus, as in examples (24) and (25), *o senhor* is the most frequent form of address used by the interviewers, who are young women, to address the interviewees, who are male and from an older generation, regardless of their social status:

- (24) E *o senhor* não sabe?
 And ART.DEF.M.SG mister NEG know-PRS.3SG?
 ‘And don’t you know?’ [PSFB interview 25]
- (25) E, por exemplo, acha que os seus filhos estão a educar os seus
 And, for example, PRT.3SG that your children are raising your
 netos da mesma maneira que *o senhor* os
 grandchildren the same way that ART.DEF.M.SG Mr raised
 educou?
 them?
 ‘And, for example, do you think that your children are raising your
 grandchildren the same way that you raised them?’ [PSFB interview 22]

The prevalence of this form of address in the interactions analysed, regardless of the social group to which the addressee belongs, seems to point to a more generalized use of *o senhor*. This may in turn lead to the banalization of its use as it becomes more automatized, consequently decreasing the prominence of the semantic-pragmatic feature of deference found in the *Corpus do Português* (Davies & Ferreira 2016–). *O senhor* thus seems to move into the semantic-pragmatic area of the form *você*. This is a shift that signals a degree of instability and plasticity of the FA, which is reflected in uses like this one:

- (26) Mas *o senhor* acha importante, por exemplo,
 But ART.DEF.M.SG mister find-PRS.3SG important, for example,
 acha importante *vocês* irem à missa?
 find-PRT.3SG important PPR.2PL go-INF.3PL to mass?
 ‘But do you [+deference] think it is important, for example, do you think
 it is important that you [-deference] all go to mass....?’ [PSFB interview 22]

The forms of address *o senhor* and *vocês* (see Duarte & Marques, accepted) participate in the construction of an anaphoric chain that brings together the interpersonal values of respect in the two FA. These are scalar uses of *o senhor*. There is a difference in the pragmatic values of respect and deference between the use of *o senhor* and *o senhor* + NFA (see Hummel & dos Santos Lopes 2020) on the traits of respect and deference). Not only titles, but also proper name and family name convey deference to varying degrees. Using *senhor* followed by the first name, last name or full name (*sr. Joaquim, senhor Silva, senhor Joaquim Silva*) is a mark of respect and establishes a growing degree of deference. In the gradation established, the example below (27) illustrates a form of respect, but not of deference. An addressee whom the speaker addresses with *senhor* + *first*

name is not in a high interpersonal position relative to the speaker. *O senhor* marks a relationship of respect, determined by a generational criterion, but not of deference.

(27) Bem, *senhor Vicente*, ficamos por aqui.

‘Well, Mr. Vicente, I think we can end here.’ [PSFB interview 12]

One of the contexts for the occurrence of *o senhor* is as an anaphoric resumption of an immediately preceding NFA (28). It is a fundamental usage to determine the semantic-pragmatic adaptability of *o senhor* as a scalar form of deferential address. As a case of anaphoric retaking by coreference, the interpersonal relationship of deference created by the NFA remains unchanged.

(28) *Senhor Engenheiro José Sócrates, o senhor insiste na*
Mr. Engineer José Sócrates, ART.DEF.M.SG Mr PRT.3SG on
co-incineração.
co-incineration.

‘Mr. Engineer José Sócrates, you insist on co-incineration.’ [debate, 2005]

In pragmatic terms, recourse to the nominal address form is a discourse strategy to ‘recognize’ the others, assigning them a specific social role in the interaction, which the form *o senhor* does not do. The derogatory irony of the FA used with critical intention in the example below (29) derives from the mismatch between the chosen form of address and the social status of the public figures mentioned.

(29) Assim, já poderia marcar mais um almoço, com *o senhor Alegre*; um lanche, com *o senhor Machete* e mais um jantar, com *o senhor Monjardino!*

‘Thus, you could schedule another lunch with Mr. Alegre; tea with Mr. Machete, and another dinner with Mr. Monjardino!’ [CETEMPúblico]

These public figures are usually referred to as *Manuel Alegre/(senhor) doutor Manuel Alegre, Rui Machete/(senhor) doutor Rui Machete, Carlos Monjardino/(senhor) doutor Carlos Monjardino*, but never referred to as *Senhor Alegre, Senhor Machete* and *Senhor Monjardino*. Thus, in this example, there is a downgrading of the referents’ image reducing them to the status almost of regular people.

4.4 Phonetic contraction of the address form (*o*) *senhor*

The phonetic phenomenon of erosion is frequent in oral language uses, especially in more informal contexts. The NP *o senhor*, in the subject position of V and performing an allocutive function, seems to be realized shorter in duration¹⁵ and it even appears reduced to the forms *seor* or *sor*. The examples (30) and (31) illustrate a very frequent usage in the corpus analysed:

- (30) Rui Rio, *aceitaria* a leitura que *o*
 Rui Rio, accept-COND.3SG the reading that ART.DEF.M.SG
seor foi um melhor líder nas
 Mr [eroded form] be-PRT.3SG a better leader in the
 duas últimas semanas que nos últimos dois anos?
 last two weeks than in the last two years?
 ‘Rui Rio, (...). Would you accept the reading that you have been a better leader in the last two weeks than in the last two years?’ (interview, 2019)
- (31) *O sor* tem falado muito do record da carga
 ART.DEF.M.SG Mr [eroded form] has spoken a lot about the record tax
 fiscal...
 burden...
 ‘You [the Mr, contracted form] have talked a lot about the record tax burden...’ (interview, 2019)

There are other reduced forms of *senhor* that are equally documented in the PSFB, like the form *se* in *se Joaquim*.¹⁶ This form can be used with both males (32) and females (33), as in *se Manel*, *se Maria*, preceding the proper name in an address form that is typical of popular varieties, with a clear generational dimension of politeness, being normally used for older addressees:

- (32) *O meu falecido pai andava pelas ruas vem aí o*
 My late father walked the streets come-PRS.3SG ADV ART.DEF.M.SG
se Machadinho cos jornais
 Mr [eroded form] Last Name [diminutive form] with the newspapers
 e tal
 and such”
 ‘My late father walked the streets “Here comes Mr Machadinho with the newspapers and such’ [PSFB, interview 22]

¹⁵This is our native speakers’ perception, as we do not measure the duration of the elocution. However, we consider this situation is similar to Posio’s (2018) finding regarding the duration of *a/uma pessoa* in grammaticalized vs. non-grammaticalized uses.

¹⁶A variant of this form is *sô*, also as a mark of informality. It has occurred in literary texts since the 19th century, as documented in the Corpus do Português, which records 65 examples.

There are a diversity of functions and syntactic positions occupied by the forms of address *senhor* and *o senhor*. *Senhor* occurs in the vocative position, combined with other nominal forms of address, organized on a gradation according to the features of [±] formality and [±] deference. It may also occur in this context with delocutive value, functioning as a full word. In the allocutive function, there is some fixation of the structure, as it only has this function if it occurs with the definite, male, singular article characteristic of the nominal address form category. Other categories of determiners like *um senhor*, *este senhor*, *aquele senhor*, *certo senhor* ‘a gentleman, this gentleman, that gentleman, a certain gentleman’, etc. always have delocutive uses. In syntactic terms, *o senhor* performs the function of subject or complement, like the personal pronouns. It is integrated into the sentence structure. It is also in this context that it occurs as the only form of address and may accumulate an anaphoric function of linking to a previous nominal form of address (see (20)). In this case, the degree of deference varies from context to context, depending on the NFA, not on the form of address *o senhor*.

The data we analysed also point to different uses and frequencies of occurrence, according to the discourse genres and the idiosyncrasies of the speakers. But there are also dimensions of change regarding the semantic-pragmatic characteristics of this FA, in connection with a semantic axis from deference to respect, originating from its lexical content as a full word. As a single FA, frequent especially from the last century onwards, the NP *o senhor* is experiencing a process of semantic bleaching, conveying a general relational value of respect. Therefore, it occurs in situations of varying formality. It marks a social relationship of distance with regard to the addressee, identified as a ‘male, adult interlocutor’. These syntactic, semantic and pragmatic particularities are accompanied by a process of phonetic erosion (see Heine & Kuteva 2004: 3),¹⁹ which gave rise to the eroded forms, *se*, *sô*, *sor*, *seor*. The contracted *sotor/stor* (*senhor* + *doutor*) is one of the most widespread forms of these eroded forms, with uses that are signs of the word’s integration into the Portuguese lexicon, occurring particularly in written contexts. In the data collected in the *Corpus do Português* (Davies & Ferreira 2016–) for current use, only the forms *sotor* and *stor* occur with 8 and 14 occurrences, respectively. They are also the only eroded forms that have been introduced into dictionaries.²⁰

¹⁹The phenomenon is very similar to what has happened to the Portuguese address pronoun *você*, where there has been a change/reduction of form that accompanies semantic change and the content.

²⁰<https://www.infopedia.pt/dicionarios/lingua-portuguesa/sotor> and other online dictionaries.

According to the four criteria established by Heine & Kuteva (2004: 17), the uses we identified in the data analysed suggest that there is an ongoing process of grammaticalization.²¹ The analysis carried out reveals processes of semantic bleaching, use in new contexts, syntactic fixation, recategorization (approximation to the pronoun category) and phonetic erosion. These characteristics of (*o*) *senhor* are related to each other, as Heine & Kuteva (2004) remind us. The semantic change highlights a more grammatical sense, although in *o senhor* as a deictic element, some part of the semantic value of the NP is maintained. In pragmatic terms, *o senhor* marks social distancing (social deixis) but retains NFA marks, like the combination with the 3rd person (Carreira 2009) and the occurrence with the definite article, a characteristic of NFA in this context.

5 Final considerations

This paper has focused on the different discourse contexts in which the form of address *o senhor* appears in EP. We consider that there is a generalization of the uses of *o senhor* which defines a respectful form of address, regardless of whether the relationship between the speakers is asymmetrical or symmetrical.

Having analysed different data, organized according to different discourse genres, we found that the discourse genre interferes in the speakers' choices of the FA, but further research is required to confirm these findings. The variability in usage that we identified, however, forces us to consider that other dimensions of verbal interaction interfere in the speakers' choices, thus, an idiosyncratic dimension should be considered in the analysis. Finally, the nominal form of address *o senhor* is in a process of grammaticalization in EP. In short, *o senhor* is a hybrid form of address, with uses that sometimes bring it closer to a NFA and sometimes to a PFA; there is a synchronic convergence of both categories' features, according to different contexts and usages. While it may be premature to speak of a stabilization of the grammatical category pronoun for *o senhor* in allocutive use, it is safe to say that there are uses with a pronoun function, a deictic function, even though the semantic bleaching is not finished. We thus underline the instability of these usages, some more grammaticalized than others, as in examples (33), (20–21) and (27).

Taking into account our preliminary expectations, the results of the present analysis show that *o senhor* is a widespread form of address marking a relationship of social distance in contemporary Portuguese. Finally, we have identified

²¹We use the term grammaticalization in a broad sense, encompassing processes also called pragmaticalization or pragmatization and discursivization.

a set of characteristics that allow us to state that as a nominal form of address, *o senhor* is in a process of grammaticalization in EP.

Some avenues for future research have become evident in the course of the analysis. We have established that the variation *o senhor/a senhora* deserves further investigation, as it is not limited to a mere morphosyntactic variation. The form (*a*) *senhora* is used in specific contexts deserving more research work. Another topic that deserves future attention is the VFA category. The use of the 3rd person singular of the verb without a subject, much more frequent in the data analysed than the NFA or PFA, is in line with the fact that EP is a null-subject language. But we also consider that it has discursive implications that have yet to be determined. As mentioned above, the duration of elocution, in previously published works, is associated with the process of grammaticalization. The analysis of the behaviour of *o senhor* in this respect may bring more data to the current discussion. Finally, the functioning of this FA in different discourse genres, not only at the oral level, but also in certain written genres, like the epistolary one, deserves further analysis. Specifically, in future research, we intend to analyse the crystallized expression *sim senhor/sim senhora* ('yes sir'/'yes ma'am') which also presents unique features and functions that may bring new information on the process of grammaticalization.

References

- Allen, Ana Sofia Ferreira. 2019. *O sistema de formas de tratamento em português europeu*. Lisbon. (Doctoral dissertation).
- Biderman, Maria Teresa. 1972–1973. Formas de tratamento e estruturas sociais. *Alfa* 18/19. 339–382.
- Brown, Robert & Albert Gilman. 1960. The pronouns of power and solidarity. In Thomas Albert Sebeok (ed.), *Style in language*, 253–276. Cambridge, MA: MIT Press.
- Bühler, Karl. 1982 [1934]. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Stuttgart/New York: Gustav Fischer.
- Carreira, Maria Helena. 1997. *Modalisation linguistique en situation d'interlocution: Proxémique verbale et modalités en portugais*. Louvain: Peeters.
- Carreira, Maria Helena. 2004. *Les formes allocutives du portugais européen: Évolutions, valeurs et fonctionnements discursifs*. Paper presented at the Coloquio de Paris of the Instituto Cervantes in Paris, France. https://web.archive.org/web/20220126214549/https://cvc.cervantes.es/lengua/coloquio_paris/ponencias/pdf/cvc_araujo.pdf.

- Carreira, Maria Helena. 2007. Le pronom d'adresse portugais «você»: Valeurs et évolution. In Alexandra Cunita, Coman Lupu & Liliane Tasmowski (eds.), *Studii de lingvistică și filologie romanică: Hommages offerts à Sanda Reinheimer Rîpeanu*, 15–19. București: Editura Universitatii din București.
- Carreira, Maria Helena. 2009. Qualification et adresse: Complexité modale et enjeux interlocutifs: L'exemple du Portugais. *Synergies Pologne* 6. 29–34.
- Castillo Lluch, Monica. 2014. Les formes nominales d'adresse en espagnol et en français: Étude contrastive de señor(a) et madame/monsieur dans deux types d'interactions institutionnelles. In Catherine Kerbrat-Orecchioni (ed.), *S'adresser à autrui: Les formes nominales d'adresse dans une perspective comparative interculturelle*, vol. 2. Chambéry: Université de Savoie.
- Cintra, Luís Filipe Lindley. 1972. *Sobre as formas de tratamento*. Lisboa: Livros Horizonte.
- Constantin de Chanay, Hugues & Catherine Kerbrat-Orecchioni. 2017. Regard et deixis personnelle: L'adresse dans les débats d'entre-deux-tours des élections présidentielles françaises. *Langue Française* 193. 93–108.
- Cook, Manuela. 1994–1995. On the Portuguese forms of address: From *vossa mercê* to *ocê*. *Portuguese Studies Review* 3(2). 78–89.
- Cunha, Celso & Luís Filipe Lindley Cintra. 1984. *Nova gramática do português contemporâneo*. Porto: Figueirinhas.
- Davies, Mark & Michael Ferreira. 2016–. *Corpus do Português: Historical genres*. <http://www.corpusdoportugues.org/hist-gen/>.
- de Paiva Sória, Maíra Vasconcelos. 2013. *Nós, a gente e o sujeito nulo de primeira pessoa do plural*. Universidade de Lisboa. (MA thesis). <http://hdl.handle.net/10451/12218>.
- Dickey, Eleanor. 1997. Forms of address and terms of reference. *Journal of Linguistics* 33(2). 255–274.
- Duarte, Isabel Margarida. 2010. Formas de tratamento: Item gramatical no ensino do Português Língua Materna. In Ana Maria Brito (ed.), *Gramática: História, teorias, aplicações*, 133–146. Porto: Centro de Linguística da Universidade do Porto.
- Duarte, Isabel Margarida. 2011. Formas de tratamento em português: Entre léxico e discurso. *Matraga* 18(28). 84–101.
- Duarte, Isabel Margarida & Maria Aldina Marques. Forthcoming. Vós and other pronominal forms of address (*tu, você, vocês*): Speakers' perceptions and differences between Brazilian and European Portuguese. In Nicole Baumgarten & Roel Vismans (eds.), *It's different with you: Contrastive perspectives on address research*. Amsterdam: John Benjamins.


- Faraco, Carlos Alberto. 1996. O tratamento você em português: Uma abordagem histórica. *Fragmenta* 13. 51–82.
- Formentelli, Maicol. 2009. Address strategies in a British academic setting. *Pragmatics* 19(2). 179–196.
- Goffman, Erving. 1981. *Forms of talk*. Oxford: Basil Blackwell.
- Hammermüller, Gunther. 2004. *Adresser ou éviter, c'est la question: Comment s'adresser à quelqu'un en portugais sans avoir recours à un pronom ou à une autre forme équivalente*. Paper presented at the Coloquio de Paris of the Instituto Cervantes in Paris, France. https://web.archive.org/web/20221007072427/https://cvc.cervantes.es/lengua/coloquio_paris/ponencias/pdf/cvc_hammermueller.pdf.
- Heine, Bernd & Tania Kuteva. 2004. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Heine, Bernd & Kyung-An Song. 2011. On the grammaticalization of personal pronouns. *Linguistics* 47(3). 587–630. DOI: 10.1017/S0022226711000016.
- Hummel, Martin & Célia dos Santos Lopes. 2020. *Address in Portuguese and Spanish: Studies in diachrony and diachronic reconstruction*. Berlin: De Gruyter. DOI: 10.1515/0123497831107.
- Ilari, Rudolfo, Carlos Franchi & Maria Helena Moura Neves. 1996. Os pronomes pessoais do português falado: Roteiro para a análise. In Ataliba de Castilho & Margarida Basílio (eds.), *Gramática do português Falado*, vol. 4. Campinas: Editora da Unicamp.
- Johnen, Thomas. 2014. Les FNA en français et en portugais: Considérations théoriques et analyses fonctionnelles dans des débats médiatiques électoraux au Brésil, au Portugal et en France. In Catherine Kerbrat-Orecchioni (ed.), *S'adresser à autrui: Les formes nominales d'adresse dans une perspective comparative interculturelle*, vol. 2, 375–416. Chambéry: Université de Savoie.
- Kerbrat-Orecchioni, Catherine. 1992. *Les interactions verbales*, vol. 2. Paris: A. Colin.
- Kerbrat-Orecchioni, Catherine. 2005. Le discours en interaction. In Paris: Armand Colin.
- Kerbrat-Orecchioni, Catherine (ed.). 2010. *S'adresser à autrui: Les formes nominales d'adresse en français*. Chambéry: Université de Savoie.
- Kerbrat-Orecchioni, Catherine (ed.). 2014. *S'adresser à autrui: Les formes nominales d'adresse dans une perspective comparative interculturelle*, vol. 2. Chambéry: Université de Savoie.
- Lehmann, Christian. 2015. *Thoughts on grammaticalization*. 3rd edn. Berlin: Language Science Press. DOI: 10.17169/langsci.b88.99.

- Marques, Maria Aldina. 2014. Cortesia, formas de tratamento e géneros discursivos: Condições de ocorrência e de uso. In Isabel Seara (ed.), *Cortesia: Olhares e reinvenções*, 145–172. Lisboa: Chiado Editora.
- Mauray-Rouan, Claire. 2005. Le flou des marques du discours est-il un inconvénient? Vers la notion de “leurre discursif”. *Marges Linguistiques* 2005(2). 163–174.
- Medeiros, Sandi Michelle de Oliveira. 1985. *A model of address form negotiation: A sociolinguistic study of Continental Portuguese*. Ann Arbor, MI: University of Texas at Austin. (Doctoral dissertation).
- Medeiros, Sandi Michelle de Oliveira. 1992. Um modelo psico-sociolinguístico de formas de tratamento. In *Atas do VIII encontro da Associação Portuguesa de Linguística*, 330–342. Lisboa: Associação Portuguesa de Linguística. <https://web.archive.org/web/20220617013445/https://apl.pt/wp-content/uploads/2017/12/1992-28.pdf>.
- Nascimento, Fernanda. 1987. A gente, um pronome da 4.^a pessoa. In *Actas do congresso sobre a investigação e ensino do português*, 480–490. Lisbon: Diálogo e Compilação.
- Nascimento, Fernanda, Amália Mendes & Maria Eugênia Duarte. 2018. Sobre formas de tratamento no português europeu e brasileiro. *Diadorim* 20. 245–262.
- Oliveira, Sandi Michelle. 2004. Para além de poder e solidariedade: Uma retrospectiva sobre formas de tratamento em Portugal (1982–2002). *Franco-British Studies* 33–34. 126–136.
- Pereira, Sandra Maria. 2003. *Gramática comparada de a gente: Variação no Português Europeu*. Universidade de Lisboa. (MA thesis). https://www.clul.ulisboa.pt/files/sandrabrito_pereira/Pereira_2003_tese.pdf.
- Posio, Pekka. 2018. Properties of pronominal subjects. In Kimberly L. Geeslin (ed.), *The Cambridge handbook of Spanish linguistics*, 286–306. Cambridge: Cambridge University Press. DOI: 10.1017/9781316779194.014.
- Posio, Pekka. 2021. *A pessoa* and *uma pessoa*: Grammaticalization and functions of a human impersonal referential device in European Portuguese. *Journal of Portuguese Linguistics* 20(1). DOI: 10.5334/jpl.254.
- Pountain, Christopher. 2003. Pragmatic and structural reflections on the expression of the second person notion in Romance with special reference to Spanish and Portuguese. *Bulletin of Spanish Studies* 80(2). 145–160. DOI: 10.1080/14753820302029.
- Preti, Dino. 2004. Papéis sociais e formas de tratamento em *A ilustre casa dos Ramires*, de Eça de Queiroz. In Dino Preti (ed.), *Estudos de língua oral e escrita*, 180–199. Rio de Janeiro: Lucerna.

- Raposo, Eduardo. 2013. Pronomes. In Eduardo Paiva Raposo, Maria Fernanda Nascimento, Maria Antónia Mota, Luísa Segura & Amália Mendes (eds.), *Gramática do português*, vol. 1, 883–918. Lisboa: Fundação Calouste Gulbenkian.
- Rossano, Federico. 2013. Gaze in conversation. In Jack Sidnell & Tanya Stivers (eds.), *The handbook of conversation analysis*, 308–329. Oxford: Wiley-Blackwell.
- Scherre, Marta, Edilene Dias, Carolina Andrade & Carolina Martins. 2015. Variação dos pronomes tu e você. In Marco Antônio Martins & Jussara Abraçado (eds.), *Mapeamento sociolinguístico do português brasileiro*, 133–172. São Paulo: Contexto.
- Traugott, Elizabeth & Bernd Heine (eds.). 1991. *Approaches to grammaticalization*, vol. 2: Types of grammatical markers. Amsterdam: John Benjamins.

Chapter 11

Variation and change in reference to discourse participants in Catalan parliamentary debate (1932–1938 and 1980–2020)

Neus Nogué-Serrano 

Universitat de Barcelona

Lluís Payrató 

Universitat de Barcelona

This chapter summarises the results of an ongoing study on the discourse strategies used by speakers to refer to the participants in political debates in the Parliament of Catalonia during two time periods: the period spanning 1932 and 1938, under the Spanish Second Republic, and from 1980, the year of the recovery of Catalonia's democratic institutions, until 2020. The study's theoretical background is the seminal work on deixis and politeness carried out by Levinson (1983) and followed by others, together with the foundational studies on participation frameworks by Goffman (1981), and the research into these subjects with reference to Catalan and other languages. Data from a corpus including the transcriptions of debates taken from the *Diari de Sessions* of the Parliament of Catalonia (more than 600,000 words) are classified and analysed.

Applying statistical reliability tests, the analysis combines qualitative and quantitative methods and highlights several trends in the use of different forms and strategies. The incorporation of Goffman's (1981) participation frameworks provides data that give a much more accurate answer to the questions set out than the analysis of just person deictic forms and honorifics. This is true not only of the study of parliamentary debates, but also of any other speech event involving more than two participants.



1 Introduction

Catalan is a western Romance language, situated between the Gallo-Romance languages (mainly, French and Occitan) and the Ibero-Romance languages (Portuguese, Galician, Spanish, Asturian, and Aragonese). It is spoken in an area including Catalonia, Valencia, Andorra, the Balearic Islands, Northern Catalonia, the eastern strip of Aragon, Carche, and Alghero, in Sardinia. It has more than 9 million speakers.

Parliamentary debate, a subgenre of parliamentary discourse, takes place in a political institutional setting, the Parliament. Members of parliament take part in the event as addressees and in some cases as addressers, and take on a variety of roles: politicians, representatives of a party or coalition, presidents either of the Parliament or of the Government, ministers, spokespersons, and so on. Interaction is highly ritualised and in Catalonia all interventions are closely moderated by the President of the Parliament, who acts as the Speaker and opens and closes the debate (see Cuenca 2014 and Ilie 2015 for a more detailed description).

In the Parliament of Catalonia, the discursive style of the debates from the 1932–1938 period, under the Spanish Second Republic, differs greatly from the style used now or in the recent past. The impression is that the language used now is less formal. But is this impression actually borne out by the facts? And what are the linguistic features that convey it?

This chapter seeks to provide an answer to these questions, focusing on a specific aspect of parliamentary debate: reference to participants.

1.1 Theoretical background

To address the questions just posed, we have combined concepts and categories drawn from different disciplines and theoretical orientations.

The first is person deixis. We adopt the framework established by Levinson (1983), and especially its adaptation in the studies of Catalan in recent decades (Payrató 2002, Cuenca 2004, 2014, Nogué 2005, 2008a,b, 2011, 2015, De Cock & Nogué 2017).

The second is politeness, and more specifically the studies into the Catalan three-degree system of honorifics: *tu-vosaltres* (2SG and 2PL, informal) – *vós* (2PL, respectful) – *vostè(s)* (3SG and 3PL, formal). The traditional use of this system was described by Coromines (1971) and more recently in the new Catalan normative grammar (GIEC 2016). The honorific form (*la*) *Vostra Senyoria*/(*sa*) *senyoria*, specific to the speech event, has also been included.

The third concept is the notion of participant and the different categories into which it can be broken down. Drawing on previous work by Bühler (1934), Jakobson (1960), and Hymes (1974), here we adopt the framework proposed by Goffman (1981). In the production of an utterance, or *production format*, Goffman distinguishes between the *animator* (the person who uses his/her voice, or hands, to produce the linguistic sounds, or letters or characters, that constitute an utterance), the *author* (the person who linguistically encodes an utterance, the one who selects the words and builds the sentences that verbalise what is meant), and the *principal* (the person or party held responsible for the message). First person always encodes reference to the principal, be it in cases where a participant only adopts this component or in cases where s/he adopts two components, or when s/he adopts all three components, which is the most frequent case.

In the reception of an utterance, or *reception format*, Goffman distinguishes between *ratified participants* (accepted in the communicative event) and *bystanders* (not accepted). Ratified participants, in turn, are split into *addressed recipients* and *unaddressed recipients*; and bystanders are split into *overhearers* (perceived) and *eavesdroppers* (not perceived). The grammatical category of second person – SG or PL – only encodes the reference to the addressed recipient(s) – 2PL, together or not with non-participants –, and the reference to unaddressed recipients is made through 3rd person strategies (Nogué 2005, 2008a). The distinction between addressed and unaddressed recipients will be highly relevant to our study.

This chapter focuses on the description of the phenomena while abstracting away from the discussions of theoretical concepts and categories involved in the analysis. Although a discussion as such would be of much interest, it would fall out of the scope and goals of this chapter to analyse the phenomena from the perspective of Brown & Levinson (1987)'s politeness model. However, as highlighted in the conclusion, we will see how the application of Goffman's participation frameworks and third person strategies to the analysis of the different strategies used to refer to participants and their diachronic evolution can help broaden our understanding of interaction, conceived as a key element in the development of any communicative event.

1.2 Data and methodology

This study consists of a corpus-driven qualitative and quantitative analysis. The corpus, divided into six subcorpora, has been taken from the *Diari de Sessions* of the Parliament of Catalonia. The first subcorpus includes six debates of a political (not legislative) nature held in the 1932–1938 period, during the Spanish Second Republic, the time the first modern parliament was convened in Catalonia. The

other five subcorpora consist of the whole text of the “general politics debate” held in the following years: 1980 (the year of the recovery of the Parliament of Catalonia after the Franco dictatorship, with CiU, a centre-right-wing coalition, in government), 1993 (when parliamentary activity was consolidated and CiU had an absolute majority), 2005 (when, for the first time in recent history, a left-wing three-party coalition was into power), 2013 (with CiU in government again and changes in the composition of the Parliament, with the incorporation of the Spanish nationalist party C’s and the radical left-wing party CUP), and 2020 (with a pro-independence centrist and social democratic coalition in government).

These debates are similar to the State of the Union or State of the State addresses in the United States, but they include the opposition’s response and further interaction in the same plenary session of the Parliament. Thus, the five subcorpora of the present-day period comprise full communicative events. The corpus contains 602,641 words, 3.6% of which are in Spanish (1980, 2013, and 2020), which is co-official in Catalonia together with Catalan. Our qualitative analysis focuses on Catalan, but Spanish is included in the quantitative analysis. The six subcorpora vary in length, containing an average of 100,440 words.

The corpus was labelled manually using the categories described in the previous section and then processed with Textstat, a programme designed at the Freie Universität Berlin, and with SPSS for the statistical analysis. Linear regression, Pearson correlation and ANOVA test are the statistical tests conducted to account for the evolution through time in the amount of tokens of some of the phenomena analysed in the qualitative analysis.

2 Qualitative analysis

In this section we discuss the phenomena related to the reference to participants from a qualitative point of view. Although some of them deal with the prototypical uses of the 1st and 2nd persons, most go beyond the prototypical uses of person deixis and beyond the 1st and 2nd persons, and indeed highlighting these phenomena is one of the main contributions of our work. This line of research can also be found in the work of Cornelia Ilie (2003, 2010, 2015) applied to parliamentary discourse.

In §2.1 we discuss the strategies for addresser reference: first, in §2.1.1, the references to the addresser alone, both in the 1st and in the 3rd person; and in §2.1.2, the strategies for the reference to the addresser groups, the groups where he or she includes him/herself; §2.2 is devoted to the strategies for the reference

to the recipient: §2.2.1, to one or several addressed recipients and §2.2.2, to one or several unaddressed recipients.¹

At the end of each strategy, a short reference is made to its temporal evolution in quantitative terms. Global quantitative data are then analysed in §3.

2.1 Addresser reference

In this section the main functions of the 1SG in parliamentary debate are presented, followed by the main alternative strategies in the 3rd person to refer to the addresser. Then, focusing on the reference to the addresser groups, two uses of the 1PL will be discussed. The section is closed by several uses of the 3rd person to refer to the addresser groups.

2.1.1 The reference to the addresser alone

It is well-known that the 1SG is prototypically used to refer to oneself. Here we focus on the specific uses of this person deictic category that can be found in parliamentary debate. The non-prototypical use of 3SG to establish reference to the speaker in this genre is also analysed.

2.1.1.1 1SG

In parliamentary debate, the prototypical use of the 1SG to refer to the addresser performs several functions. The most genre-specific ones are the following.

The MPs use the 1SG to manage their own discourse, with a metadiscursive purpose, using different kinds of *verba dicendi* (1);² they make statements that convey performative speech acts that involve them individually (2); together with other words that are semantically related and structures such as “*com a* (‘as’) + POSITION”, they emphasize the role they are playing in a specific utterance (3); and finally they interact directly: see (4a) for an interaction between the President of Catalonia and the Leader of the Opposition and (4b), where the President

¹The examples are labelled as follows, in brackets: the name (or names) that identify the addressers, their party’s initials (see the list of abbreviations on page 341) and the year of the subcorpus. In some cases, the position (*Prime Minister, President of the Parliament, President of Catalonia...*) is given instead of the party’s initials in order to clarify the example. Usually the President of Catalonia – also called *President de la Generalitat* – is also the Prime Minister, but on some occasions (1932–1933 and 2005, in our corpus) the two posts were occupied by different people.

²By means of parliamentary metadiscourse, “MPs provide supplementary indications about the intentionality, implications, and goals of their own discourse” (Ilie 2015: 12; see also Ilie 2003).

and another MP present contrasting points of view, which is reflected in the use of explicit personal pronouns (*jo*, ‘I’, vs. *vostè* ‘you formal’).

- (1) assumim el que fins ara era competència de l’Estat, és a dir, *subratllo* això
‘we are taking on something that until now has been a competence of the
State, that is, *I underline* that’ (Pujol, President of Catalonia, 1980)
- (2) i *em disculpo*, perquè potser hauria hagut d’esperar
‘and *I apologize*, because maybe I should have waited’
(Iceta, PSC-PSOE, 2005)
- (3) és el *meu* deure i el *meu* compromís *com a president*
‘it is my *duty* and my *commitment as President*’
(Torra, President of Catalonia, 2020)
- (4) a. després *li parlaré* d’alguns temes concrets
‘*I will talk to you* [sing.] later about some specific issues’
(Mas, Leader of the Opposition, 2005)
- b. *vostè* parla que Itàlia no és exemple. Home, *jo* no l’*he posat* com a
model a seguir
‘you say that Italy is not an example. Well, *I did not put* it as a role
model’ (Maragall, President of Catalonia, 2005)

The use of the 1SG throughout our corpus follows a decreasing tendency (Table 1). As we will see later on (§2.1.2), MPs more and more prefer to include themselves in groups than to speak in their own behalf alone.

Table 1: The evolution of the 1SG (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
1SG	2,129	2,070	2,119	2,081	1,839	1,380

2.1.1.2 3SG

In Catalan the 3SG allows the addresser to refer to him/herself through a variety of non-prototypical strategies (see Nogué 2011: 124–127, 2015: 226–228). In parliamentary language, making use of a full NP for participant reference is both an

indirect reference strategy and a marker of formality.³ At the same time, it conveys a certain distance from oneself to focus on the institution (5) or the position held (6). In example (7), moreover, the deictic reference is preserved by means of a demonstrative. Finally, the use of *un servidor* ('your humble servant'), a 3SG strategy without deictic inscription in Catalan, conveys mainly modesty and politeness (8).

- (5) *La Presidència no necessita, agraint-ho molt, cap consell*
 'The Presidency, while very grateful, does not need any advice'
 (Comanys, President of the Parliament, 1932)
- (6) *El president no ha perdut la confiança en el seu Govern*
 'The President has not lost confidence in his Government'
 (Maragall, President of Catalonia, 2005)
- (7) *en aquesta tasca tindran sempre, si la volen, la col·laboració d'aquest Diputat*
 'you will always have this member of parliament's cooperation in this task if you want it'
 (Benet, independent, 1980)
- (8) *aquesta exposició d'un servidor*
 'this presentation by your humble servant' (Rigol, Labour Minister, 1980)

As can be observed in Table 2, this strategy is recorded most frequently in the 1932–1938 period, followed by the debate in 1980, when it seems that to some extent the MPs wanted to recover the stylistic tradition predating the break caused by Franco's dictatorship. Even though in 2005 the President of Catalonia, Pasqual Maragall, used this strategy quite often, between 1993 and 2020 it was rarely used by others. Thus, the present trend is to use it only when a distance effect and a focus on the institutional position is sought, while it is used less and less as a marker of formality alone. Furthermore, its use by Maragall may also be considered a feature of his own communicative style. In Table 2 only NPs have been considered, without including verbal morphemes, pronouns and possessives that can be in an anaphoric relation to them.

In Table 2, the figures in italics highlight the uses of *el president* ('the President') in 1980 and in 2005 that correspond to the same addresser each year, the President of Catalonia. In a way, these can be regarded as outliers. In fact, in 1993 the President was the same as in 1980, hence, during his presidency he seems to

³We use the term *full noun phrase* to refer to phrases with a noun as its nominal head (in contrast to pronoun or infinitive NPs).

have changed his discursive style, at least in this respect. The eight tokens of “demonstrative + POSITION” in 1980 (*aquest Diputat*, ‘this MP’), also correspond to the same addresser. Excluding these outliers, the reduction of the total tokens of these strategies in recent decades is even clearer (see the second figure in the global data for 1980 and 2005).

Global quantitative data, including verbal morphemes, pronouns and possessives, are summarized in Table 3.

Table 2: The evolution of 3SG strategies to refer to the addresser (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
<i>la Presidència</i>	14	4	1	0	0	0
<i>el president (de)</i>	1	15	1	15	2	1
demonstr. + N/Rel	7	8	1	1	0	0
others	6	10	0	4	5	3
TOTAL	28	37 / 14	3	20 / 5	7	4

Table 3: The evolution of 3SG strategies to refer to the addresser (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
Reference to the addresser	71	48	9	41	9	15

2.1.2 Addresser-group reference

Beyond the prototypical uses of the 1PL, in this section we focus on two uses of this category: the first one is widespread in Catalan but serves specific purposes in political discourse; the second one is quite new, and only found in some genres. Some non-prototypical uses of 3SG and 3PL are also analysed.

2.1.2.1 1PL with a full NP subject

The combination of a verb in the 1PL with a full NP subject, also in PL (9), “allows the speaker to underline affiliation to a group which the addressee may or may not belong to, and to simultaneously give a clear, not merely deictic, definition of the group” (De Cock & Nogué 2017: 107).

- (9) a. *tots els Diputats encara estem pendents de quin és l'íntim pensament de Lliga Catalana*
 'we members of parliament are (1PL) all still waiting for the private thoughts of the Lliga Catalana' (Lluhí, Prime Minister, 1933)
- b. *Els catalans hem de poder decidir*
 'We Catalans must (1PL) be able to decide' (Navarro, PSC-PSOE, 2013)

This structure can be found through all our corpus and is also common in other registers. Catalan shares it with Spanish, Occitan, and Basque, while English, Italian, and other languages need a 1PL pronoun specifying the subject NP (*We Italians are very friendly, Noi italiani siamo molto gentili*). It is also used with the 2PL: *Els catalans heu de poder decidir* ('You Catalans must be able to decide') (De Cock & Nogué 2017: 108).⁴

On the other hand, this structure has recently adopted a specific variant, mainly in political and trade-union discourse, although it is also spreading in other registers: now the NP subject is in SG and the referent is usually the name of the political party or trade union, a collective noun, by means of which the addresser includes him/herself. In our corpus, the two first tokens of this variant are found in 1980 (10a) and 1993 (10b), and there are some other later tokens, as in (10c). In this variant the lack of agreement applies not only to the grammatical category of person, but also to the category of number.

- (10) a. *Centristes de Catalunya [...] aprofitem la invitació del President*
 'We Centristes de Catalunya [the name of a political party] [...] take (1PL) advantage of the President's invitation' (Cañellas, CC, 1980)
- b. *Iniciativa per Catalunya hem elaborat [...] un document, amb quaranta o cinquanta mesures*
 'We Iniciativa per Catalunya [...] have (1PL) prepared a text with forty or fifty measures' (Saura, IC, 1993)
- c. *la CUP també faríem els mateixos quatre blocs però canviant-ne els títols*
 'We the CUP would (1PL) also make the same groups but would change their titles' (Fernàndez, CUP, 2013)

⁴Basque can emphasize the inclusion in the group with a specific morpheme, *-o-*, in the NP which also appears in other structures, all of them called *plural hurbila* (proximate or close plural) (Hualde & de Urbina 2003: 122 and Zubiri 2012: 68–69): *Italiarrak oso jatorrak gara* or *Italiarrok oso jatorrak gara* ('We Italians are very friendly').

2.1.2.2 Assembly 1PL

In the 2013 and 2020 subcorpora a particular use of the 1PL is found which had not appeared before: the CUP MPs use the 1SG less than the other MPs, in favour of the 1PL (11).

- (11) a. *Provarem de dir-ho tot sense deixar-nos res en nom de la CUP*
'We'll try to say everything without omitting anything on behalf of the CUP' (Fernández, CUP, 2013)
- b. Per iniciar la *nostra* resposta al seu discurs, inicialment *mirarem* d'apuntar algunes dimensions de la crisi que, des del *nostre* punt de vista, és important assenyalar
'To begin *our* answer to your speech, in the beginning *we shall try* to mention several dimensions of the crisis that, from *our* point of view, it is important to point out' (Riera, CUP, 2020)

Besides the quantitative aspect, however, the qualitative analysis of the two grammatical categories reveals that these addressers try to use the 1SG only in some metadiscursive utterances (12), the first use shown in §2.1.1.1.

- (12) a. No *entenc* la *meva* lletra...
'I can't read my own writing' (Fernández, CUP, 2013)
- b. I *acabo* –si *em* dona un segon més, president– emplaçant-los [...] a frenar el despropòsit de l'acord [...] per fer possible el Hard Rock
'And I *finish* – if you give *me* one more second, President – urging you to stop this nonsensical agreement to make the Hard Rock possible' (Sánchez, CUP, 2020)

Finally, in (13) we can see a case of self-correction in this use of the 1PL, which reflects that it is not completely spontaneous.

- (13) I, després, *conec* experiències, *coneixem* experiències de la CUP, sobretot, arrelades al territori i a les comarques
'And afterwards, I *know* some experiences, *we know* some of the CUP's experiences, above all, rooted in the territory and the counties'
(Fernández, CUP, 2013)

The reason for this communicative behaviour can be found in the ideology of the group these addressers represent (pro-independence, anti-capitalist, ecologist, and feminist), and above all, in its assembly-based decision-making processes. In (11a) we observe the use of the 1PL in an utterance that makes the illustrated strategy explicit. The preference for the 1PL, thus, linguistically reflects

the fact that as MPs the addressers speak *on behalf of* the assembly, and not on their own behalf or on behalf of the hypothetical leaders of their political group.

Table 4 shows that the number of tokens of the 1PL follows an upward trend that compensates the opposite trend for the 1SG (§2.1.1.1) and the 3SG and the 3PL (§2.1.2.3 and §2.1.2.4).

Table 4: The evolution of the 1PL (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
1PL	2,377	2,965	2,754	2,543	3,037	3,816

2.1.2.3 3SG with a full NP

The use of a full NP in SG allows reference to the addresser groups. The NP head is also a collective noun, usually in our corpus the name of a party, a parliamentary group, the Government, the majority, or any parliamentary minority (14).

- (14) a. *per a concretar d'una manera ja més ferma el punt de vista del Govern, he de dir que aquest Govern us exposarà l'obra que pensa*
 'to establish more firmly the point of view of *the Government*, I should say that *this Government will present* to you the work *it is thinking of*'
 (Lluhí, Prime Minister, 1932)
- b. *el 80 per cent del Parlament té clar que no podem mantenir l'estatu quo actual*
 '80% of *the Parliament understands* that we cannot maintain the current status quo'
 (Mas, President of Catalonia, 2013)

In an example such as (14a), the use of the demonstrative (*aquest Govern*, 'this Government') preserves deictic reference and makes the inclusion of the addresser in the reference explicit, whereas in (14b) the inclusion of the President in the group he is talking about is achieved through inference. Through this strategy, a distance effect is achieved.

Furthermore, when the NP is in PL, the addresser can choose either the strategy we have seen in §2.1.2.1, where the inclusion in the group is explicit, or the strategy we will see immediately below.

This strategy is most used in the 1932–1938 period and follows a downward trend nowadays, as Table 5 shows. More direct strategies, mainly the 1PL, compensate for this reduction.

Table 5: The evolution of 3SG strategies to refer to the addresser groups (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
Reference to the addresser groups	782	348	268	429	250	180

2.1.2.4 3PL with a full NP

A full NP in PL also allows reference to the addresser groups. In contrast to the strategy discussed in §2.1.2.1, where the verb is in the 1PL, in this case the use of the 3rd person obtains a distance effect from the group and from the addresser him/herself (15).

- (15) a. *el doctor Martí i Julià i aquest que ara us parla propugnaven per a donar al catalanisme liberalista d'aquella època un gran sentit d'universalitat*
 'Doctor Martí i Julià and the person [lit. this] who is addressing you wanted to give the liberal Catalanism of that period a great sense of universality' (Serra i Moret, USC, 1932)
- b. *Aquests cent vint diputats i diputades, aquest Parlament, mereixen respecte*
 'These hundred and twenty MPs (men and women), this Parliament, deserve respect' (Carod-Rovira, ERC, 2005)

The addressers of the examples in (15) are part of the group they are talking about, but the use of the 3rd person allows them to distance themselves and talk as if they did not belong to this group.

As the previous strategy, this one also shows a downward trend, as seen in Table 6. More direct strategies, mainly the 1PL, compensate for this reduction.

Table 6: The evolution of 3PL strategies to refer to the addresser groups (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
Reference to the addresser groups	47	21	16	32	6	10

2.2 Recipient reference

In this section, the main strategies for the reference to the recipient in Catalan parliamentary discourse are presented. We will begin with the strategies to refer to the addressed recipient(s) and then we will move on to the strategies to refer to the unaddressed recipient(s).

2.2.1 Addressed recipient(s) reference

In our corpus, the reference to the addressed recipient(s) is the kind of participant reference that includes the widest range of strategies, both prototypical and (especially) non-prototypical.

2.2.1.1 *Tu* ('you' SING, informal)

In our corpus, the *tu* ('you' SING, informal) form is not used to refer to the addressed recipient. All the political parties and parliamentary groups, from right-wing to left-wing and from 1932 to 2020, follow an unwritten norm for participant reference in parliamentary debates and avoid this form (Payà 2022). The colloquial *tu* is perceived as inappropriate in the context of high formality associated with that communicative event. Moreover, comparing parliamentary debate with other meetings of MPs which are not carried out in public, Payà (2022) observes that in these other speech events the *tu* form is also used, so the private-public opposition has to be taken into account too. This norm is only broken on two different occasions, in the last two subcorpora (2013 and 2020), as we will see immediately below.

In any case, in the debates of the present-day period (interestingly, not during the 1932–1938 period) a number of 2SG tokens are found. They can be included in the following uses:

- a. In some discursive markers, often (though not always) in direct reported speech: *mira* ('look'), *escolta('m)* ('listen (to me)') and *espera* ('wait') (16).

- (16) I vostès diuen: “*Espera*, nosaltres decidirem, però a la nostra manera, no?”
 ‘And you (2PL, formal) say: «*Wait* (2SG, informal), we will decide, but our own way, OK?»’ (Mas, President of Catalonia, 2013)

- b. In its prototypical use, in direct reported speech in which the addresser quotes a conversation with another person, a public person – usually another politician – or an anonymous person. Here, it must be underlined

that, even when the conversational partner is another politician and they talk about politics, the different setting provided by the reported speech, outside the Parliament, makes it easy to switch to the informal form *tu*. In (17), Carlos Solchaga was Spain's Minister for the Economy and Finance.

(17) l'any passat vàrem dir a en Solchaga: “El plantejament que *tu fas* és irreal, és voluntarista.”

‘last year we told Solchaga: «the proposal *you* (2SG, informal) *make* is unreal, it is just wishful thinking.»’

(Pujol, President of Catalonia, 1993)

c. With a generic value, which is also a strategy of mitigated reference to the addresser (Nogué 2011: 124 and 2015: 226). This use, mostly found in informal registers (Nogué 2008b: 213), appears for the first time in our corpus in 1993, with a single token; it reappears in 2005, with two tokens; and it increases remarkably in 2013, when several MPs use it, especially Artur Mas (CiU, President of Catalonia) (18), who acts here as an outlier. In the 2020 subcorpus we find this use again but the number of tokens is lower than in 2013.

(18) de què serveix tenir el 50 per cent de participació en l'IVA, si després resulta que quan *incrementes* els IVA *tu no tens* cap rendiment addicional [...]?

‘what is the use of having a 50 per cent share of VAT, if afterwards when *you increase* VAT *you don't gain* any additional revenue [...]?’

(Mas, President of Catalonia, 2013)

Most probably, beyond the reference to participants, the generic use of the 2SG can be considered as a marker of informality in the stylistic evolution of parliamentary discourse – and in our corpus also an individual feature of one participant, the President of Catalonia in 2013.

d. As we said before, the prototypical use of 2SG appears twice in our corpus, one in the 2013 subcorpus (19a) and the other in the 2020 subcorpus (19b).

(19) a. En Junqueras diu: “Estimem Espanya, però no ens en fíem,” –no?– “del Govern espanyol.” Oriol, estimo Catalunya, però no em fio gens del Govern de Convergència i Unió. [...] I *saps* per què? –*saps* per què? Perquè resulta que en polítiques fiscals, en polítiques laborals, [...] sempre estan a l'altre bàndol

‘Junqueras says: «We love Spain but we don’t trust» – no? – «the Spanish Government». Oriol, I love Catalonia but I don’t trust the Convergència i Unió Government at all. [...] And *do you know* (2SG, informal) why? – *do you know* (2SG, informal) why? Because it turns out that in tax policies, in labour policies, [...] they are always on the other side’ (Herrera, ICV, 2013)

- b. seria la proposta transaccional entre el Grup de la CUP - Crida Constituent i el nostre grup, sobre el Hard Rock Cafe –ai!, “cafe” no, *perdona*

‘it would be the transactional proposal between the Group of the CUP - Crida Constituent and our group, about the Hard Rock Cafe – ah! not «cafe», *sorry* (2SG, informal, in Catalan)’

(Segovia, CC-P, 2020)

In (19a) a direct interaction of the addresser with an MP of another political group is found. When the speaker addresses him, he seems to “forget” the formal situation they are in. In fact, he seems to forget it from the beginning of the utterance, when he uses his colleague’s family name with an article (*en Junqueras*) and his first name as a vocative (*Oriol*), which are clearly informal and extremely unusual in parliamentary debate. In (19b), after his failing to name properly a giant casino and leisure complex project, the MP recurs to *perdona*, the informal version of a very frequent formula to apologize.

Even as unique tokens, they do not appear in the first subcorpora but in the most recent ones; the present-day social setting seems to allow a kind of *slippage* that would not have been easy to imagine only thirty years ago. Another question arises here: has the left-wing orientation of the parties involved in these two examples anything to do with these tokens? A larger corpus would be needed to answer it properly.

Table 7 summarizes the evolution of the 2SG in our corpus. From 1980 on, a sustained upward trend is observed that peaks in 2013 due to its use with a generic value by Artur Mas, as just mentioned.

Table 7: The evolution of 2SG (informal) (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
2SG (informal)	0	10	9	26	80	34

2.2.1.2 *Vós* ('you' PL, respectful)

In the Catalan three-degree honorific system (*tu-vós-vostè*, §1.1), *vós*, which refers to a single addressed recipient although it is morphologically 2PL, was used regularly during the 1932–1938 period: the presidents of Catalonia and of the Parliament and the MPs of the different parties and coalitions used it. It conveys respect and is structurally similar to its French equivalent *vous* (a more detailed description can be found in Coromines 1971: 88–89, Nogué 2011: 134–135, 2015: 232–233, and GIEC 2016: 195–196).⁵ In (20) the speaker addresses only the MP Josep M. Espanya (ERC) although the forms in italics are in the 2PL.

- (20) És això justament el que jo *us* demanava i em plau que ho *veieu* així
 'This is precisely what I asked *you* (2PL, respectful) for and I am pleased
 that *you* (2PL) *see* it this way' (Martínez Domingo, Lliga, 1933)

In that period, this form alternated with the formulaic form (*la*) *Vostra Senyoria* (*Your Lordship*) (see §2.2.1.4 below), which was much more frequent: 40 tokens of *vós* and 200 tokens of (*la*) *Vostra Senyoria* (per 100,000 words). In the present-day period, only 4 tokens of *vós* are found in 1980, and only one MP (Àngel Colom, ERC) used it systematically in 1993 (also an outlier).⁶ These data are summarized in Table 8.

Table 8: The evolution of *vós* (respectful) (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
<i>Vós</i> (respectful)	40	4	366	0	0	0

2.2.1.3 *Vostè* (3SG, formal)

Recorded only occasionally in the 1932–1938 period, *vostè* has become more and more used as a strategy to refer to the addressed recipient in Catalan parliamentary debate. It comes from the formulaic form *vostra mercè* (*your grace*), from the *vós* form (*vostra* is the stressed feminine possessive of *vós*); when it is the subject,

⁵Nogué (2022) offers a general overview of the present unsteady situation of the Catalan honorifics system, and Payà (2022) of the use of *vós* in the Catalan Administration.

⁶In 1993, in addition to Colom, only the President of the Parliament (3 tokens) and the President of the Government (2 tokens) use it, and only when answering him, hence influenced by his own usage.

it agrees in the 3rd person with the verb (21). Although the Catalan Administration chose *vós* for its relations with the citizens forty years ago (when Catalonia recovered its political institutions and Catalan become the main language of communication both within the Administration and between the Administration and the citizens), today *vostè* is the general preferred strategy for formal interaction, especially in oral contexts (Nogué 2022, Payà 2022). Parliamentary debates are no exception: all parties and coalitions use it, including the anti-capitalist left-wing, as can be observed in the examples (for a more detailed description, see Nogué 2011: 134–136, 2015: 232–233).

- (21) a. Senyor Macià Alavedra, no ens ofenguem, *vostè està* a la dreta i jo a l'esquerra. Si *vostè es considera* de dreta popular, jo em considero d'esquerra popular
 'Mr. Macià Alavedra, do not be offended, *you* (3SG, formal) *are* on the right and I am on the left. If *you* (3SG, formal) *consider yourself* (3SG, formal) popular right-wing, I consider myself popular left-wing'
 (Gutiérrez Díaz, PSUC, 1980)
- b. *Vostè parlava* de colideratge, una tesi que surt molt del marquès d'ESADE, si em *permet* dir-ho així
 'You (3SG, formal) *talked* about co-leadership, an idea that goes out a lot from the marquis of ESADE,⁷ if *you* (3SG, formal) *allow* me the expression'
 (Fernández, CUP, 2013)

This use of *vostè* in parliamentary debate is fully consistent with its use as a marker of formality and politeness in other communicative contexts in present-day Catalan society.

Like the corresponding form in SG, the PL form *vostès* is found only occasionally in the 1932–1938 period. In contrast, in the present-day period it is increasingly used by all the parties of the Parliament of Catalonia, as in (22).

- (22) Aquest Parlament s'ha de reactivar, i *vostès*, més que cap altre grup polític, *tenen* la responsabilitat de reactivar la vida d'aquest Parlament. Moltes gràcies per la *seva* atenció
 'This Parliament has to be reactivated, and *you* (3PL, formal), more than any other political party, *have* the responsibility of reactivating this Parliament. Thank you very much for *your* (3PL, formal) attention'
 (Obiols, PSC-PSOE, 1993)

⁷The expression *marquès d'ESADE* is a play on words on the *marquis de Sade*. ESADE is a prestigious Business School in Barcelona.

The above example also reveals that this form, like other forms of reference to the addressed recipient, can have different referents in consecutive utterances (and even in the same utterance): in the first one, the addresser is referring to the parliamentary group in government (CiU) and, in the second, to all the MPs.⁸

The evolution in the use of *vostè(s)* can be observed in Table 9. The 2013 and 2020 figures suggest that the upward trend observed before may have come to a standstill.

Table 9: The evolution of *vostè(s)* (formal) (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
<i>Vostè</i> (formal, sing.)	4	312	858	1,388	1,712	1,588
<i>Vostès</i> (formal, pl.)	4	401	692	1,004	1,243	1,235

2.2.1.4 (la) Vostra Senyoria (Your Lordship)

As already noted (§2.2.1.2), *(la) Vostra Senyoria (Your Lordship)* is the most frequent strategy in the Second Republic subcorpus for referring to the addressed recipient. Like *vostè* (see §2.2.1.3), it derives from the *vós* form; in subject position it agrees with the verb in the 3SG and is usually represented in the Catalan Parliament's *Diari de Sessions* by means of the abbreviation *V. S.* (VV.SS. in plural) in the 1932–1938 period (23).

- (23) Senyor Lluhí: en l'article 14.è de l'Estatut s'estableix, d'una manera precisa i categòrica, una cosa que també *ha reconegut V. S.*
 'Mr Lluhí: the 14th article of the Catalan Constitution established, in a precise and categorical way, something that *Your Lordship* has also recognised'
 (Ventosa, Lliga, 1932)

This form, as a marker of high formality, is strongly associated today with a specific speech event: a trial in court (De Cock & Nogué 2017: 116). In the corpus from the present-day period, no SG token of this honorific form is found.

The *(les) Vostres Senyories* plural form is used systematically in the 1932–1938 subcorpus, although it is used to a lower extent than *vosaltres*, as we will see in §2.2.1.5 (24).

⁸A multimodal analysis of fragments like this – including gestures, body-position and above all gaze – would allow us to confirm this more precisely.

(24) El 12 d'abril *les VV. SS. es quedaven a casa seva*

'On April 12th *Your Lordships stayed at home*'

(Lluhí, Prime Minister, 1933)

This PL variant is occasionally found, with only two tokens, one in Catalan (25a) and the other in Spanish (25b), in the 1980 subcorpus, recalling the tradition previous to the Franco dictatorship. However, in the case of these tokens its morphological form is related to *vostè* (*ses senyories*, *sus señorías*; *ses* and *sus* are 3rd person possessives), not to *vós*. This change in grammatical person may once again reflect the strength of the emergence of *vostè(s)* in the present-day period, and also the influence of the practices of the Spanish Parliament, where it is a usual form of address (De Cock & Nogué 2017).

(25) a. em limitaré [...], per no cansar *ses senyories*, a mencionar d'una manera bastant puntual cadascun dels punts

'I will restrict myself [...], so as not to bore *Their Lordships*, to mentioning each of the points quite briefly' (Vicens, ERC, 1980)

b. a nosotros nos interesa todo, pero *sus señorías saben* que, de forma especial, lo que hace referencia a la política cultural y social

'we are interested in everything, but as *Their Lordships know*, we are especially interested in matters of cultural and social policy'

(Acosta, GA, 1980)

The evolution in the use of these forms is summarized in Table 10.

Table 10: The evolution of (*la*) *Vostra Senyoria* and (*les*) *Vostres/Ses Senyories* (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
(<i>la</i>) <i>Vostra Senyoria</i>	200	0	0	0	0	0
(<i>les</i>) <i>Vostres/Ses Senyories</i>	43	2	0	0	0	0

2.2.1.5 *Vosaltres* ('you PL')

In the Catalan three-degree system of honorifics, the form *vosaltres* ('you' PL) corresponds both to the PL of *tu* ('you' SG, informal) and to the PL of *vós* ('you' PL, respectful). In the 1932–1938 subcorpus, it is used as a PL of *vós* (26), and in contrast to the reference to a single addressed recipient in that period, it is the

MPs' preferred strategy: 748 tokens (for 100,000 words), compared with only 43 for (*les*) *Vostres Senyories* (*Your Lordships*).

- (26) Honorables Diputats: La *vostra* presència ací, com a representants que *sou* del poble i membres d'aquestes Corts catalanes
'Honourable Members: *your* (2PL, respectful) presence here, as representatives (lit. that you (2PL, respectful) are) of the people and as members of this Catalan parliament'
(Macià, President of Catalonia, 1932)

The *vosaltres* PL form is used only occasionally in the present-day period. Given the absence of the honorific *vós* from 2005 onwards (see section §2.2.1.2), it must be interpreted as a PL of the informal form *tu* used to refer to all MPs or to some of them (27). These tokens, even if they are only occasional, are linked with a certain loss of formality in present-day Catalan parliamentary debates.

- (27) *les nostres ciutats metropolitananes, de les quals alguns de vosaltres sou o heu estat* alcaldes
'our metropolitan cities, of which *some of you are or have been* (2PL, informal) mayors'
(Junqueras, ERC, 2013)

In present-day parliamentary debate, these tokens of *vosaltres* can also be seen as an unmarked strategy which, due to its plural form and its lack of an individual referent, does not share all features of *tu* – linked to informal and colloquial speech events – and, thus, allows the addresser to avoid the features associated with *vostè*: formality, respect and distance. This is why a sustained increase of this use in the Parliament of Catalonia is to be expected in the near future.

As can be seen in example (28), in some cases this form is also found in fragments of direct reported speech, like *tu* (2SG, informal, §2.2.1.1).

- (28) *els ho hem de dir, haurem de dir amb coratge: "Escolteu, aquí no n'hi haurà mai, de regs. Per tant, feu el que vulgueu"*
'we have to tell them, we will have to say with courage: «Listen (2PL), there will never be any irrigation systems here. So *do* (2PL) whatever you (2PL) want»'
(Pujol, President of Catalonia, 1980)

The evolution of the 2PL *vosaltres* is summarized in Table 11.

Table 11: The evolution of 2PL (*tu* and *vós*) (informal or respectful) (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
2PL (<i>tu</i> and <i>vós</i>) (informal or respectful)	748	33	43	19	49	24

2.2.1.6 Vocatives

The use of a vocative to refer to a single addressed recipient is a strategy that is seldom used in the 1932–1938 subcorpus: in that period, fewer types and also fewer tokens occur than in the present-day subcorpora. The reason can be found in the fact that vocatives are a direct way of addressing the recipient; in earlier times, less direct, more formal strategies of address were preferred. Today, a greater diversity of forms and a greater number of tokens of SG vocatives is used by MPs to address the President of Catalonia, the President of the Parliament and other MPs.⁹

The main structures that constitute the SG vocatives of our corpus are the following (they have been ordered according to a chronological criterion, from older to more recent, which combines with structure and formality):

1. *(Molt) honorable (senyor(a)) president(a)* [(Very) Honourable (Mr./Ms.) President]. The different variants of this strategy, the most complex structurally, are the most formal vocatives used in the Parliament of Catalonia.
2. *Senyor(a)* (Mr./Ms.) + POSITION. Above all, *senyor(a) president(a)* (Mr./Ms. President); but also *senyor(a) diputat -ada* (Mr./Ms. Member of parliament), *senyor(a) conseller(a)* (Mr./Ms. Minister), *senyor secretari* (Mr. Secretary), *senyor portaveu* (Mr. Spokesman)...
3. *Senyor(a)* (Mr./Ms.) + FAMILY NAME(S). This is the general vocative used in formal situations and it is used throughout our corpus, from 1932 to 2020. Some examples are: *senyor Lluhí*, *senyor Gutiérrez Díaz*, *senyor Mas*, *senyora Rovira*. Very occasionally, this vocative includes the first name (*senyor Jordi Pujol*).

⁹A detailed study of vocatives in parliamentary debate is beyond the scope of this chapter. We will only highlight the general trends found in the corpus. We do not distinguish between the use of upper case (used most in the two first subcorpora) and lower case letters (used most from 1993 onwards).

4. POSITION alone: *president(a)* (masculine or feminine), *diputat/diputada* (*member of parliament*, masculine or feminine), *conseller(a)* (*minister*, masculine or feminine). This structure appears for the first time in 1993, but in that year it was used by a single MP (Colom, ERC) only to refer to the President of Catalonia and to the President of the Parliament; in 2005 and 2013, it was used by several MPs and in 2020 it was already the preferred form of vocative to address the presidents and the MPs. These vocatives, with their very simple structure, which has no honorific form and only expresses the position, are among the clearest markers of the trend towards less formality in the Catalan parliamentary debates in modern times.

In contrast to the vocatives in SG, the highest number of tokens of vocatives in PL is found in the Second Republic subcorpus. This is probably because the reference to a group makes this strategy more indirect and, thus, a more acceptable one in that context and in that time, and because of a more frequent use of the phatic *senyors diputats* (*Gentlemen members*) to address all MPs. The subsequent evolution must be related to a redistribution of the reference strategies found in the present-day period, especially to the surge in the use of *vostès* (§2.2.1.3), and also to a less frequent use of the phatic (*senyors*) *diputats* / (*senyores i senyors*) *diputats*.

The main structures we find in the PL vocatives of our corpus are the following (a chronological criterion has been followed here):

1. *Senyors diputats* (*Gentlemen members*). This is, by far, the most frequently used PL vocative during the 1932–1938 period, when in the Parliament were only men, and also appears occasionally in the first present-day period subcorpora.
2. *Senyores i senyors diputats* (*Ladies and gentlemen members*). This structure is found from 1980, when women began to take seat as MPs in the Parliament of Catalonia, until 2005, when it began to alternate with structure 3 below. This is probably one of the first contexts in which masculine and feminine forms appear more or less systematically in coordination (instead of only the masculine ones, which can have a generic value in Catalan).
3. *Senyores diputades i senyors diputats*; or *senyors diputats i senyores diputades*; or *senyores diputades, senyors diputats*; or *senyors diputats, senyores diputades* (all combinations of versions of '*Ladies and gentlemen members*'). This structure splits men and women into two coordinated NPs and is found especially in the 2013 subcorpus.

4. *Diputats i diputades*; or *diputades i diputats*; or *diputades, diputats*; or *diputats, diputades* (all combinations of versions of ‘members’ (masculine) and ‘members’ (feminine)). Like structure 3 above, this vocative separates men and women in two coordinated NPs and only expresses their position. This is the preferred strategy in the 2020 subcorpus, and, together with singular structure 4, seems to be the last step towards a radical simplification of vocatives in the Parliament of Catalonia’s debates, which is associated with a more general drop in formality.

The quantitative tendencies of both SG and PL vocatives can be observed in Table 12.

Table 12: The evolution of SG and PL vocatives (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
SG vocatives	62	237	260	232	348	269
PL vocatives	92	70	63	62	57	51

2.2.1.7 3SG with a full NP

Goffman’s (1981) distinction between addressed and unaddressed recipients (see, §1.1) is highly relevant when we analyse the use of the 3rd person to refer to these two different kinds of recipients.

The 3SG is used for the non-prototypical reference to an addressed recipient with purposes similar to those found when it is used to refer to the addresser (§2.1.1.2): as an indirect strategy, it is a marker of formality and a way to convey distance (29).

- (29) a. *Però recordi el senyor Ventosa que...*
 ‘But Mr Ventosa must remember (lit. imperative) that...’
 (Lluhí, Prime Minister, 1933)
- b. *El Diputat senyor Benet vol replicar la intervenció?*
 ‘Does the Member of parliament Mr Benet want to reply to this intervention?’
 (Barrera, President of the Parliament, 1980)

Similarly, with a collective noun head (party, parliamentary group...), this strategy also allows reference to one group of the Parliament as the addressed recipient (30).

- (30) a. *Ho accepta així la minoria regionalista?*
 ‘Does the regionalist minority accept it in this way?’
 (Companys, President of the Parliament, 1932)
- b. *El Grup Socialista no hi veuria inconvenient?*
 ‘Would the Socialist Group agree?’
 (Barrera, President of the Parliament, 1980)

Table 13 summarizes the evolution of these two strategies through time.

Table 13: The evolution of 3SG strategies to refer to one addressed recipient and to a group of addressed recipients (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
Reference to one addressed recipient	64	22	5	0	0	0
Reference to a group of addressed recipients	15	8	13	0	0	0

Tokens of the first strategy are found most of all in the Second Republic subcorpus and in the 1980 debate; in 1993 it drops significantly, and it does not appear at all in the 21st-century debates (2005, 2013, and 2020).¹⁰

The second strategy is also used from the 1932–1938 period to 1993. The high distance and formality effect obtained explains its extremely low frequency and the complete absence of tokens from 1993 onwards. In declarative utterances (not in interrogatives and imperatives), the boundary between reference to an addressed or an unaddressed recipient is considerably blurred.

2.2.1.8 3PL with a full NP

Recipients can also be addressed directly through a structure in the 3PL, usually with a full NP in subject or other syntactic positions (31).

¹⁰In the United Kingdom’s House of Commons, “the third person pronoun is the officially acknowledged pronominal form of address” and in the Swedish Riksdag “the third person pronoun used to be the recommended form of parliamentary address. However, the use of the second person pronoun – both plural (‘ni’) and singular (‘du’) – is increasingly frequent in Swedish parliamentary debates” (Ilie 2010: 891).

- (31) a. Resten suspeses les sessions. Per a la vinent, *els senyors Diputats seran advertits a domicili*
 ‘The sessions are suspended. For the next one, *the members will be notified at home*’ (Comanys, President of the Parliament, 1932)
- b. *Als altres portaveus els va bé aquesta agrupació?*
 ‘Do the other spokespersons agree with this grouping?’
 (Xicoy, President of the Parliament, 1993)

In (31a) the President of the Parliament addresses all MPs to give them some important information: how they will be notified the date for the next plenary session. In (31b) it is also the President of the Parliament who addresses a group of spokespersons in an attempt to secure their agreement. Neither the declarative (31a) nor the interrogative (31b) modality of this strategy appear in the 21st-century debates in our corpus, as observed in Table 14. More direct strategies are used instead, especially vocatives together with the *vostè(s)* form of address.

Table 14: The evolution of 3^{PL} strategies to refer to a group of addressed recipients (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
Reference to a group of addressed recipients	3	2	4	0	0	0

2.2.2 Unaddressed recipient(s) reference

2.2.2.1 3^{SG} with a full NP

The 3^{SG} is the unmarked or prototypical strategy for the reference to an unaddressed recipient: as the 2^{SG} only grammaticalizes the reference to an addressed recipient (see §1.1), it cannot be used to refer to this second type of ratified recipient. Catalan, like other languages, uses the 3rd person to refer to them (32).

- (32) a. *Té la paraula el molt honorable senyor Pasqual Maragall*
 ‘The Very Honourable Mr. Pasqual Maragall has the floor’
 (Benach, President of the Parliament, 2005)
- b. *Resta elegit President de la Cambra el Diputat senyor Josep Irla i Bosch.*
 ‘The Member of parliament Mr. Josep Irla i Bosch has been elected
 President of the Chamber’
 (Serra i Hünter, temporary President of the Parliament, 1938)

- c. a això *el senyor Junqueras* potser no hi *ha fet* referència, però jo també ho vull esmentar
'Perhaps *Mr Junqueras* did not refer to this, but I want to mention it too'
(Mas, President of Catalonia, 2013)

As the unmarked strategy for the reference to an unaddressed recipient, the 3SG with a full NP is found in all the debates in our corpus (indeed, in all genres and registers with more than two participants; see Nogué 2011: 140–141, 2015: 233–234). What is more, example (32a) illustrates a fixed formula to call on someone to speak, used throughout our corpus by the President of the Parliament as the moderator of the debate. A number of tokens are thus found in all the debates. Even so, this strategy is more frequent in the Second Republic debates, whereas in recent times there is a downward trend, in spite of a small increase in 2005. In this case, the difference in the number of tokens is probably due to a general redistribution of the strategies for referring to participants in a parliamentary debate which, as we will see in §3, is reflected in the quantitative analysis of the corpus.

The use of a full NP in SG, with a collective noun head (party, parliamentary group, plenary session of the Parliament, Government, parliamentary majority, a minority...), also makes it possible to refer to the different groups taking part in the parliamentary debate as unaddressed recipients (33). In (33a) Mr Ventosa is not a member of the Government, but of the opposition; in (33b) Mr Bargalló is a member of the Government who refers indirectly to the majority that supports it.

- (33) a. *El Govern no vol rectificar*
'The Government does not want to rectify' (Ventosa, Lliga, 1933)
b. *El Govern sap que té el suport de la majoria d'aquesta cambra*
'The Government knows that it has the support of the majority of this chamber'
(Bargalló, Prime Minister, 2005)

The downward trend observed in Table 15 suggests that more direct forms of reference are preferred nowadays.

2.2.2.2 3PL with a full NP

Similarly, the 3PL is the unmarked or prototypical strategy for the reference to several unaddressed recipients: as the 2PL only grammaticalizes the reference to one or several addressed recipients (see §1.1), it cannot be used to refer to this second type of ratified recipient.

11 Variation and change in reference to discourse participants

Table 15: The evolution of 3SG strategies to refer to one unaddressed recipient and to a group of unaddressed recipients (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
Reference to one unaddressed recipient	1,230	954	405	743	194	360
Reference to a group of unaddressed recipients	488	321	260	397	153	292

In (34a), for example, the President of Catalonia refers to a group of MPs but he does not address them directly; likewise, in (34b) the President of the Parliament mentions the next speakers in the debate without addressing them directly.

- (34) a. *aquest drama que es viu en aquests pobles, que els Diputats de la demarcació de Lleida coneixen perfectament*
 ‘the dramatic situation in those towns, which *the members from Lleida* know perfectly well’ (Pujol, President of Catalonia, 1980)
- b. *Prosegueix el debat amb les intervencions dels representants dels grups parlamentaris*
 ‘The debate continues with the speeches of *the representatives of the parliamentary groups*’ (Benach, President of the Parliament, 2005)

As observed in Table 16, a distribution matching the pattern identified for the SG structure (§2.2.2.1, Table 15) is observed in the quantitative data, albeit with lower global figures.

Table 16: The evolution of 3PL strategies to refer to a group of unaddressed recipients (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
Reference to a group of unaddressed recipients	244	244	49	149	27	72

3 Quantitative analysis

In this section we present an overview of the quantitative data of the study. We will begin with general participant reference, focusing on two statistically significant linear regressions and one Pearson correlation. Next we will see the quantitative results in detail and how they show significant differences and lead to different conclusions depending on whether 3rd person strategies for participant reference are included or excluded.¹¹

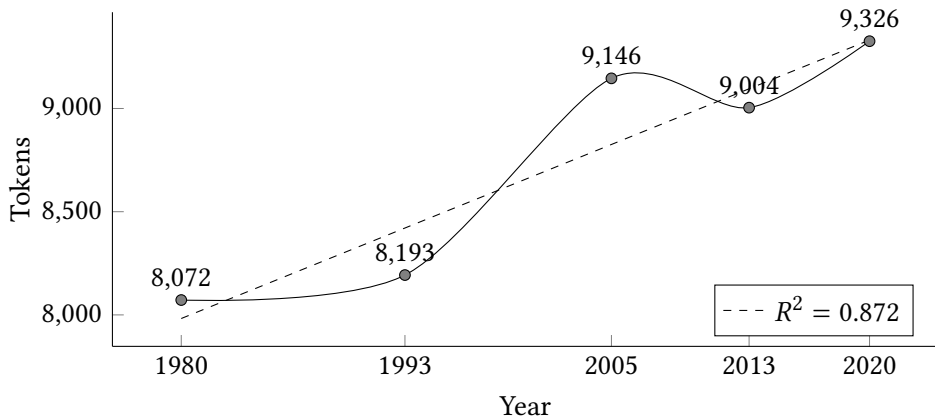


Figure 1: General evolution of participant reference including 3rd person (1980–2020)

Figure 1 shows an upward trend in the global amount of participant reference in the present-day period (1980–2020) in the form of a linear regression (significance values calculated with ANOVA test, $p < 0.02$): from 1980 on, the MPs make more references to themselves, both as addressers and as recipients. So there is more participant inscription in the discourse today than in the past.

If we focus on the evolution of the internal distribution of the reference to the addresser alone and the reference to one addressee (both addressed and un-addressed recipient), we also find a patterned evolution in the form of a linear regression (significance values calculated with ANOVA test, $p < 0.03$) (Figure 2): from more references to the addresser (in the 1932–1938 period and in 1980) to increasingly more references to the addressee (until 2005–20). Hence, as for individual reference, in the past MPs referred more to themselves than to other MPs, whereas now they refer more to other MPs than to themselves. Thus, the higher participant inscription we saw in Figure 1 includes a tendency toward a

¹¹We are very grateful to Pau Francesch (University of Birmingham) for his help in the quantitative analysis of the data.

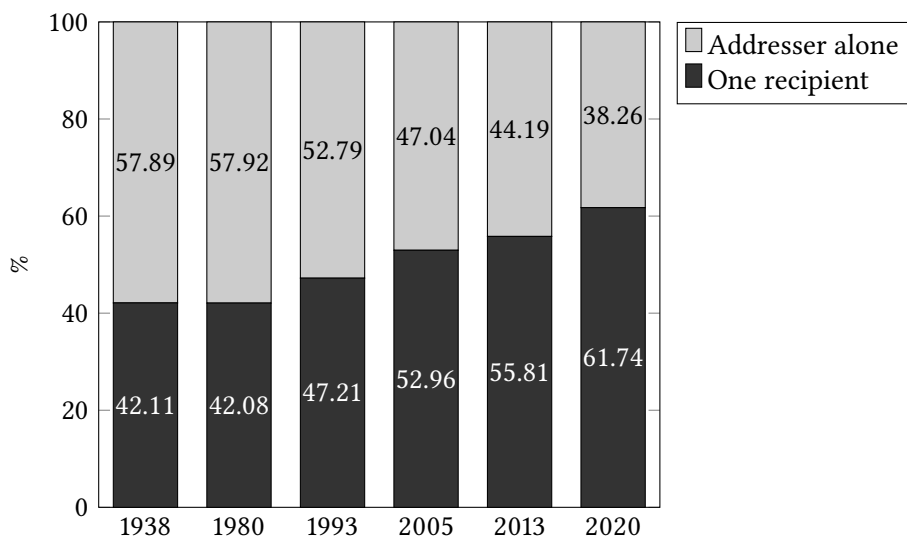


Figure 2: General distribution of the reference to an addresser alone and one recipient (%)

more addressee-oriented reference, at least when the speaker's discourse has a single recipient, either addressed or unaddressed. These data can be read as a patent rise in the interactivity of Catalan parliamentary debate. The fact that the general politics debate concerns the action of the Government, presented and defended by its President, can help explain such an increase in both absolute and relative terms (De Cock 2014: 261–262).

We focus now on the addresser reference and compare the reference to the speaker alone and the reference to the addresser groups (Figure 3). In this case, we can see a Pearson correlation in the data (significance values calculated with Pearson correlation, $p < 0.012$): the two variables show a similar amount of tokens for decades but, from 2013 onwards, they develop in opposite directions: the reference to the addresser groups grows while the reference to the speaker alone falls. That is, in the last decade, MPs, presidents and ministers include themselves in a group (the Government, the Parliament, the party, the country, and so on) more often than before, and at the same time they refer less and less to themselves alone. These figures seem to mirror an evolution in the conception of politics and government from a more individual to a more collective one.

Table 17 includes the general quantitative results (both broken down into variables and global) without taking into account the 3rd person strategies; Table 18 includes the 3rd person results; and Table 19, the changes in global data due to their incorporation in the analysis.

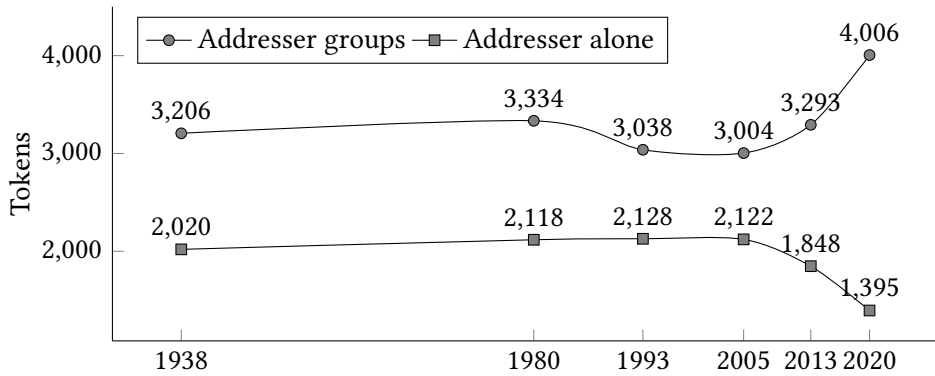


Figure 3: The evolution of addresser-alone and addresser-group reference

Table 17: Global data without 3rd person (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
1SG	2,129	2,070	2,119	2,081	1,839	1,380
1PL	2,377	2,965	2,754	2,543	3,037	3,816
Total addresser	4,506	5,035	4,873	4,624	4,876	5,196
2SG (informal)	0	10	9	26	80	34
Vós (respectful)	40	4	366	0	0	0
Vostè (formal)	4	312	858	1,388	1,712	1,588
(la) Vostra Senyoria	200	0	0	0	0	0
SG vocatives	62	237	260	232	348	269
Total one recipient	306	563	1,493	1,646	2,140	1,891
2PL (<i>tu</i> and <i>vós</i>) ^a	748	33	43	19	49	24
Vostès (formal)	4	401	692	1,004	1,243	1,235
(les) Vostres/Ses Senyories	43	2	0	0	0	0
PL vocatives	92	70	63	62	57	51
Total several recipients	887	506	798	1,085	1,349	1,310
Total recipient(s)	1,193	1,069	2,291	2,731	3,489	3,201
TOTAL	5,699	6,104	7,164	7,355	8,365	8,397

^a(informal or respectful)

11 Variation and change in reference to discourse participants

Table 18: 3rd person strategies for the reference to the addresser and to the addressee(s) (number of tokens per 100,000 words)

Reference to		1932–					
		1938	1980	1993	2005	2013	2020
3SG	addresser	71	48	9	41	9	15
3SG	addresser groups	782	348	268	429	250	180
3PL	addresser groups	47	21	16	32	6	10
Total 3 rd person addresser		900	417	293	502	265	205
3SG	one addressed recipient	64	22	5	0	0	0
3SG	one unaddressed recip.	1,230	954	405	743	194	360
3SG	a group of addr. recip.	15	8	13	0	0	0
3SG	a group of unaddr. recip.	488	321	260	397	153	292
3PL	a group of addr. recip.	3	2	4	0	0	0
3PL	a group of unaddr. recip.	244	244	49	149	27	72
Total 3 rd person addressee		2,044	1,551	736	1,289	374	724
Total 3 rd person strategies		2,944	1,968	1,029	1,791	639	929

Table 19: Total addresser, total recipient(s) and global data without and with 3rd person (number of tokens per 100,000 words)

	1932–1938	1980	1993	2005	2013	2020
Total addresser without 3 rd person	4,506	5,035	4,873	4,624	4,876	5,196
Total addresser with 3 rd person	5,406	5,452	5,166	5,126	5,141	5,401
Total recipient(s) without 3 rd person	1,193	1,069	2,291	2,731	3,489	3,201
Total recipient(s) with 3 rd person	3,237	2,620	3,027	4,020	3,863	3,925
Global data without 3 rd person	5,699	6,104	7,164	7,355	8,365	8,397
Global data with 3 rd person	8,643	8,072	8,193	9,146	9,004	9,326

The global data concerning the reference to the addresser and to the addresser groups reveal the following trends:

1. A slight global downward trend.
2. A clear increase in the preference for the inclusion in a group with a subsequent drop in the individual references to oneself, as we saw in Figure 3.
3. A gradual rise in the deictic forms of reference (in the 1st person, SG or PL), more direct and less formal, to the exclusion of 3rd person strategies, which are less direct and more formal.
4. In spite of the trend above, the 3rd person strategy that has decreased the least is the use of the SG to refer to a group: mainly, the party, the parliamentary group or the Government.

As for the reference to the recipient, a distinction must be made between references to addressed and to unaddressed recipients.

In the reference to the addressed recipient, the following trends are observed:

1. The disappearance of *vós* in the present-day period. This form was only used systematically during the 1932–1938 period.
2. The disappearance in the present-day period of (*la*) *Vostra Senyoria*, which was also only used systematically during the 1932–1938 period, both in individual and collective references.
3. The decline of most 3rd person non-prototypical strategies in the three last subcorpora (2005, 2013, and 2020). These strategies had been used during the 1932–1938 period, and some tokens are still found in 1980 and 1993. The evolution towards more direct and less formal forms is clear.
4. The systematic use of *vostè(s)* in the present-day period, with a pronounced and sustained growth in both individual and collective references.
5. The sustained increase in the number of vocatives referring to a single recipient. Together with the introduction of more simple vocative forms in recent years (see §2.2.1.6), this growth illustrates once again the rising preference for more direct and less formal forms of reference.
6. The reduction in the number of vocatives referring to addressed recipient groups, which is compensated by the soaring use of *vostè* just mentioned in 4.

In the reference to the unaddressed recipient, the following trends are observed:

1. A general reduction in this kind of reference. Increasingly, MPs prefer to conceptualise the recipient, often a political adversary, as the addressed recipient, and they replace an indirect strategy of reference with a direct one: mainly *vostè(s)*.
2. Within this overall trend, a smaller decrease is observed in the case of the reference to groups through a SG NP: political parties, parliamentary groups, Government...
3. A reduction, also smaller, in the case of the reference to a single unaddressed recipient is observed. The speech formula used by the President of the Parliament to give the floor to an addresser (*Té la paraula el diputat / la diputada...* – ‘The MP... has the floor’) explains the maintenance of this strategy.

Finally, the comparison of the global figures in Tables 17 and 18, summarised in Table 19, reveals straightforwardly how taking into account 3rd person strategies for the analysis provides us with a more accurate view of the reality we want to describe and explain.

Table 17 suggests a sustained increase in participant reference tokens, but Table 19 shows that the real growth is much more moderate. It also shows a change in the preferred strategies: for the reference to the addresser, there is a trend towards the inclusion in groups to the detriment of individual references; for the reference to the recipient, and above all to the addressed recipient, over time we find a more reduced use of 3rd person strategies, the decrease and later disappearance of the use of *vós* and (*la*) *Vostra Senyoria*, and a surge in the use of *vostè(s)*. *Vostè(s)* is also distant and formal, but it is a more direct strategy of reference than 3rd person strategies.

4 Conclusions

Both the qualitative and the quantitative analyses of the corpus show that person deixis only (1st and 2nd person, including honorifics) is not enough to explain the reference to participants; the combination with Goffman's (1981) participation frameworks allows the incorporation into the study of strategies for referring to unaddressed recipients. These strategies are added to other non-prototypical strategies of reference, also in the 3rd person.

This first general conclusion, which is theoretical and methodological, goes beyond the study of participant reference in parliamentary debate and is highly relevant to an understanding of participant reference in general.

As for the specific results of the study, the conclusions can be summarised as follows:

1. Throughout the period analysed, the strategies for referring to the recipients (addressed and unaddressed) and the groups they belong to present a greater variety than those for referring to the addresser and the addresser groups.
2. From a structural point of view, vocatives evolve from more complex (*Molt Honorable Senyor President*) to simpler forms (*president*), and, from a functional point of view, from more formal and indirect to less formal and direct. The increase in the use of SG vocatives is related to the extension of more direct forms of reference to one recipient, and the reduction of PL vocatives can be explained by the extension of the use of *vostès* and by a less frequent use of phatic (*senyors*) *diputats* / (*senyores i senyors*) *diputats*.
3. The loss of *vós* and (*la*) *Vostra Senyoria* (in SG and in PL), and the reduction of the use of 3rd person forms, is compensated by the extension of the use of *vostè(s)*. Hence, actually there is little variation in the total number of tokens of forms of participant reference, although a slight upward trend is observed. This growth also entails an increase in the degree of personalisation of the discourse of parliamentary debate.
4. These conclusions suggest that from the 1932–1938 period until now, and also within the present-day period, there is a major stylistic evolution from more indirect and formal strategies to more direct and less formal ones. This final conclusion is especially meaningful in an institutional setting, and it is consistent with wider processes that have affected many other registers of Catalan in that period, which can be summarised in a constant movement towards less formality within the continuum defined by the two extremes of solemnity (or the highest degree of formality, not absent from the parliamentary debate) and the most informal colloquiality (which would not be expected in this speech event).

Two general questions, among others, remain open for further research. First, whether the movement towards less formality found in the Parliament of Catalonia is also found in other traditions of parliamentary discourse. Besides, it would

be interesting to analyse to what extent this tendency is found in other genres of formal discourse, such as non-parliamentary political discourse, discourse of mass media, or academic discourse.

And second, whether participant reference in other parliamentary traditions presents the same properties and evolution. The different rules and uses found in the United Kingdom's House of Commons, the Swedish Riksdag (Ilie 2010), the Spanish Congreso de los Diputados (De Cock & Nogué 2017), and the Parliament of Catalonia suggest that traditions regarding participant reference vary in several important aspects, and that contrastive analyses are needed to shed light upon the relation between these uses and the corresponding sociocultural contexts.

Abbreviations

The abbreviations used in the text follow the Leipzig Glossing Rules. Additional abbreviations for political parties, coalitions and parliamentary groups are:

CC-P	Catalunya en Comú Podem
CC	Centristes de Catalunya
C's	Ciudadanos
CUP	Candidatura d'Unitat Popular
ERC	Esquerra Republicana de Catalunya
CiU	Convergència i Unió
GA	Grup Andalusista
IC	Iniciativa per Catalunya
ICV-EUA	Iniciativa per Catalunya Verds - Esquerra Unida i Alternativa
Lliga	Lliga Regionalista/Lliga Catalana
MP	Member of parliament
PSC-PSOE	Partit dels Socialistes de Catalunya
PSUC	Partit Socialista Unificat de Catalunya
USC	Unió Socialista de Catalunya

Comments and acknowledgments

A previous version of this research was published in Nogué (2018). This revised, extended version includes a new subcorpus (updated to 2020) for both qualitative and quantitative analysis. In addition, a brief introduction to Catalan has been added. This chapter also includes a new section on the use of the first person singular in parliamentary debate (§2.1) and several minor updates to the

previous version. We are deeply grateful to the anonymous reviewers for their suggestions and comments. This chapter is supported by the project *Ambiguity at the grammar-pragmatics interface: structural and pragmatic factors* (reference PID2019-104453GA-I00), granted by the Spanish Ministerio de Ciencia, Innovación y Universidades to the Universitat de Barcelona.

References

- Brown, Penelope & Stephen C. Levinson. 1987. *Politeness*. 2nd edn. Cambridge: Cambridge University Press.
- Bühler, Karl. 1934. *Sprachtheorie*. Stuttgart & New York: Gustav Fisher Verlag.
- Coromines, Joan. 1971. *Lleures i converses d'un filòleg*. Barcelona: Club Editor.
- Cuenca, M. Josep. 2004. El receptor en el text: El vocatiu. *Estudis Romànics* 26. 39–64.
- Cuenca, M. Josep. 2014. The use of demonstratives and context activation in Catalan parliamentary debate. *Discourse Studies* 16(6). 729–752.
- De Cock, Barbara. 2014. *Profiling discourse participants: Forms and functions in Spanish conversation and debates* (Pragmatics & Beyond New Series 246). Amsterdam: John Benjamins.
- De Cock, Barbara & Neus Nogué. 2017. The pragmatics of person reference: A comparative study of Catalan and Spanish parliamentary discourse. *Languages in Contrast* 17(1). 96–127. DOI: 10.1075/lic.17.1.05dec.
- GIEC. 2016. *Gramàtica de la llengua catalana*. <https://giec.iec.cat>. Barcelona: Institut d'Estudis Catalans. (18 August, 2022).
- Goffman, Erving. 1981. *Forms of talk*. Oxford: Basil Blackwell.
- Hualde, José Ignacio & Jon Ortiz de Urbina. 2003. *A grammar of Basque*. Berlin: Mouton de Gruyter.
- Hymes, Dell. 1974. *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvania Press.
- Ilie, Cornelia. 2003. Discourse and metadiscourse in parliamentary debates. *Journal of Language and Politics* 2(1). 71–92.
- Ilie, Cornelia. 2010. Strategic uses of parliamentary forms of address: The case of the UK Parliament and the Swedish Riksdag. *Journal of Pragmatics* 42. 885–911.
- Ilie, Cornelia. 2015. Parliamentary discourse. In Karen Tracy, Cornelia Ilie & Todd Sandel (eds.), *The international encyclopedia of language and social interaction*. Hoboken, NJ: John Wiley & Sons.
- Jakobson, Roman. 1960. Linguistics and poetics. In Thomas A. Sebeok (ed.), *Style in language*, 350–377. Cambridge, MA: MIT Press.

- Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- Nogué, Neus. 2005. *Dixi de persona i marcs participatius en català*. Departament de Filologia Catalana, Universitat de Barcelona. (Doctoral dissertation).
- Nogué, Neus. 2008a. *La dixi de persona en català*. Barcelona: Publicacions de l'Abadia de Montserrat.
- Nogué, Neus. 2008b. La dixi de persona en el discurs acadèmic oral en català. *Caplletra* 44. 195–218.
- Nogué, Neus. 2011. Person deixis. In Lluís Payrató & Josep Maria Cots (eds.), *The pragmatics of Catalan*, 115–144. Berlin/Boston: De Gruyter Mouton.
- Nogué, Neus. 2015. Catalan. In Konstanze Jungbluth & Federica Da Milano (eds.), *Manual of deixis in Romance languages*, 206–239. Berlin/Boston: De Gruyter.
- Nogué, Neus. 2018. *L'evolució de la referència als participants en el debat parlamentari en català (1932–2013)*, vol. 23. *Quaderns de Filologia: Estudis Lingüístics*: 283–307. DOI: 10.7203/QF.23.13532.
- Nogué, Neus. 2022. El sistema de tractaments del català. In Neus Nogué, Josep Besa, Lluís Payrató & Montserrat Sendra (eds.), *Els sistemes de tractament: Una visió interlingüística i aplicada*. Barcelona: Edicions UB.
- Payà, Marta. 2022. Les formes de tractament en l'àmbit parlamentari. In Neus Nogué, Josep Besa, Lluís Payrató & Montserrat Sendra (eds.), *Els sistemes de tractament: Una visió interlingüística i aplicada*. Barcelona: Edicions UB.
- Payrató, Lluís. 2002. L'enunciació i la modalitat oracional. In Joan Solà, Maria-Rosa Lloret, Joan Mascaró & Manuel Pérez Saldanya (eds.), *Gramàtica del català contemporani*, 4th edn. Barcelona: Empúries, 1149–1220.
- Zubiri, Ilari. 2012. *Gramàtica didàctica del euskera*. Bilbo, Bizkaia: ikasBook.

Name index

- Abreu, Laurel, 26, 27, 50, 75, 82
Achard, Michel, 110
Afonso, Susana P. C., 107, 118, 138
Aijón Oliva, Miguel Á., 2, 6, 43, 45,
48, 57, 206, 209, 214
Al-Kasey, Tamara, 76
Alarcón, Irma, 78, 83, 94
Alarcos Llorach, Emilio, 43
Albertuz, Francisco J., 241, 248, 251
Alexiadou, Artemis, 181
Alfano, Lidia Rodríguez, 213
Allen, Ana Sofia Ferreira, 12, 294
Álvarez, Rosario, 43, 47
Amaral, Eduardo Tadeu Roque, 3, 10,
11, 15, 107, 108, 111, 116, 118–
121, 129, 130, 135, 136, 139
Ammara, Um-e, 253
Ariel, Mira, 2
Ávila-Jiménez, Bárbara, 73
- Barbosa, Afranio G., 130
Barbosa, Pilar, 4, 6
Barcala, Mario, 240
Barrajón, Elisa, 204
Barsalou, Lawrence W., 116
Bassa Vanrell, Maria del Mar, 205
Bayley, Robert, 26, 71, 73, 74, 77
Bazenga, Aline, 149
Beatty-Martínez, Anne L., 79
Bentivoglio, Paola, 4, 8, 42, 45, 52, 70,
73, 74
Bessett, Ryan M., 6, 8, 25–29, 32, 36
- Bhaskararao, Peri, 43
Biber, Douglas, 210, 264
Biderman, Maria Teresa, 283, 286
Bini, Milena, 76
Bittner, Maria, 187
Blackwell, Sarah, 77
Botero, Fernando, 207
Briz Gómez, Antonio, 213
Briz, Antonio, 241
Brown, Esther L., 8, 62
Brown, Penelope, 13, 309
Brown, Robert, 13, 14, 276, 277
Bühler, Karl, 275, 276, 309
Bybee, Joan, 43
- Cabredo Hofherr, Patricia, 10, 108,
110, 111, 119, 238, 239, 257
Caffi, Claudia, 212, 226
Cairns, Helen Smith, 73, 74
Callaghan, Matthew, 75
Callou, Dinah, 159
Cameron, Richard, 4, 5, 50, 73–76,
168, 211, 225
Campos, Héctor, 43
Cappeau, Paul, 111, 118
Carreira, Maria Helena, 14, 15, 275–
277, 293, 301
Carrilho, Ernestina, 9
Carvalho, Ana M., 1, 27, 34, 36, 51, 57,
70, 71, 73, 75, 76, 204
Carvalho, Danniell, 193
Carvalho, Janayna, 6, 118

Name index

- Casielles Suárez, Eugenia, 204
Castañeda, Luz Estella, 207
Casteleiro, João Malaca, 158
Castillo Lluch, Monica, 279
Cerrón-Palomino, Álvaro, 26–28, 30,
32–34, 36
Cestero Mancera, Ana María, 213,
226, 229
Chafe, Wallace, 264
Chambers, Jack K., 150
Chierchia, Gennaro, 184
Child, Michael, 73, 75, 76
Chomsky, Noam, 4, 178, 187
Christoffersen, Katherine, 25, 28
Cinque, Guglielmo, 158, 181
Cintra, Luís Filipe Lindley, 14, 274–
277, 280
Coelho, Sueli M., 111, 118
Company Company, Concepción,
45, 111, 113, 114, 119, 204
Constantin de Chanay, Hugues, 288
Cook, Manuela, 274
Coromines, Joan, 308, 322
Correia, Clara Nunes, 115
Costa, João, 151, 159, 167
Coveney, Aidan, 134, 149
Cuenca, M. Josep, 308
Cuervo, Cristina, 43
Cunha, Celso, 280

D'Alessandro, Roberta, 181
Davies, Mark, 282–286, 295, 300
De Cock, Barbara, 11, 12, 209, 216, 217,
239, 245, 264, 265, 308, 314,
315, 324, 325, 335, 341
De Jonge, Bob, 3, 13
de Omena, Nelize Pires, 148, 159
de Paiva Sória, Maíra Vasconcelos,
152, 154, 168, 274

de Prada Pérez, Ana, 75, 76, 79
De Smet, Hendrik, 240
de Urbina, Jon Ortiz, 315
Delbecque, Nicole, 61
Dewaele, Jean-Marc, 71, 241, 265
Díaz, Axel Hernández, 264
Díaz-Campos, Manuel, 218, 220, 222
Dickey, Eleanor, 283
Dieck, Marianne, 206, 209
Dobrushina, Nina, 152, 170
dos Santos Lopes, Célia, 295
Du Bois, John W., 60
Duarte, Isabel Margarida, 11, 14, 15,
275, 276
Dufter, Andreas, 45, 48
Dumont, Jenny, 70, 73
Dussias, Paola, 79

Egerland, Verner, 113, 177–179, 181–
187, 192, 194, 196–199
Elvira, Javier, 113, 121
Emberson, Jan E., 76
Enghels, Renata, 47
Enríquez, Emilia V., 2, 4–7, 61, 73
Erker, Daniel, 73, 74
Erteschik-Shir, Nomi, 2
Expósito Loureiro, Andrea, 45

Faraco, Carlos Alberto, 274
Fauconnier, Giles, 60
Fenger, Paula, 177–179, 181, 186–188,
194, 195
Fernández Acosta, Diana, 207
Fernández Ramírez, Salvador, 204
Fernández Rei, Elisa, 46
Fernández Soriano, Olga, 9, 204, 239
Fernández, Susana, 204, 205
Ferreira, Marcelo Barra, 115
Ferreira, Michael, 282–286, 295, 300

- Ferreira, Victor S., 75, 76
Finegan, Edward, 210
Flores-Ferrán, Nydia, 26, 27, 50, 73–76, 82, 206, 209, 211, 212, 216, 225, 226, 228
Formentelli, Maicol, 276, 278, 279
Freixeiro Mato, Xosé Ramón, 43, 45
- Galvan, Javier A., 76
Galves, Charlotte, 193, 195
García Pineda, Érick, 57
García, Carmen, 13
García-Miguel, José M., 241, 248, 251
Gast, Volker, 10, 108, 110–113, 118, 138
Geeslin, Kimberly L., 9, 70–73, 75, 77, 80–83, 93, 94, 96
Gelabert-Desnoyer, Jaime, 204, 245
Gelbes, Silvia Ramírez, 169
George, Darren, 82
Gerstner-Link, Claudia, 114–116, 119
Giacalone Ramat, Anna, 11, 111, 113, 114, 116
GIEC, 12–14, 308, 322
Gili Gaya, Samuel, 45
Gilman, Albert, 13, 14, 276, 277
Givón, Talmy, 2, 8, 42, 43, 113
Goffman, Erving, 287, 288, 307, 309, 329, 339
Gómez Torrego, Leonardo, 204
González Vergara, Carlos, 209
González-Rátiva, María Claudia, 208
Goussev, Valentin, 152, 170
Grice, H. P., 76
Gries, Stephen T., 76
Guantiva Acosta, Ricardo, 205
Gudmestad, Aarnes, 70–73, 75, 77, 80–83, 93
Guirado, Kristel, 205, 209
Gundel, Jeanette K., 2
- Gurzynski-Weiss, Laura, 71
Gutiérrez Ordoñez, Salvador, 43, 61
Gutiérrez-Bravo, Rodrigo, 43
Gutiérrez-Rivas, Carolina, 204–206, 209
Guy, Gregory, 5, 73, 74
- Hale, Ken, 187
Halliday, Michael, 248, 253
Hammermüller, Gunther, 275, 276
Harris, James W., 79
Haspelmith, Martin, 111, 113
Haverkate, Henk, 204
Heim, Irene, 2
Heine, Bernd, 279, 281, 300, 301
Hennemann, Anja, 6
Henriques Pestana, Yoselin, 10–12
Herbeck, Peter, 2, 5–7, 62
Hernanz, María Luisa, 204, 209, 217
Heylighen, Francis, 241, 265
Hochberg, Judith, 4, 5, 73
Hoekstra, Jarich, 181
Hollænder Jensen, Mikkel, 204
Holmberg, Anders, 6
Hopper, Paul J., 51, 137, 210, 214, 215, 217, 223
Hualde, José Ignacio, 13, 315
Hummel, Martin, 12, 295
Hurtado, Luz Marcela, 26, 27, 60, 204–207, 209, 210, 214, 220, 222, 225, 229, 231
Hymes, Dell, 309
- Ilari, Rudolfo, 280
Ilie, Cornelia, 308, 310, 311, 330, 341
Isabelli, Casilde, 76
- Jaeger, T. Florian, 78, 83, 94
Jakobson, Roman, 309

Name index

- Jaramillo, June A., 28
Johnen, Thomas, 275, 279
Jónsson, Jóhannes Gísli, 198
- Kanwit, Matthew, 71
Kato, Mary A., 6, 109, 110
Kerbrat-Orecchioni, Catherine, 275,
276, 279, 281, 287, 288, 299
Kibrik, Andrej A., 2
Kitagawa, Chisato, 136
Kittilä, Seppo, 43
Klee, Carol A., 78, 83, 94
Kluge, Bettina, 111, 253, 260, 262
Koch, Peter, 151
Koenig, Jean-Pierre, 119
Kootstra, Gerrit Jan, 76
Krifka, Manfred, 114–116, 119, 184
Kuteva, Tania, 281, 300, 301
- Labov, William, 45, 211, 217, 239
LaFond, Larry, 76
Lamiroy, Béatrice, 61
Lamoglia Duarte, Maria Eugenia,
109
Lapesa, Rafael, 14
Laslop, María Eugenia, 264
Lehmann, Christian, 281
Lehrer, Adrienne, 136
Leon-Castro, Marta, 209
Levinson, Stephen C., 2, 13, 307–309
Liceras, Juana M., 76
Limerick, P., 26, 27
Linford, Bret, 77, 83, 94
Lipski, John M., 13
Llompарт, Miquel, 57
Long, Avizia Y., 71
Lopes, Célia Regina dos Santos, 3,
6, 111, 113, 132, 134, 148, 159,
194
- Lozano, Cristobal, 71, 77
Luján, Marta, 4, 45
- Malchukov, Andrej L., 9, 10
Mallery, Paul, 82
Malovrh, Paul, 78, 83, 94
Marchís Moreno, Mihaela, 47
Marques, Maria Aldina, 11, 14, 15, 276,
281
Martins, Ana Maria, 107, 121, 123, 124,
135, 138, 147, 149, 154, 156–
158, 160, 161, 165–167
Masullo, Pascual J., 8
Matthiessen, Christian, 248, 253
Maurer, Gail, 119
Mauri, Caterina, 116
Maury-Rouan, Claire, 288
Mayol, Laia, 4, 5
Medeiros, Sandi Michelle de
Oliveira, 275, 277, 278, 280
Mihatsch, Wiltrud, 3, 10, 11, 15, 107,
108, 111, 116, 118–121, 130,
135, 136, 139
Moltmann, Friederike, 111, 119
Monge, Félix, 217
Montes Giraldo, José Joaquín, 207
Montes-Alcalá, Cecilia, 27
Montrul, Silvina, 76
Morales, Amparo, 4–6, 42, 45, 73,
205
Moser, Karolin, 12
Mougeon, Raymond, 71
Muñiz Cachón, Carmen, 204, 228
- Naro, Anthony, 158, 163
Nascimento, Fernanda, 274, 276, 280,
286
Navarro, Mario, 47
Neeleman, Ad, 187

- Negrão, Esmeralda V., 6, 109
Nieuwenhuijsen, Dorien, 3, 13
Nogué-Serrano, Neus, 12, 14, 15, 308,
309, 312, 314, 315, 320, 322–
325, 332, 341
Nunes, Jairo, 154
Nunes, José Joaquim, 126

Ochs, Elinor, 264
Oesterreicher, Wulf, 151
Ogawa, Akio, 10
Olave-Arias, Giohanny, 209
Ornelas de Avelar, Juanito, 11, 107,
111, 118, 195
Orozco, Rafael, 3, 5, 7, 11, 12, 32, 36,
60, 204–207, 209, 212, 214,
222, 225, 230, 231
Ortega-Santos, Iván, 205, 206, 209,
210, 214, 220, 229
Otero, Carlos-Peregrín, 47
Otheguy, Ricardo, 4–6, 26, 27, 29, 61,
70, 73, 74, 76, 96, 204, 230,
239

Papastathi, Maria, 10, 11, 118, 238,
239, 251, 257, 262, 265, 266
Paredes Silva, Vera Lucia, 168
Payà, Marta, 319, 322, 323
Payrató, Lluís, 15, 308
Pease-Alvarez, Lucinda, 26, 73, 74, 77
Pereira, Sandra, 151, 154, 155, 159, 167,
274
Pérez-Leroux, Ana T., 76
Pericchi, Natalia, 45
Phinney, Marianne, 76
Pickering, Martin J., 75, 76
Pierre, Émeline, 11, 12, 239–241, 253,
255, 262, 264, 265
Poplack, Shana, 70, 75

Posio, Pekka, 1, 2, 4–7, 10, 11, 42,
45, 62, 107–110, 120, 121, 123,
126, 130, 132, 135, 136, 138–
140, 152, 153, 155, 157, 158,
160–163, 206, 209, 214, 215,
229, 239, 260, 264, 274, 297
Pountain, Christopher, 279
Pozas Loyo, Julia, 111, 113, 114, 119,
204
Preston, Dennis, 71
Preti, Dino, 280

Quesada Pacheco, Miguel Ángel, 12
Quesada, Margaret L., 77

Ramírez, Dora, 206
Ramos, Miguel, 42, 45
Ranson, Diana, 5
Raposo, Eduardo, 158, 280
Reese, Samuel, 240
Regueira Fernández, Xosé L., 46
Reinhart, Tanya, 2
Repede, Doina, 209
Ricos Vidal, Amparo, 204
Rigau, Gemma, 4, 5
Rivas, Javier, 8, 47, 49, 57
Rivero, María L., 43
Rizzi, Luigi, 4
Robinson, Joy L., 12–14
Rodríguez Alfano, Lidia, 205
Rodríguez Louro, Celeste, 76
Rojas, Evelin Hugo, 209
Rossano, Federico, 288
Rothman, Jason, 76, 77
Rumeu, Márcia, 194

Salaberry, M. Rafael, 94
Sankoff, David, 31
Sansò, Andrea, 11, 111, 113, 114, 116

Name index

- Santos, Welber N., 122
Scherre, Marta, 275
Schneedecker, Catherine, 111, 118
Scrivner, Olga, 218, 220, 222
Serrano, María José, 2, 6, 43, 45, 206,
209, 214
Shin, Naomi, 6, 26, 27, 29, 62, 70, 73,
74, 239
Siewierska, Anna, 9–11, 110, 111, 118,
157, 170, 204, 238, 239, 251,
257, 262, 264–266
Silva, Luis Filipe L., 111, 118
Silva-Corvalán, Carmen, 5, 26, 27, 32,
36, 57, 70, 73, 74, 168
Silveira, Agripino S., 152, 159
Snider, Neal, 78, 83, 94
Sodaci, Hande, 76
Solà, Jaume, 4
Song, Kyung-An, 279
Soto Andión, Xosé, 49
Stark, Elizabeth, 45, 48
Steffen, Joachim, 12
Subbarao, Karumuri V., 43

Táboas Baylín, Susana, 9, 239
Tagliamonte, Sali A., 33, 45, 218, 220,
222
Tamminga, Meredith, 211
Taylor, Michael, 159
Thompson, Sandra A., 51, 210, 214,
215, 217, 223
Todolí, Júlia, 12
Toribio, Almeida Jacqueline, 9
Torres Cacoullós, Rena, 4–6, 8, 26,
27, 29, 32, 34, 36, 42, 45, 50,
52, 54, 57, 60, 62, 70, 73–76,
211, 223, 225
Traugott, Elizabeth, 281

Travis, Catherine E., 4–6, 8, 26, 27,
29, 32, 34, 36, 42, 45, 50, 52,
54, 57, 60, 62, 70, 73–76, 82,
152, 159, 211, 223, 225
Trudgill, Peter, 150
Turner, Mark, 60

Uber, Diane R., 3, 13, 15
Uriagereka, Juan, 158

van der Auwera, Johan, 10, 108, 110–
113, 118, 138
VanPatten, Bill, 94
Vázquez Rozas, Victoria, 45, 46, 48,
240
Vianna, Juliana Barbosa de Segadas,
152, 194
Vilkuna, Maria, 10, 157, 158, 162, 163

Weerman, Fred, 187
Weiner, Judith, 239
White, Lydia, 76

Xove, Xosé, 43

Young, Richard, 71

Zentella, Ana Celia, 26, 27, 61, 73,
204, 230
Zilles, Ana M. S., 3
Zubiri, Ilari, 315

Referring to discourse participants in Ibero-Romance languages

This volume brings together contributions by researchers focusing on personal pronouns in Ibero-Romance languages, going beyond the well-established variable of expressed vs. non-expressed subjects. While factors such as agreement morphology, topic shift and contrast or emphasis have been argued to account for variable subject expression, several corpus studies on Ibero-Romance languages have shown that the expression of subject pronouns goes beyond these traditionally established factors and is also subject to considerable dialectal variation. One of the factors affecting choice and expression of personal pronouns or other referential devices is whether the construction is used personally or impersonally. The use and emergence of new impersonal constructions, eventually also new (im)personal pronouns, as well as the variation found in the expression of human impersonality in different Ibero-Romance language varieties is another interesting research area that has gained ground in the recent years. In addition to variable subject expression, similar methods and theoretical approaches have been applied to study the expression of objects. Finally, the reference to the addressee(s) using different address pronouns and other address forms is an important field of study that is closely connected to the variable expression of pronouns. The present book sheds light on all these aspects of reference to discourse participants. The volume contains contributions with a strong empirical background and various methods and both written and spoken corpus data from Ibero-Romance languages. The focus on discourse participants highlights the special properties of first and second person referents and the factors affecting them that are often different from the anaphoric third person. The chapters are organized into three thematic sections: (i) Variable expression of subjects and objects, (ii) Between personal and impersonal, and (iii) Reference to the addressee.