ESTABLISHING A VARIABLE CONTEXT FOR LEXICAL SUBJECTS IN SPANISH[1]

EL ESTABLECIMIENTO DE UN CONTEXTO VARIABLE PARA LOS SUJETOS LÉXICOS EN ESPAÑOL

Philip P. Limerick[2]

---

2   ORCID https://orcid.org/0000-0003-2271-2175
Eastern Kentucky University, Richmond, Kentucky, Estados Unidos. philip.limerick@eku.edu

Abstract

The dominant trend in variationist studies of Spanish subject expression is to focus on pronominal subjects, excluding lexical subjects (LSs). Due to such lack of attention to LSs in previous research, the current paper aims to fill this gap and to gain a better understanding of variation between LSs and subject pronouns (SPs). While previous research that analyzes LSs makes a significant contribution to the scholarship on subjects and enriches our understanding of the functions of LSs, a *variable context* for LSs in the variationist tradition has not heretofore been established. The current paper proposes a variable context methodology for LSs by investigating cases where LSs are produced (e.g. *mi mamá trabaja* 'my mom works'), particularly in contexts in which SPs (overt or null) could have alternatively been produced (e.g. *ella/∅ trabaja* 'she works'). Overall frequencies, constraints, and pragmatic functions of LSs are discussed.

Keywords: *lexical subjects; pragmatics; subject expression; syntactic variation; variable context.*

Resumen

La tendencia dominante en los estudios variacionistas de la expresión del sujeto en español es centrarse en los sujetos pronominales, excluyendo los sujetos léxicos (LS). Debido a tal falta de atención, el presente trabajo pretende llenar este vacío y entender mejor la variación entre los LS y los sujetos pronominales (SP). Mientras las investigaciones previas que analizan los LS hacen una contribución significativa a la investigación sobre los sujetos, hasta ahora no se ha establecido un *contexto variable* para los LS en la tradición variacionista. El presente trabajo propone una metodología de contexto variable para los LS mediante la investigación de casos en los que se producen los LS (ej.: *mi mamá trabaja*), particularmente en contextos en los que los SP (explícitos o nulos) podrían haberse producido alternativamente (ej.: *ella/∅ trabaja*). Se discuten las frecuencias generales, restricciones y funciones pragmáticas de los LS.

Palabras clave: *contexto variable; expresión del sujeto; pragmática; sujetos léxicos; variación sintáctica.*

I. INTRODUCTION

The dominant trend in variationist studies of Spanish subject expression is to focus on pronominal subjects, leaving out lexical subjects (LSs) (exceptions include Silva-Corvalán, 1994; Dumont, 2006; Gudmestad & Geeslin, 2010). Due to such lack of attention to LSs in previous research, this paper aims to fill this gap and to gain a better understanding of variation between LSs and subject pronouns (SPs). While previous research that analyzes LSs makes a significant contribution to the scholarship on subjects and enriches our understanding of the functions of LSs, a *variable context* for LSs in the variationist tradition has not heretofore been established. A key process in determining the contexts where variation is possible between two (or more) variants and which contexts it is not is known as circumscribing the «variable context» (Labov, 1966). For SP variation, this means determining all contexts where an overt SP occurred but could have been omitted (the null SP), and also cases where a null SP occurred, but an overt SP could have been produced (see, e.g., Otheguy & Zentella [2012] for an extensive discussion). However, since LSs are an additional choice a speaker could make in place of third-person overt or null SPs (e.g. *el chico* instead of él or Ø), they would also fit inside the variable context for third-person referents. The current paper proposes a variable context methodology for LSs. To fully understand the variation between LSs and pronouns, this investigation focuses on cases where LSs are produced (e.g. mi mamá *trabaja* 'my mom works'), particularly in contexts in which SPs (overt or null) could have alternatively been produced (e.g. ***ella***/∅ *trabaja* 'she works').

While some studies acknowledge the importance of LSs, the authors simply explain (and validly justify) that they exclude such forms to maintain consistency and enable comparability with previous research (e.g. Geeslin, Linford, Fafulas, Long & Díaz-Campos, 2013; Silva-Corvalán, 2015). Although not expressed in the literature to my knowledge, perhaps another reason for this exclusion is related to limitations on examining grammatical person. That is, LSs can only be analyzed in third-person contexts whereas pronominal subjects apply to all persons. Moreover, according to Labov's (1966) Principle of Accountability, all variants of a linguistic variable should be considered. If following this principle strictly, it would require LSs to be included in analyses of third-person subject expression. An analysis of only pronominal subjects is partially informed by this principle but does not take into account «all the relevant forms in the subsystem of grammar» (Tagliamonte, 2012, p. 10). Building on the few previous studies that have analyzed LSs and extending them by establishing a variable context, the current study seeks to answer the following research questions:

- RQ1: What is the overall distribution of LSs relative to pronouns within the variable context?
- RQ2: Are LSs governed by the same linguistic constraints as overt pronouns?
- RQ3: What are the pragmatic functions of LSs when used in contexts where pronouns would have sufficed?

To answer these questions, both a quantitative and qualitative analysis are executed using conversational data collected by the author via sociolinguistic interviews with speakers of Mexican Spanish in Georgia. The following section briefly reviews some of the literature on both SPs and LSs. Then, the methodology employed is outlined including a preliminary proposal for a new LS variable context. Following the methodology section, the results of the quantitative and qualitative analyses are presented and a discussion of the results follows. Finally, the conclusions drawn as well as potential avenues for future research are discussed.

2. Previous Research: Pronominal Subjects and LSs

Previous studies have repeatedly shown that the following factors (among others) significantly influence subject expression, particularly in the case of pronominal subjects: grammatical person/number (Silva-Corvalán, 1994; Otheguy & Zentella, 2012; Orozco, 2015), switch reference (Cameron, 1994; Shin & Otheguy, 2009; Carvalho & Child, 2011), tense-mood-aspect (TMA, Cameron, 1994; Travis, 2007; Carvalho & Bessett, 2015), morphological ambiguity (Erker & Guy, 2012; Lastra & Martín Butragueño, 2015; Michnowicz, 2015), verb class (Travis, 2007; Otheguy & Zentella, 2012; Orozco, 2015), priming (Cameron, 1994; Flores-Ferrán, 2002; Travis, 2005), and polarity (Lastra & Martín Butragueño, 2015; Geeslin & Gudmestad, 2016; Limerick, 2018)[1].

The current study will focus on three of these factors: grammatical number, switch reference, and morphological ambiguity[2].

Grammatical person/number of the verb has been shown to be the strongest predictor of variable subject expression cross-dialectally (Orozco, 2015). Specifically, first-person singular and third-person singular verbs tend to favor overt pronouns (e.g. Silva-Corvalán, 1994; Flores-Ferrán, 2002; Shin, 2012; Lastra & Martín Butragueño, 2015). In fact, most studies have found that all singular forms in general are more likely to appear with overt SPs compared to plural forms (Orozco, 2015). Similarly, with regard to variation of LSs with pronouns, it has been found that singular referents favor LSs while plural referents favor pronouns (Dumont, 2006).

Switch-reference has also shown a strong influence on subject variation (e.g. Bentivoglio, 1987; Cameron, 1994; Silva-Corvalán, 1994; Bayley & Pease-Alvarez, 1997; Travis, 2005; Prada Pérez, 2009; Carvalho & Child, 2011; Otheguy & Zentella, 2012; Michnowicz, 2015; Orozco, 2015). Specifically, where there is a switch in subject referent, the SP is often overt; when there is no switch, null SPs are preferred. This pattern is generally thought to have a functional influence that has to do with referential tracking (Shin & Otheguy, 2009) and facilitating interpretation of the antecedent for the listener (Cameron, 1994). In addition to overt pronominal subjects, LSs also tend to be used more frequently when there is a change in subject referent (Silva-Corvalán, 1994).

Another important factor in determining the manifestation of subject forms is verbal morphology. Specifically, many researchers have shown that, when examining verb forms that are morphologically ambiguous, these forms favor overt SPs while unambiguous forms prefer nulls (e.g. Travis, 2007; Prada Pérez, 2009; Erker & Guy, 2012; Lastra & Martín Butragueño, 2015; Michnowicz, 2015). Ambiguous verb forms are those that have indistinct person inflections in the first and third-person singular, specifically in the imperfect, subjunctive (both present and past), conditional, and pluperfect. The idea is that speakers use overt SPs to clarify the referent for the listener. However, some scholars have found the opposite effect and argue that true ambiguity is rare and that it is the context that clarifies the antecedent. For instance, Ranson (1991) found fewer overt SPs with ambiguous verb forms relative to unambiguous forms and posits that contextual markers such as previous mention of the referent or background knowledge help to clarify the speaker's intended referent when the SP is null. Regarding LSs, a similar pattern appears to persist in that LSs are preferred with ambiguous verb forms (Silva-Corvalán, 1994).

As previously stated, pronominal subjects are much more researched than LSs in the variationist literature; however, it is important to discuss the exceptional studies that have treated LSs in particular. For example, Dumont (2006) studied LSs among five speakers of New Mexican Spanish. This researcher examined the variation between LSs and pronouns, finding that the significant predictors for such variation were previous realization, distance from previous mention of coreferential subject, information flow, verb class, and grammatical number. Specifically, the use of LSs were favored when the referent was mentioned for the first time, when the previous mention of the referent was also a LS, with motion verbs, and with singular referents. In terms of pragmatic function, Dumont suggests two main motivations for their use: (1) question/answer pair [repetition effect], and (2) to express contrast (when repeated).

Furthermore, Silva-Corvalán (1994) analyzed subject expression in Los Angeles Mexican Spanish, including both pronouns and LSs. In examining potential factors that influenced expressed versus unexpressed subjects, she found that switch reference as well as morphological ambiguity played a significant role. While this researcher did not isolate cases of LSs as variants in the analysis (LSs were simply included with the overt pronouns), she did find that these two factors were significant, suggesting that LSs are governed in a similar way as pronouns by switch reference and morphological ambiguity. Additionally, she found that the switch reference constraint was operative for pronouns only and also with the additional inclusion of LSs, and that the constraint was slightly stronger when LSs were included. This suggests that LSs are more sensitive to a switch in subject referent than are overt SPs. The following section will present the methodology utilized in the current investigation, including the data employed for the analysis of subjects as well as a proposal for a variable context that includes LSs.

3. Methodology

3.1. The corpus

The data for the current analysis were taken from a corpus of sociolinguistic interviews conducted in Roswell, Georgia (Atlanta suburb) in 2015 by the author. The sample for the present investigation consists of 20 first-generation Mexican immigrants in Georgia who were born in various regions of Mexico: Mexico City (8), Acapulco, Guerrero (2), the state of Guerrero (1), Juando, Mexico (1), the state of Zacatecas (1), Cuernavaca, Morelos (1), the state of Morelos (1), Tampico, Tamaulipas (1), San Juan del Río, Querétaro (1), Monterrey, Nuevo León (1), the state of Colima (1), and the state of San Luis Potosí (1). They consist of 12 females and 8 males, and their ages range from 25 to 60. Their average length of residency in the U.S. is 12 years. In terms of education levels, they range from primary school to university. The speakers have a variety of occupations, nearly half of them being small business owners. The interviews lasted between 30 minutes and one hour. They were informal, conversational, and addressed topics of personal history, local community life, differences between the speakers' home countries and the U.S., and experiences adapting to life in Roswell, among others.

3.2. The variable context

In studies that analyze pronominal subjects exclusively, the following types of contexts are typically excluded in order to isolate only cases where variation between an overt and null SP can occur in Spanish: verbs within subject headed relative clau-

ses; verbs appearing with full noun phrases; existential structures (e.g. *haber*; *ser*); *hacer* + time expressions; verbs with inanimate referents; impersonal *se* expressions; imperatives; set phrases where an overt or null SP is categorical (e.g. *¿Qué sé **yo**?* 'I don't know'; Ø *digamos* 'let's say'); among others (see Otheguy & Zentella, 2012 for an exhaustive outline and illustration of these contexts). Since speakers do not generally alternate between an overt and null SP in the above cases, these structures are excluded. To provide a point of comparison for LSs in the present investigation, I first analyzed third-person pronominal subjects, excluding the same types of cases discussed above and conducting a quantitative multivariate analysis in *Rbrul* (Johnson, 2009)[3]. Then, I located all cases of LSs in the 20 interviews (N=362) and began to establish a variable context, which will now be outlined and illustrated in detail. In order to isolate only the cases in which variation can occur between all three subject forms (LSs, overt SPs, and null SPs), I have excluded the LSs that appeared in the following contexts:

- when they introduce a new referent into the discourse (i.e. first mention cases)
- when the previous mention of the referent is at a distance such that a pronoun would not suffice to identify the referent (5+ clauses back)
- when there is a competing referent
- when the subject refers to a collective referent.

Within each of the four environments, LSs were used (near) categorically (97% - 100% of the time). Each of these contexts will now be explained and illustrated in turn.

For first mention cases, the speaker introduces a referent for the first time in the discourse. In example 1 below, the speaker's friend (*un amigo*) is a new referent that had not been previously mentioned in the interview. His friend could not have been referred to with *él* or Ø because the listener would, presumably, not have been able to interpret the antecedent. Thus, first mention cases, generally speaking, can only be realized as LSs or else the listener is likely not able to identify the intended referent[4].

(1) First mention:
    despúes este **un amigo** me invitó a...este, a trabajar con él en...(1.5) en poner carpeta, en los apartamentos [M27Mex][5]
    'after that uh a friend asked me to…uh, work with him in…(1.5) in putting carpet, in apartments'

In other cases, such as example 2, a LS that is *not* a new referent was used, but its previous mention occurred at a distance, with multiple intervening clauses between the target structure and its previous reference[6].

(2)   Distance from previous mention:
       hace dos años <u>mi madre</u> fallece, de cáncer, ujum, este tengo, todos mis her-
       manos, tienen esposas e hijos, sí, en total somos una familia de, 35, personas,
       entre hermanos, nietos, uhm, esposas, esposos, ajá, sí, sí es una familia de, de,
       ajá, pero ya no existe **mi madre** y mi padre tampoco [F52Mex]
       'two years ago my mother dies, from cancer, uh huh, uh I have, all my
       brothers, have wives and children, yeah, overall we're a family of 35, people,
       among siblings, grandchildren, um, wives, husbands, uh huh, yeah, yeah it's
       a family of, of, uh huh, but my mother no longer exists and neither does my
       father'

Thus, the use of *mi madre* (bolded) could not have varied with *ella* or Ø because the use of a pronoun by the speaker would likely make it difficult for the listener to recover the intended antecedent (*mi madre*). This relates to the notion of *accessibility* in relation to pronominal subjects (Givón, 1983; Ariel, 1994). Specifically, the greater the distance between referents and their antecedents, the lower the accessibility and salience of such referents. Thus, less accessible subject referents are more likely to be marked with more coding material to enable the identification of the referent (Givón, 1983), which in (2) is a LS instead of a pronoun due to the degree of distance from the previous mention of *mi madre*. In other words, the distance (5 clauses back) of the first case of *mi madre* (underlined) from the target structure makes a LS necessary for the subsequent reference to the speaker's mother. Other cases of distant previous mention involved a greater number of clauses intervening between the reference and the target (e.g. 10+ clauses), including cases where the previous mention was found in an entirely different speech turn.

Furthermore, competing referents also make the use of LSs necessary. The speaker in example 3 produced the subject form *mi papá* and could not have used a pronoun to refer to her father because of the presence of the previous subject *mi abuelo*. The use of *él* or Ø in place of *mi papá* would indicate reference to her grandfather. That is, although a pronoun could have occurred in place of *mi papá*, it would not have refe-rred to the same subject (her father)[7]. On the other hand, the subsequent mention of *mi*

papá in this example (underlined) was included in the present analysis since it does not have a competing referent, is not a first mention, and does not have a distant previous mention. Specifically, there is continuity in reference in that the previous subject *él* also refers to the speaker's father. In this case, the speaker could have alternatively said *él continúa* or Ø *continúa*.

(3)   Competing referent:
        mi abuelo tenía este, descendencia de españoles...ah...(1) **mi papá** viene de
        una familia de nueve hermanos, él es uno de los mayores...(2) y bueno mi papá
        continúa con… [F39Mex]
        'my grandfather had uh, Spanish descent…uh…(1) my dad comes from a
        family of nine siblings, he is one of the oldest…(2) and well my dad continues
        with…'

Finally, the use of collective referents, as in example 4, presents a case where variation with overt and null SPs is not possible (Otheguy & Zentella, 2012). In this example, neither *la policía* nor *la gente* could have been replaced by *ella* or Ø since collective subjects cannot be referred to with pronouns in Spanish[8]. In sum, due to the (near) categorical use of LSs in the above contexts, these cases were excluded.

(4)   Collective referent:
        entonces **la policía** está más acechando por aquellos lados… y todo eso es que
        hace que **la gente** se venga más para acá [F60Mex]
        'so the police are more vigilant in those areas… and all that it's just that it
        makes people come more toward here'

After examining the cases of LSs (N=362) in the 20 interviews and making the exclusions discussed above, 39 total tokens of LSs were left for analysis within the variable context. The frequencies of all three subject forms were calculated, and the data were coded for three independent factors: Grammatical number (singular vs. plural), switch reference (same vs. switch), and morphological ambiguity (ambiguous vs. unambiguous) to determine their potential role in explaining subject variation. In addition, a contextual analysis of the pragmatic functions of LSs was carried out. In the following section I present both the quantitative and qualitative results of the analyses.

4. RESULTS

4.1. Quantitative Analysis

Regarding the distribution of subject forms in the current data, there were 796 total third-person verb tokens within the variable context, 39 of which were LSs, 189 were overt SPs, and 568 were null SPs (see Table 1).

Table 1. Overall distribution of third-person subjects

LSs	Overt SPs	Null SPs
39/796 (5%)	189/796 (24%)	568/796 (71%)

As mentioned above, a quantitative analysis of third-person pronominal forms was executed first. The independent factors included were grammatical number, switch reference, and morphological ambiguity, and it was found that all except for morphological ambiguity had a statistically significant influence on third-person pronoun usage. As would be expected, overt pronouns were favored with singular verbs and when there was a switch in subject reference.

Table 2 below presents the constraint hierarchies for each factor group. The first column shows each factor group along with their particular levels, and the second column presents the factor weights (FW) for each constraint from highest to lowest probability of appearing with an overt SP. When a FW is closer to 1, this indicates a relative favoring of overt SPs. When it is closer to 0, it generally indicates a disfavoring of overt SPs (see Tagliamonte, 2006, p. 145, 156).

Table 2. Hierarchy of constraints (third-person SPs)

Factor	Factor weight	% Overt	N tokens	*p*-value
Number				1.02e-13
singular	.69	38%	371	
plural	.31	12%	386	
RANGE	38			
Switch reference				0.000303
switch	.58	31%	279	
same	.42	21%	478	
RANGE	16			
Speaker (random) Std. Dev. .54				

In the subsequent analysis, LSs were included with the SPs from the previous model, the dependent variable being subject form (LS vs. pronoun) (see Table 3 below). It was revealed that the same two factors exerted a significant influence, but with opposite ranking. That is, for LSs, the strongest predictor was switch reference followed by Number. The direction of effect remained the same in that LSs, like overt SPs, were favored with singular verbs (FW = .63) and in switch reference contexts (FW = .72). Thus, for third-person pronouns, number is the strongest predictor while for LSs, switch reference is the most important factor. A LS is much more likely to be produced when there is a change in reference from one subject to the next than when there is no change in reference (9% LS vs. 2% LS in coreferential contexts). The implication is that to facilitate referential tracking for the listener and to bring the subject referent into higher salience when it is different from the previous subject, speakers' greatest preference is to use LSs. These findings for the effect of switch reference and grammatical number on LSs are consistent with previous research (Silva-Corvalán, 1994; Dumont, 2006).

Table 3. Hierarchy of constraints (third-person LSs vs. pronouns)

Factor	Factor weight	% LS	N tokens	*p*-value
Switch reference				3.94e-07
switch	.72	9%	308	
same	.29	2%	488	
RANGE	43			
Number				0.00571
singular	.63	7%	397	
plural	.37	3%	399	
RANGE	26			
Speaker (random) Std. Dev. .69				

While a quantitative analysis is very helpful in revealing the probabilistic constraints of LSs and in comparing these to the typical constraints on SPs, we must also consider the discourse context where the LSs are produced. This will shed more light on the pragmatic functions of LS use and the potential motivation for speakers to use them in place of overt or null pronouns. We turn to such an analysis in the following section.

4.2. Qualitative Analysis: Pragmatic functions of LSs

While one of the primary uses of LSs is to introduce new referents into discourse (Silva-Corvalán, 1994), there were several cases in which a LS was used in the current data where a pronoun could have also occurred, thus moving beyond the mere introduction of new referents to additional uses (see Dumont, 2006). The most immediate evidence of this is seen simply by the appearance of the same LS in two consecutive clauses or otherwise close distances, an observation consistent with that of Dumont (2006), as seen in the following example:

(5) <u>mi mamá</u> siempre fue ama de casa…**mi mamá tenía** *más trabajo con nosotros* cuidando, somos cinco hermanos [M41Mex]
     'my mom was always a housewife...my mom had more work caring for us, there are five of us [siblings]'

Thus, the second mention of *mi mamá* 'my mom' does not introduce the referent since the referent was already introduced by the speaker in the immediately preceding clause. Despite Silva-Corvalán's (1994, p. 148) assertion that «if the subject is coreferential with the subject of the preceding sentence, [...] a full subject NP is not acceptable in Spanish», we do see evidence of speakers producing LSs in such contexts[9]. However, in the majority of cases in the Georgia data (95%) the previous coreferential mention was in object position rather than subject position as in example 6:

(6) es lo que le explicaba a <u>mi mamá</u> porque **mi mamá dice** *«*Carolina, hablas como, like… spanglish*»*... [F34Mex]
     'it's what I was explaining to my mom because my mom says «Carolina, you speak like, like...Spanglish»'

Thus, these latter environments (where the previous clause contains a coreferential mention in non-subject position) are simply more common for LSs than are the former environments (where the previous coreferential mention is a subject). In fact, the previous coreferential mention was in subject position only 2% of the time (10/488). This means that, in considering the distribution with pronomimal subjects, the great majority (98%) of immediately preceding coreferential subject contexts yielded a subject pronoun rather than a LS (478/488). Regarding previous mention in non-subject position, LSs occurred 14% of the time (9/64) while pronouns were produced in 86% of cases (overt: 22%; null: 64%). The use of LSs when their referent's first mention was in object posi-

tion could be explained by the notion of *accessibility* and *salience*; referents in object position are thought to be less accessible and less salient than those in subject position (Ariel, 1994). Therefore, speakers may use LSs (instead of pronouns) in these contexts to bring the referent into higher salience. This particular context also illustrates the results of the quantitative analysis for the switch reference constraint, namely that LSs were favored when there was a change in subject referent (switch from *yo* [*explicaba*] to *mi mamá* in example 6)

Similarly, another function of LSs found in the current data was to establish the referent as a topic of more than one clause (see Silva-Corvalán, 1994), as in example 7:

(7)   Sí, muy importante más para la comunidad latina porque… **los latinos** son ah, una cultura que se caracteriza por ser... mm espiritual,
'Yeah, very important more for the Latino community because...Latinos are uh, a culture that is characterized as being...mm spiritual'
I: Uh-huh.
R: En México, si no... me equivoco el noventa y ocho por ciento... son católicos,
'In Mexico, if I'm not...mistaken ninety-eight percent...are Catholic'
I: Sí.
'Yeah'
R: aquí lo que me he fijado es que... muchos mexicanos o latinos que llegan católicos, a veces se van convirtiendo en otras religiones… [F34Mex]
'here what I've noticed is that...a lot of Mexicans or Latinos that arrive Catholic, sometimes they end up converting to other religions'

In this example, the subject *los latinos* is used to establish Latinos as the topic to be continued in the subsequent clauses (*son católicos*, *latinos que llegan católicos*, *se van convirtiendo*). A further illustration of LSs to establish a topic is the following:

(8)   R: ...yo siento que Roswell no es pobre, Roswell es rico…y este... ha cambiado bastante… en, en los policías también [M32Mex]
'I feel like Roswell isn't poor, Roswell is rich...and umm...it has changed quite a bit...in terms of, the police too'
I: Mhm
R: Ahorita **los policías** ya no son tan malos…Um, que ellos hacen su trabajo

'Now the police aren't that bad anymore...Um, they just do their job'
I: Sí.
'Yeah'
R: Tienen que hacer su trabajo… [M32Mex]
'They have to do their job…'

The speaker in example 8 first mentions *los policías* 'the police' as a non-subject, and then repeats it in subject position, thereby establishing the police as his topic. He then continues to refer to the police in the following two clauses with an overt SP and then a null SP. Furthermore, in establishing such topics in the above two examples, speakers may be simultaneously making the subject referents more salient/accessible, similar to the cases discussed above.

As Silva-Corvalán (1994) notes, speakers can also make the communicative choice to highlight the subject referent by means of an expressed subject, for example in situations that are counter to expectation. Example 9 below from the present data demonstrates this function for the LS *ese señor* 'that man'. In the preceding discourse, the speaker talks about a man he knows who never went to school and is illiterate. The first mention of the man is made by a noun phrase in object position followed by several cases of anaphora with overt and null SPs. The speaker then uses a noun phrase in subject position (LS) to refer to the man presumably to draw the listener's attention toward the subject referent at a point in the discourse where unexpected information is introduced: despite the man not being able to read or write, *ese señor ha salido adelante* 'that man has done well for himself'[10]:

(9) conozco a <u>un señor</u> que nunca fue a la escuela, él nunca- él no sabe X T:12:08 no sabe ni leer ni escribir, X T:12:13 ve, símbolos más o menos pero en sí él no sabe leer ni escribir, pero es- **ese señor ha salido** a, adelante es... su vida es muy interesante, este es muy trabajador y ya ha sacado su familia adelante y todo,

'I know a man that never went to school, he never- he doesn't know [unintelligible] he doesn't even know how to read or write, [unintelligible] he sees, symbols more or less but he doesn't know how to read nor write, but th- that man has done well for himself he's...his life is very interesting, umm he's very hardworking and now he has done well for his family and everything'

Furthermore, although appearing only once in this particular context in the Roswell data, LSs are used in question/answer sequences. According to Dumont (2006), this type of usage indicates a «repetition effect between speaker and interlocutor» (p. 286):

(10)   I: ¿De dónde es <u>tu esposa</u>?
       'Where is your wife from?'
       R: **Mi esposa es** de, nacida en Virginia [M27Mex]
       'My wife is from, born in Virginia'

Note that in example 10 the interviewee could have used an overt or null SP but produced a LS. The motive for this could be due to the interviewer's previous use of the same subject. Beyond simply a «repetition effect», then, this exemplifies an inter-speaker priming effect. That is, the use of *mi esposa* 'my wife' (and not *ella* 'she') by the interviewee is influenced by the interviewer's immediately preceding use of *esposa* as part of the question. Further, it may also be the case that the use of LSs is influenced by priming within a single speaker's turn, as is the case with pronominal subjects (e.g. Travis, 2005, 2007). In other words, a speaker's own previous use of a LS may lead them to repeat the LS (instead of using a pronoun) in subsequent clauses. A quantitative analysis of a larger data set would shed some light on this issue. As stated above, only one LS was produced in the context of question/answer sequences. Pronominal subjects, then, were the dominant trend in this environment (19% overt SP and 78% null SP vs. 3% LSs).

5. DISCUSSION

Revisiting the initial research questions posed in the introduction to this article, I will now discuss the findings of the current study and implications for future research.

5.1. RQ1: What is the overall distribution of LSs relative to pronouns within the variable context?

Once exclusions were made for tokens of LSs that did not exhibit variation with pronouns in the current data, it was found that 39 LSs were produced within the variable context. Out of a total of 796 third-person subjects, then, the rate of LSs was 5%, the rate of overt SPs was 24% and the rate of null SPs was 71%. It is difficult to compare the LS rate with previous studies since other researchers simply included all cases of LSs instead of limiting to those inside a variable context based on variationist sociolinguistic methods. These differing methodologies would likely result in disparate rates of LSs.

Nevertheless, the LS rate found in this study could serve as a baseline for future studies, enabling an overall comparison of how frequently speakers produce LSs relative to SPs.

5.2. RQ2: Are LSs governed by the same linguistic constraints as overt pronouns?

Overall, LSs inside the variable context are guided by the same constraints as overt SPs, at least with regard to third-person subjects. In the multivariate analysis of SPs exclusively as well as the analysis with LSs included, the same two constraints guided the subject variation: Grammatical number and switch reference. Moreover, the factor of morphological ambiguity was not significant in either analysis. The main difference lies in the finding that the two significant factors (grammatical number and switch reference) ranked differently for LSs in that Switch reference exhibited a stronger constraint than grammatical number, demonstrating that the preference to use a LS when there is a switch in subject referent and a pronoun when there is no switch accounts for the most variance. The opposite is true for pronominal subjects in that grammatical number plays a larger role and accounts for more of the variation between overt and null SPs while switch reference plays a secondary role.

5.3. RQ3: What are the pragmatic functions of LSs when used in contexts where pronouns would have sufficed?

LSs in the current data, and as appearing within the variable context circumscribed above, tend to occur immediately following or at close distances from their previous coreferential lexical noun phrases; this suggests that they exhibit additional functions aside from simply introducing new referents into discourse for the first time, a finding consistent with that of Dumont (2006). One primary function of such occurrences may be to heighten the salience of their referents and also to establish the subject referent as the topic of subsequent discourse. Additional uses of LSs in the data involved highlighting the referent in relation to unexpected situations as well as repeating a recently mentioned lexical noun phrase (a potential priming effect). Further examination of a more robust dataset of lexical subject tokens within the variable context is necessary to uncover such patterns in a more systematic way.

6. CONCLUSION

In terms of methodological practices, by establishing a preliminary variable context for variation between LSs and pronouns this analysis takes a step forward in better explaining the motivations for speakers to produce LSs. On one hand, this study co-

rroborates previous research on LSs (Silva-Corvalán, 1994; Dumont, 2006). Crucially, the linguistic constraints and pragmatic functions hold for LSs even after accounting for a more restricted variable context. On the other hand, the present analysis diverges in that morphological ambiguity of the verb did not show a significant influence on subject variation. This may suggest that, once the variable context is circumscribed and competing referents are excluded, ambiguity no longer plays a role.

Before concluding, an important limitation to the present work must also be addressed. First, it was found that LSs accounted for only 5% of the dataset of third-person subjects. This seems to indicate very little variability between LSs and pronouns in the current data. Therefore, rather than representing a final and conclusive model for variationist methodology in subject expression research, the variable context proposal presented here should instead be interpreted as a preliminary first step in moving toward a variable context through a case study of 20 speakers of Mexican Spanish. Future work needs to employ both a larger and more dialectally diverse data set of LSs, that may expand on (or limit) the current proposal. For instance, would more LS/pronoun variation be observed with more robust data and with a different variety of Spanish?

At any rate, the methodology and findings of this study can serve as a point of departure for future studies to shed more light on the envelope of variation where LSs and pronouns are situated, the potential role of other linguistic constraints of LS usage aside from the three factors employed here (e.g. verb class, priming, polarity, clause type), as well as an extension of the discourse-pragmatic motivations for a speaker choosing a LS over a pronoun. To facilitate this, future research should employ a more robust dataset that contains more examples of LSs. In addition, it would be interesting to examine potential dialectal differences in terms of LS usage; thus, future work should also employ a more diverse speaker sample in terms of national/ethnic origin.

7. References

Ariel, M. (1994). Interpreting Anaphoric Expressions: A Cognitive versus a Pragmatic Approach. *Journal of Linguistics, 30,* 3-42. https://doi.org/10.1017/S0022226700016170

Bayley, R., & Pease-Alvarez, L. (1997). Null Pronoun Variation in Mexican-descent Children's Narrative Discourse. *Language Variation and Change, 9*, 349-371. https://doi.org/10.1017/S0954394500001964

Bentivoglio, P. (1987). *Los sujetos pronominales de primera persona en el habla de Caracas.* Central University of Venezuela.

Blackwell, S. E. (1998). Constraints on Spanish NP Anaphora: The Syntactic versus the Pragmatic Domain. *Hispania, 81*(3), 607-618. https://doi.org/10.2307/345683

Cameron, R. (1994). Switch Reference, Verb Class and Priming in a Variable Syntax. In *Papers from the Regional Meeting of the Chicago Linguistic Society: Parasession on Variation in Linguistic Theory, 30*, 27-45.

Carvalho, A. M., & Bessett, R. M. (2015). Subject Pronoun Expression in Spanish in Contact with Portuguese. In A. M. Carvalho, R. Orozco, & N. Lapidus Shin (eds.), *Subject Pronoun Expression in Spanish: A Cross-dialectal Perspective* (pp. 143-65). Georgetown University Press.

Carvalho, A. M., & Child, M. (2011). Subject Pronoun Expression in a Variety of Spanish in Contact with Portuguese. In J. Michnowicz, & R. Dodsworth (eds.), *Selected Proceedings of the 5th Workshop on Spanish Sociolinguistics* (pp. 14-25). Cascadilla Proceedings Project.

Carvalho, A. M., Orozco, R., & Shin N. L. (eds). (2015). *Subject Pronoun Expression in Spanish: A Cross-dialectal Perspective*. Georgetown University Press.

Dumont, J. (2006). Full NPs as Subjects. In N. Sagarra, & A. J. Toribio (eds.), *Selected Proceedings of the 9th Hispanic Linguistics Symposium* (pp. 286-296). Cascadilla Proceedings Project.

Erker, D., & Guy, G. R. (2012). The Role of Lexical Frequency in Syntactic Variability: Variable Subject Personal Pronoun Expression in Spanish. *Language, 88*, 526-557. https://doi.org/10.1353/lan.2012.0050

Flores-Ferrán, N. (2002). *Subject Personal Pronouns in Spanish Narratives of Puerto Ricans in New York City: A Sociolinguistic Perspective*. Lincom.

Frascarelli, M., & Jiménez-Fernández, A. L. (2019). Understanding Partiality in *pro*-Drop Languages: An Information-Structure Approach. *Syntax, 22*, 162-198. https://doi.org/10.1111/synt.12184

Geeslin, K. and Gudmestad, A. (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. https://doi.org/10.1075/sic.13.1.03gee

Geeslin, K., Linford, B., Fafulas, S., Long, A., & Díaz-Campos, M. (2013). The L2 Development of Subject Form Variation in Spanish: The individual vs. the Group. In J. Cabrelli Amaro et al. (eds.), *Selected Proceedings of the 16th Hispanic Linguistics Symposium* (pp. 156-174). Cascadilla Proceedings Project.

Givón, T. (1983). Topic Continuity in Discourse: An introduction. In T. Givón (ed.), *Topic Continuity in Discourse: A Quantitative Cross-Language Study* (pp. 1-41). John Benjamins. https://doi.org/10.1075/tsl.3

Gudmestad, A., & Geeslin, K. L. (2010). Exploring the Roles of Redundancy and Ambiguity in Variable Subject Expression: A Comparison of Native and Non-Native Speakers. In C.

Borgonovo, M. Español-Echevarría, & P. Prévost (eds.), *Selected Proceedings of the 12th Hispanic Linguistics Symposium* (pp. 270-283). Cascadilla Proceedings Project.

Johnson, D. E. (2009). Getting off the Goldvarb Standard: Introducing Rbrul For Mixed-Effects Variable Rule Analysis. *Language and Linguistics Compass, 3*(1), 359-383. https://doi.org/10.1111/j.1749-818X.2008.00108.x

Labov, W. (1966). *The Social Stratification of English in New York City.* Center for Applied Linguistics.

Lastra, Y., & Martín Butragueño, P. (2015). Subject Pronoun Expression in Oral Mexican Spanish. In A. M. Carvalho, R. Orozco, & N. Lapidus Shin (eds.), *Subject Pronoun Expression in Spanish: A Cross-Dialectal Perspective* (pp. 39-57). Georgetown University Press.

Limerick, P. P. (2018). *Subject Expression in a Southeastern U.S. Mexican Community.* [Doctoral thesis, University of Georgia, United States of America].

Michnowicz, J. (2015). Subject Pronoun Expression in Contact with Maya in Yucatan Spanish. In A. M. Carvalho, R. Orozco, & N. Lapidus Shin (eds.), *Subject Pronoun Expression in Spanish: A Cross-dialectal Perspective* (pp. 101-119). Georgetown University Press.

Orozco, R. (2015). Pronominal Variation in Colombian Costeño Spanish. In A. M. Carvalho, R. Orozco, & N. Lapidus Shin (Eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. 17-37). Georgetown University Press.

Otheguy, R., & Zentella, A. C. (2012). *Spanish in New York: Language Contact, Dialectal Leveling, and Structural Continuity.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199737406.003.0001

Prada Pérez, A. de. (2009). *Subject Expression in Minorcan Spanish: Consequences of Contact with Catalan.* [Doctoral thesis, Pennsylvania State University, United States of America].

Ranson, D. L. (1991). Person Marking in the Wake of /s/ Deletion in Andalusian Spanish. *Language Variation and Change, 3*, 133-152. https://doi.org/10.1017/S0954394500000491

Shin, N. L., & Otheguy, R. (2009). Shifting Sensitivity to Continuity of Reference: Subject Pronoun use in Spanish in New York City. In M. LaCorte, & J. Leeman (eds.), *Español en Estados Unidos y otros contextos: sociolingüística, ideología y pedagogía* (pp. 111-136). Iberoamericana. https://doi.org/10.31819/9783865279033-007

Shin, N. L. (2012). Variable Use of Spanish Subject Pronouns by Monolingual Children in Mexico. In K. Geeslin, & M. Díaz-Campos (eds.), *Selected Proceedings of the 14th Hispanic linguistics symposium* (pp. 130-141). Cascadilla Proceedings Project.

Silva-Corvalán, C. (1994). *Language Contact and Change: Spanish in Los Angeles.* Clarendon.

Silva-Corvalán, C. (2015). The Acquisition of Grammatical Subjects by Spanish-English Bilinguals. In A. M. Carvalho, R. Orozco, & N. Lapidus Shin (eds.), *Subject Pronoun*

Expression in Spanish: A Cross-dialectal Perspective (pp. 211-229). Georgetown University Press.

Tagliamonte, S. A. (2006). *Analysing Sociolinguistic Variation*. Cambridge University Press. https://doi.org/10.1017/CBO9780511801624

Tagliamonte, S. A. (2012). *Variationist Sociolinguistics: Change, Observation, Interpretation*. Wiley-Blackwell.

Travis, C. E. (2005). The yo-yo Effect: Priming in Subject Expression in Colombian Spanish. In R. Gess, & E. J. Rubin (eds.), *Selected Papers from the 34th Linguistic Symposium on Romance Languages (LSRL)* (pp. 329-349). John Benjamins. https://doi.org/10.1075/cilt.272.20tra

Travis, C. E. (2007). Genre Effects on Subject Expression in Spanish: Priming in Narrative and Conversation. *Language Variation and Change, 19*, 101-135. https://doi.org/10.1017/S0954394507070081

8. ABBREVIATIONS

LS   lexical subject
SP   subject pronoun
FW   factor weight

NOTES

1 From a different perspective, see Frascarelli & Jiménez-Fernández (2019), who take an information-structure approach within an experimental framework.

2 For a more comprehensive discussion of these factors, see, e.g., Otheguy and Zentella (2012) and Carvalho, Orozco, and Shin (2015).

3 See Limerick (2018) for a more extensive discussion of the variable context of SPs with this particular dataset.

4 There were a few exceptions to this first mention generalization observed in the current data. For example, one speaker referred to her son for the first time by using él. This was possible since her son was part of the extralinguistic context (sitting beside her). However, due to only finding two or three cases of such reference in the entire dataset, first mentions were excluded since, for the most part, they did not exhibit variation. Such examples of pronouns for first mention have also been attested in previous research (e.g. Blackwell, 1998, p. 616).

5 This code indicates the sex, age, and national origin of the speaker.

6 LSs were the preferred variant of the speakers when their previous coreferential mentions were at a distance of 5+ clauses back in the discourse. Therefore, for the purposes of the present investigation, when such distance was present in the interviews, the LS was excluded since variation with a pronoun was not observed. This pattern for distance is consistent with Dumont's

(2006) analysis of LSs in that she found that a greater number of intervening clauses between two coreferential subjects resulted in a higher use of LSs and a lower use of pronouns, especially at a distance of 10+ clauses (p. 289).

7   A LS was considered to have a «competing» referent if the previous clause contained a referent different from that of the LS and in a context where the use of a pronoun could not have sufficed to discern between the two referents in question, as in example 3 (Total N competing referents = 5).

8   The only exception to this is when a subsequent reference to the collective is made with third-person plural inflection and/or a plural SP (e.g. *La gente* tiene que aprender el idioma y ***(ellos)*** ***tienen*** *que respetar la ley* 'people have to learn the language and they have to respect the law'). Moreover, as an anonymous reviewer interestingly pointed out, the possibility of referring to a collective with plural subjects/verbs may depend on the dialect as well as the particular collective noun used (e.g. *Orquesta* is not typically referred to with a plural subject/verb).

9   Though no precise explanation of what it means to not be *acceptable* is given by Silva-Corvalán (1994, p. 148) regarding the LS in this context, she discusses comparative examples of subject variation/obligatoriness in the same paragraph as being not *optional* or *allowed*, leading one to interpret *acceptable* in these terms. There is no mention of the words *felicitous*, *odd*, or similar terms to denote pragmatic unacceptability. Thus, though it seems that the meaning of *acceptable* here is perhaps more in line with *grammatical*, the more accurate explanation for the use of LSs in this context and in terms of natural occurring speech is that LSs are less *felicitous* or generally *disfavored* after their coreferential mention in the preceding clause.

10   An anonymous reviewer pointed out that this particular instance of an LS in a context in which there is a given topic established in the previous discourse (*un señor*) could reflect a tendency for some varieties of Mexican Spanish to be moving toward partial pro-drop status (see, e.g., Frascarelli & Jiménez-Fernández 2019). I thank the reviewer for this insight, and this issue should be explored in future research.