B O R E A L I S
An International Journal of Hispanic Linguistics

A LONGITUDINAL INVESTIGATION OF THIRD-PERSON SUBJECT EXPRESSION IN ADDITIONAL-LANGUAGE SPANISH

Aarnes Gudmestad[a] & Thomas A. Metzger[b]
Virginia Polytechnic Institute and State University[a] & The Ohio State University[b]

ABSTRACT. The goal of the present investigation was to examine the variable use of third-person subject forms in additional-language Spanish over a period of 21 months that included an academic year in Spain or Mexico. The data came from the additional-language learners of Spanish in the LANGSNAP corpus (N = 27). We analyzed production data from three points in time: before the participants went abroad, at the end of their nine-month stay, and nine months after the sojourn abroad ended. We coded all contexts of third-person subject forms (K = 9,379) for the subject form used, seven linguistic variables, and five extralinguistic variables. The analysis revealed that learners used a diverse inventory of subject forms, namely unexpressed subjects, five types of pronouns, lexical noun phrases, and verbal phrases. The mixed-effects model demonstrated that learners' variable use of third-person subjects is complex. In brief, four linguistic factors (referent continuity, clause type, specificity, and perseveration) and one extralinguistic factor (task) predicted use. We moreover observed stability in their use of third-person subject forms over time, meaning that no significant changes occurred while learners were abroad.

Keywords: subject expression, SLA, Spanish, study abroad, variation

RESUMEN. El objetivo de la presente investigación fue examinar el uso variable de las formas de sujeto en tercera persona en el español como lengua adicional durante un período de 21 meses que incluyó un año académico en España o México. Los datos provinieron de los aprendices de español como lengua adicional en el corpus LANGSNAP (N = 27). Se analizaron los datos de producción de tres momentos determinados: antes de que los participantes fueran al extranjero, al final de su estancia de nueve meses y nueve meses después de finalizar su estadía en el extranjero. Se codificaron todos los contextos de las formas de sujeto en tercera persona (K = 9,379) para la forma de sujeto utilizada, siete variables lingüísticas y cinco variables extralingüísticas. El análisis reveló que los aprendices utilizaron un inventario diverso de formas de sujeto, a saber, sujetos no expresados, cinco tipos de pronombres, frases nominales léxicas y frases verbales. El modelo de efectos mixtos demostró que el uso variable de los sujetos en tercera persona por parte de los aprendices es complejo. En resumen, cuatro factores lingüísticos (la continuidad del referente, el tipo de cláusula, la especificidad y la perseveración) y un factor extralingüístico (la tarea) predijeron el uso. Además, se observaron estabilidad en el uso de las formas de sujeto en tercera persona a lo largo del tiempo, lo que significa que no se produjeron cambios significativos mientras los aprendices vivían en el extranjero.

Palabras clave: expresión de sujeto, adquisición de segundas lenguas, español, estudios en el extranjero, variación

1. Introduction

One of the most well-studied linguistic phenomena in research on Spanish is subject expression. Much of this attention stems from the classification of Spanish as a pro-drop language, with researchers seeking to differentiate between contexts in which subject forms are overtly realized versus those in which they are unexpressed. This large body of work spans different theoretical frameworks, such as variationism (e.g., Carvalho, Orozco, & Shin, 2015) and Universal Grammar

 Borealis: An International Journal of Hispanic Linguistics, 2025, 14 / 2. pp. 159-184. https://doi.org/10.7557/1.14.2.8234

(e.g., Lozano, 2002; Rothman & Iverson, 2007). It also consists of examinations of different Spanish-speaking populations, including speakers in different regions (e.g., Carmeron, 1993; Carvalho et al., 2015), bilinguals in the United States (e.g., Flores-Ferrán, 2004; Prada Pérez, 2018), and additional-language learners (Brown, Quan, & Rivas, 2025; Geeslin, Linford, & Fafulas, 2015; Long, 2021). While there now exists a wealth of knowledge about subject expression in Spanish, the scope of understanding is primarily limited to animate referents and the contrast between overt personal pronouns and unexpressed subjects (e.g., Shin & Otheguy, 2009). This pair of forms is all that is available for first-person-singular and second-person subjects, which are also animate, but the inventory of third-person forms is more diverse and notably includes the lexical noun phrase (LNP) as a possible subject form, as in *El gato duerme mucho* 'The cat sleeps a lot', and inanimate referents, as in *Esta casa es muy moderna* 'This house is very modern' (e.g., Dumont, 2006). In order to more fully understand subject expression in Spanish, the current study focuses on third-person referents and takes as its starting point all third-person subjects, regardless of form or animacy status.

Moreover, from the perspective of second language acquisition (SLA), third-person subject expression is worthy of exploration due to its multifunctionality, which refers to form-function relationships in which more than one form can convey a given function (Andersen, 1990). This means that additional-language learners are faced with the learnability issue of developing the ability to use and interpret multiple forms to express the function of third-person verbal subjects of finite verbs. Third-person subjects may pose more challenges for learners than first-person or second-person subject expression because the inventory of possible forms is greater. With this in mind, the present study seeks to examine additional-language learners' use of third-person subjects over a period of 21 months, which includes an academic year spent in a Spanish-dominant country, in order to investigate possible development.

2. Background

We begin this section with a brief introduction to the approach we adopt in the current study. We then review research on subject expression in Spanish, giving specific attention to third-person subjects. Because we analyze learners' use of subject expression during a stay abroad, we conclude with a concise overview of study abroad research in SLA.

2.1. Variationist SLA

Variationist SLA, the framework adopted in the current study, seeks to account for learners' variable language behavior and its development. With this approach, researchers traditionally investigate the linguistic and extralinguistic factors that influence the variable occurrence of a linguistic structure (often called a variable structure) and its corresponding linguistic forms (i.e., variants). For example, Pozzi and Bayley (2020) studied learners' production of prepalatal fricative (i.e., the variable structure, represented with the graphemes <y> and <ll>), and the variation between two variants: [ʃ] and [ʒ]. The learners in their investigation spent a semester studying in Buenos Aires, Argentina. Pozzi and Bayley found that the linguistic factor of word type and the extralinguistic factors of task and time predicted additional-language use. Regarding word type and task, learners were more likely to use the voiceless [ʃ] compared with the voiced [ʒ] in words with a vowel-ll-vowel combination (e.g., *taller* 'workshop') and in the interview, and they were less likely to use this variant with the word *yo* 'I' and in the reading task compared to the voiced [ʒ]. With the other extralinguistic factor of time, learners were more likely to use [ʃ] as their time in Buenos Aires increased. Thus, with this framework, researchers are able to explain the

systematic and complex ways in which learners vary language and how these patterns change along the developmental trajectory.

We adopt a variationist approach to study third-person subject expression for two reasons. The first is that research on different speaking populations has shown evidence of variation in the use of subject forms (e.g., Carvalho et al., 2015). This means that the input that learners receive, which plays a key role in language development, likely contains variable usage patterns (Geeslin with Long, 2014). It also suggests that an approach that provides conceptual and methodological tools to understand variation in language aligns well with subject expression because it is a variable linguistic phenomenon. The second reason that we believe the variationist approach is well-suited to the goal of advancing knowledge about third-person subject expression is that it has shown to be an effective framework for investigating the development of multifunctionality in additional languages (Kanwit, 2017). Third-person subject expression is a clear case of multifunctionality because there are various forms that express third-person referents (e.g., unexpressed subjects, personal pronouns, demonstrative pronouns, LNPs). To date, however, knowledge about how additional-language learners vary their use of third-person subject forms is limited (see Gudmestad, House, & Geeslin, 2013, for an exception).

2.2. Subject expression

Beginning broadly with subject expression in Spanish, variationist research has documented the role that several linguistic factors play in explaining variable patterns of use. We review the seven factors that have been investigated in previous studies that we examine in the current study. Although the findings considered here generally come from the investigation of the binary distinction between unexpressed subjects and personal pronouns, we illustrate them with an example in (1) of third-person subject expression from our data that includes the use of LNPs (see *Method* for details on the data). With this example, we focus on the subject in bold; other subjects are underlined. First, one of the most well-studied variables is referent continuity (also called switch reference, Carvalho et al., 2015), which pertains to whether the referent in subject position is the same as or different from the referent of the subject of the preceding tensed verb. In the example below, we focus on the LNP subject in the phrase *de donde su gato ehm dormía normalmente* 'where the cat uhm normally slept'. The preceding subject in the phrase *lo puso eh cerca del lugar* 'she put it uh close to the location' is an unexpressed subject and refers to the cat's caretaker Nati. Thus, there is a change or switch in referent between the two subjects. Studies have consistently demonstrated that unexpressed subjects are more likely to occur if there is no change in referent from the previous subject content (e.g., Bayley & Pease-Álvarez, 1997; Cameron, 1995; Geeslin et al., 2015; Otheguy & Zentella, 2012). Second, clause type refers to whether the subject form occurs in a main or subordinate clause. The LNP subject in bold in (1) is in a subordinate clause. Some investigations have shown that personal pronouns are linked to independent clauses, whereas others have not found a significant effect for this variable (e.g., Abreu, 2009; Gudmestad & Edmonds, 2023; Torres Cacoullos & Travis, 2010).[1] Third, object pronoun pertains to whether or not the finite verb is preceded by an object pronoun. In (1) there is no preverbal object pronoun between the subject *su gato* and the verb *dormía*. Limited evidence suggests that the presence of a pre-verbal object pronoun predicts the use of unexpressed subjects (e.g., Geeslin & Gudmestad, 2016); other work has not found an effect for preverbal object pronouns (e.g., Gudmestad & Edmonds, 2023). Fourth, number distinguishes between singular and plural subject contexts, with

[1] We limit our focus here to a clause type variable that is binary. See Otheguy, Zentella, and Livert (2007), for example, for research that analyzed other types of clauses.

evidence suggesting that personal pronouns are more likely to occur with singular, compared to plural, referents (e.g., Bayley & Pease-Álvarez, 1997; Flores-Ferrán, 2004; Long, 2021; Otheguy & Zentella, 2012). The LNP subject *su gato* is singular. Fifth, specificity differentiates between referents that are identifiable (e.g., can be named) from those that are not in the discourse. The particular role that this factor plays in variable subject expression has been found to vary across grammatical persons and varieties of Spanish (e.g., Cameron, 1993; Lapidus & Otheguy, 2005). The example in (1) comes from a task in which the participant is retelling a story presented to her in pictures. The referent is a specific character in the story—a cat named Pancho. The final two variables are priming and perseveration and constitute two ways of operationalizing a "birds of a feather flock together" effect (Pereira Scherre & Naro, 1991, p. 24), meaning that forms tend to be repeated or that the occurrence of a certain form increases the chance that it will be used again. In the present investigation, priming focuses on the subject form of the referent in the preceding clause, regardless of what the referent is (e.g., Brown et al., 2025; Cameron & Flores-Ferrán, 2004; Orozco, 2015; Travis, 2007). In (1) the subject form of the preceding referent to the phrase *lo puso eh cerca del lugar* is an unexpressed subject, so in this case the use of an unexpressed subject did not appear to encourage the use of another unexpressed subject with the following verb *dormía*. Perseveration refers to the subject form of the previous mention of the same referent (e.g., Geeslin & Gudmestad, 2016; Gudmestad & Edmonds, 2023). The previous mention of the same referent (*su gato*) is also an LNP (in the phrase *que su gato debería tener mucha hambre*).

(1) *Pues Nati estaba muy triste y Ø imaginaba que [/] que su gato debería tener mucha hambre. Y por eso ehm Ø puso alguna comida en su cuarto de [/] de jugar. Eh Ø lo puso eh cerca del lugar de donde* ***su gato*** *ehm dormía normalmente.* 'Well Nati was very sad and she was imagining that [/] that her cat must be very hungry. And for that reason uhm she put some food in his play room. Uh she put it uh near the location where the cat normally slept.'
Participant 158, narrative task, in-stay 3

Turning specifically to third-person subjects, as previously stated, third-person subject expression differs from other grammatical persons in two primary ways. One is that the inventory of possible subject forms is greater than what is available for first- and second-person subjects, and the other is that third-person subjects can be animate or inanimate. The examples in (2) through (5) illustrate these characteristics.

(2) LNP
  a. *Lucía estudia química.* 'Lucy studies chemistry.'
  b. *La química es interesante.* 'Chemistry is interesting.'
(3) Pronouns (demonstrative, indefinite, interrogative, personal, and relative)
  a. *Ésta es interesante.* 'This is interesting.' (demonstrative)
  b. *Nadie estudia química aquí.* 'No one studies chemistry here.' (indefinite)
  c. *¿Quién estudia química?* 'Who studies chemistry?' (interrogative)
  d. *Ella estudia química.* 'She studies chemistry.' (personal)
  e. *La química es la disciplina que me interesa más.* 'Chemistry is the discipline that interests me the most.' (relative)
(4) Unexpressed subject: Ø *Estudia química. 'Ø (She/he) studies chemistry.'*
(5) Verbal phrase (VP): *Estudiar química merece la pena.* 'Studying chemistry is worth it.'

A limited body of variationist research has examined the multifunctionality of third-person subject expression, meaning that the scope of analysis was not restricted to unexpressed subjects and personal pronouns (see Blackwell & Quesada, 2012, and Lozano, 2016, for other theoretical approaches). The Spanish-speaking populations investigated include speakers living in New Mexico (Dumont, 2006), first-generation Mexican immigrants living in Georgia (Limerick, 2021), Spanish speakers of diverse origins living in the United States (Geeslin & Gudmestad, 2008; Gudmestad & Geeslin, 2021; Gudmestad et al., 2013; Linford, Long, Solon, Whatley, & Geeslin, 2016), and additional-language learners (Geeslin & Gudmestad, 2008; Gudmestad et al., 2013; Linford et al., 2016). This collection of studies examined oral production data. We highlight what we see as the four most important observations that have emerged from these studies. The first is that in contexts of third-person finite verbs, researchers have documented the use not only of LNPs but also of different types of pronouns, such as demonstrative and indefinite pronouns (Geeslin & Gudmestad, 2008; Gudmestad et al., 2013; Gudmestad & Geeslin, 2021). The second observation is that although the LNP is commonly used at first mention of a referent, it is not categorically restricted to that linguistic context (Dumont, 2006; Gudmestad & Geeslin, 2021; Limerick, 2021), which suggests that an LNP, like personal pronouns and unexpressed subjects, can be connected to various linguistic properties. Third, researchers have made different decisions about whether to include inanimate referents in their analysis (Dumont, 2006; Geeslin & Gudmestad, 2008; Gudmestad et al., 2021; Gudmestad & Geeslin, 2021) or to exclude them (Limerick, 2021; Linford et al., 2016).[2] Differences in decisions appear to be attributed to whether pronouns were used with inanimate referents in the respective datasets. For example, whereas Linford et al. (2016) found the referents of pronouns to be categorically animate, Gudmestad and Geeslin (2021) observed that antecedents of pronouns in their data were both animate and inanimate. Finally, among studies that examined unexpressed subjects, pronouns, and LNPs in one statistical model, with the goal of predicting when speakers are more or less likely to use a particular subject form, linguistic factors such as perseveration, specificity, number, referent continuity, and object pronoun, have been shown to be important (Gudmestad & Geeslin, 2021, Gudmestad et al., 2013; Linford et al., 2016). Collectively, this finding indicates that the same linguistic factors that have been examined in investigations of variable subject expression focused on animate referents and that the binary opposition of unexpressed subjects and personal pronouns are also useful for understanding the multifunctionality of third-person subject expression. We therefore continue to explore these factors in the current study.

2.3. Study abroad in SLA

Study abroad has garnered significant attention in SLA (see, for example, Collentine & Freed, 2004, and Sanz & Morales-Front, 2018) for being a valuable context to investigate additional-language learning because it affords learners opportunities to use and be exposed to the target language in a wider array of situations than solely in a language classroom (Geeslin, 2023). Despite the potential it provides for regular exposure and interaction in the target language, SLA research has not uniformly shown evidence of language development over the course of a stay abroad. One source of difference pertains to the components of language. For instance, studies on sociolinguistic competence tend to show evidence of development, whereas research on grammatical competence has exhibited varying results (Churchill & DuFon, 2006; Llanes, 2011).

[2] Geeslin and Gudmestad (2008) opted for a hybrid approach in which inanimate referents were examined in portions of their analysis.

Regarding variable subject expression in Spanish and study abroad, all existing research to our knowledge has demonstrated that learners' usage patterns change during their time abroad (Brown et al., 2025; Denbaum, 2020; Gudmestad & Edmonds, 2023, Linford, 2016, Linford, Zahler & Whatley, 2018)—either in the frequency of occurrence of forms or in the factors that influence usage. Research centered on targetlikeness found evidence of learners moving both toward and away from native-speaker norms (Denbaum, 2020; Linford, 2016; Linford et al., 2018). Gudmestad and Edmonds (2023) diverged from a focus on nativelikeness and instead examined learner language independently (Ortega, 2014). Thus, their observations about development were focused exclusively on the changes in subject expression that occurred while abroad, without making assessments about whether these changes constituted a move toward nativelikeness (see Grammon 2022, 2025 for concerns about nativeness in variationist SLA). In Gudmestad and Edmonds' study of first-person-singular subjects, they found that learners increased their use of personal pronouns during their academic year abroad. They also found that the way in which perseveration influenced use changed over time. Learners' inclination to use a personal pronoun immediately following the use of another personal pronoun grew stronger longitudinally.

Another reason for differing results about the impact of study abroad is that learners' individual experiences while abroad vary. For example, some study abroad participants remain closely connected, with the help of technology, to their life at home, which can reduce contact with the target language in the host country (Kinginger, 2012). These varied experiences can shape the effect that a stay abroad has on language development, and researchers have increasingly been examining the relationship between language development while abroad and individual characteristics of the learners. Research on subject expression in Spanish, for instance, has found that amount of contact with Spanish (Linford et al., 2018), amount of contact with English (Gudmestad & Edmonds, 2023), and the type of placement (i.e., student or worker, Gudmestad & Edmonds, 2023) influence subject expression while abroad. In the present investigation, we aim to shed further light on Spanish subject expression in a study abroad context; we consider both language development longitudinally and individual characteristics that may shape the experience.

3. The Current Study

We analyzed longitudinal production data spanning 21 months, during which additional-language learners of Spanish spent an academic year in Spain or Mexico. The goal of the present investigation was to examine the use of third-person subject forms and to identify the linguistic and extralinguistic factors that influence learners' variable use of these subject forms. Our three research questions are in (a) through (c).

(a) What third-person subject forms do learners use and with what frequency do they use third-person subject forms?
(b) What linguistic and extralinguistic factors influence the variable use of third-person subject forms?
(c) To what extent does the variable use of third-person subject forms change longitudinally?

3.1. Method

3.1.1. The corpus

The data came from LANGSNAP (http://langsnap.soton.ac.uk), a publicly available corpus that tracks learners of Spanish and French longitudinally. We focus on the additional-language Spanish

data and the first phase of the corpus (Mitchell, Tracy-Ventura, & McManus, 2017), in which the participants were in their third year of undergraduate study at a British university and spent the academic year in Mexico or Spain. We analyzed all additional-language Spanish participants in the corpus (N = 27). Twenty were women and seven were men. The average age of the 25 participants who provided this information was 20.76 years (SD = 1.45, *range* = 20-25). English was the first language for 25 participants; one participant indicated English and Polish as first languages; and another listed Polish as her first language. Twenty-six participants shared details on their experience with additional languages. They had studied Spanish for an average of 5.62 years (SD = 3.25, *range* = 2-14). Regarding other languages, participants had experience with French (n = 20), German (n = 8), Italian (n = 2), English (n = 1, the first-language Polish speaker), Latin (n = 1), and Portuguese (n = 1).[3] Three participants shared that they did not have experience with languages other than Spanish and English. While abroad, nine participants spent the school year in Mexico, while 19 were in Spain. They chose from three types of placement for their time abroad: exchange student (n = 9), teaching assistant (n = 16), or workplace intern (n = 2).

The corpus consists of six data-collection points listed in (6). For the current study, we analyzed the first, fourth, and six data-collection periods: pre-stay, in-stay 3 (henceforth, end-stay), and post-stay 2 (henceforth post-stay). Thus, we examine language use over a span of 21 months. At each data-collection phase, the participants completed three production tasks. The oral semi-guided interview consisted of questions about the participants' abroad experience and lasted about 20 minutes. For the oral narration, participants told a story that was guided by pictures. At pre-stay and end-stay, the story was about a missing cat. At post-stay, the narrative focused on two brothers. The written task was an argumentative essay, with the goal length of approximately 200 words. The prompt at pre-stay and end-stay was whether gay couples should be able to marry and adopt children. The topic for post-stay was whether soft drinks and junk food should be taxed at a higher rate in order to promote a healthier diet. We analyzed all production data at the three data-collection points under investigation. We also examined the elicited imitation (EI) task that participants completed at pre-stay. This instrument was employed as a proficiency measure (see Ortega, 2000). The learners listened to 30 sentences in turn and were asked to repeat them aloud. They received points for accuracy and completeness (120 maximum). The participants' average score was 85.15 (*range* = 59-108, SD = 11.73).

(6)

Pre-stay	May 2011
In-stay 1	October 2011
In-stay 2	February 2012
In-stay 3	May 2012
Post-stay 1	September 2012
Post-stay 2	February 2013

[3] Given the participants' varied language backgrounds, it was not possible to systematically explore the possible influence of other additional languages on the variable use of third-person subject expression in Spanish.

3.1.2. Data coding and analysis

In the current study, we examine third-person subject expression in contexts where the referent is identifiable. The dependent variable is the subject form used. Third-person subjects allow for various forms, as previously indicated in (2) through (5). We also coded for seven independent linguistic variables and four independent extralinguistic variables, which we examined as fixed effects in the analysis. Table 1 provides details on each. The linguistic factors we investigated were selected because they have been examined in previous subject-expression research, including studies on the multifunctionality of third-person subjects. Regarding specificity, because studies on variable subject expression have largely focused on animate referents, we follow Geeslin and Gudmestad (2008) and apply the categories of specific, non-specific, and group (i.e., a referent in which the group can be named but the individual entities in that group cannot be named) to animate referents only and classified all inanimate referents together as a fourth category of this variable. The extralinguistic variables under investigation enabled us to examine language development (time) and the potential role that certain individual characteristics may play in the use of third-person subject forms (task, EI score, placement type, and country). We also considered participant and verb in third-person subject contexts as random effects in the multinomial mixed effects model.

Table 1. Fixed effects

Fixed effects	Categories/description
Number	Singular or plural third-person subject context
Object pronoun	Presence or absence of a preverbal object pronoun
Clause type	Whether the subject is in an independent or dependent clause
Referent continuity	Whether the referent of the preceding tensed verb is the same or different
Specificity	Whether the referent is specific, non-specific, part of a group, or inanimate
Priming	Subject form of the referent in the previous clause
Perseveration	Subject form of the preceding mention of the same referent
Time	Pre-stay, end-stay, post-stay; measured numerically (Pasta, 2009)
Task	Interview, narration, essay
EI score	59-108; numeric
Placement type	Exchange student, teaching assistant, intern
Country	Spain, Mexico

We analyzed the data quantitatively in four steps, using R (R Core Team, 2021). First, we conducted crosstabulations to examine the aggregate frequency of subject forms. Second, we identified the average rate of use of subject forms at each data-collection period. Third, we began our analysis with a data-exploration step in which we explored the relationship between the dependent variable and each independent variable separately. We also checked for collinearity between independent variables and found no strong correlations. Fourth, we fit a mixed-effects model with a multinomial dependent variable using the mclogit package in R (Elff, 2022), following closely the statistical procedures put forth in Gudmestad and Metzger (2025). We chose the unexpressed subject as the referent point of the dependent variable because it was the most frequent form. Regarding our model selection procedure, we began with a full exploratory model

comprising all main effects of interest and removed effects without significance in either the dependent-variable comparisons (LNP versus unexpressed subject and pronoun versus unexpressed subject). We then conducted best subset selection with the remaining main effects, considering 511 candidate models and selecting the model with the lowest Bayesian Information Criterion (BIC).[4] During this process, we also examined interactions between time and other fixed effects in order to make observations about possible language development over the 21-month period. All models under consideration had random intercepts for both participant and verb (see the Appendix for the two best and two worst models based on the BIC).[5]

3.2. Results

We begin our reporting of the results with the inventory of third-person subject forms that learners used. Table 2 identifies eight types of forms. The most frequent form was the unexpressed subject, which occurred 46.1 percent of the time. The second most frequent form was the LNP, which participants used a third of the time. Learners also produced five types of pronouns; collectively, they constituted 15.3 percent of the data. They also used VPs 8.7 percent of the time.

Table 2. Frequency of third-person subject forms

Form	#	%
Unexpressed subject	4320	46.1
Personal pronoun	193	2.1
Demonstrative pronoun	224	2.4
Indefinite pronoun	185	2.0
Interrogative pronoun	56	0.6
Relative pronoun	768	8.2
LNP	2817	30.0
VP	816	8.7
Total	9379	100

Next, we turn to an assessment of frequency of use over time. Table 3 provides learners' average rate of use of subject forms at each data-collection period. Given the relatively low frequency of each pronoun form, we group all five pronoun types together for the remainder of the analysis. The use of unexpressed subjects appears to be stable over time, with a slight decrease (2%) between pre-stay and end-stay, on the one hand, and post-stay, on the other hand. Learners' use of pronouns increased while abroad (6.1%) and, for the most part, this increase was maintained after

[4] The BIC is a likelihood-based quantity that assesses the suitability of a model relative to a set of candidate models based on fit and complexity (Burnham, 2004). In a previous study, Gudmestad and Metzger (2025) explored 93 models under consideration. However, in the current study with a much larger pool of candidate fixed effects, there are 511 fixed-effect structures under consideration. Given this larger risk of overfitting, we have opted to use the BIC rather than the Akaike Information Criterion (AIC) because the BIC adds a higher penalty to more complex models (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002). We note, though, that the AIC remains a mathematically and statistically defensible choice as well, and recommendations in this context are not unanimous (Bolker, Brooks, Clark, Geange, Poulsen, Stevens, & White, 2009).

[5] Although Barr, Levy, Scheepers, and Tily (2013) recommended considering a maximal random effect structure, a caveat is that models must exhibit numeric stability to be considered, and we found that, with our data, many candidate models with maximal random effect structures (i.e., random slopes for task within participant) did not converge to solutions.

returning home (decrease of 1.4%). Although there was a slight decrease in the use of LNPs while abroad (2.3%), the rate of LNP use at pre-stay and post-stay was similar (0.4% difference). Finally, between pre-stay and end-stay, learners decreased their use of VPs by 3.3 percent, which was followed by a slight increase at post-stay (1.1% increase), and over the 21-month period, the use of the VP decreased 2.2 percent. Together, these findings do not appear to indicate dramatic change over time, as the largest change in frequency of any form between data-collection periods was about six percent (i.e., pronouns between pre-stay and end-stay).

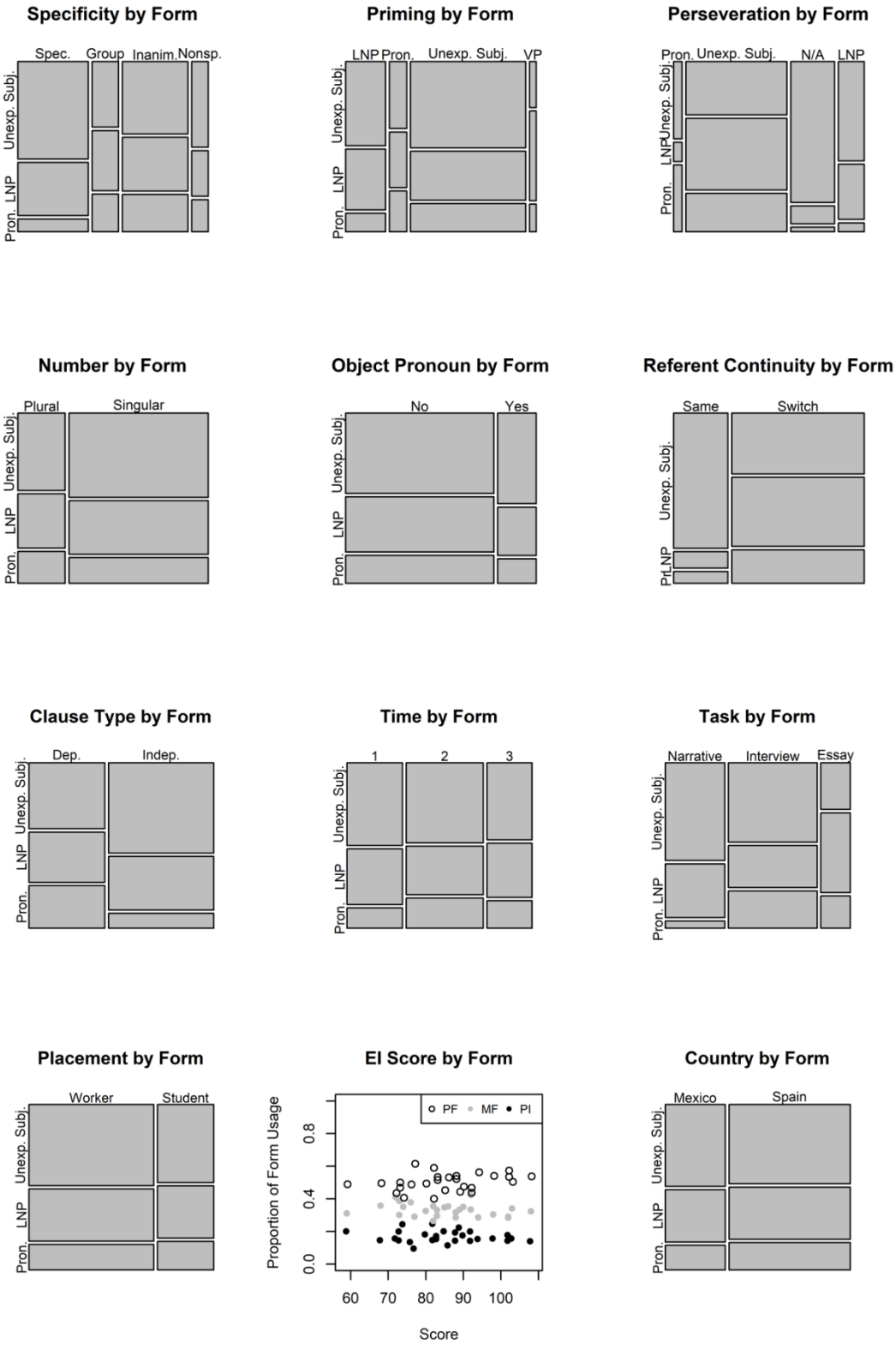
Table 3. Frequency of third-person subject forms over time

Form	Unexpressed subject			Pronoun			LNP			VP		
Time	Pre-stay	End-stay	Post-stay	Pre-stay	End-stay	Post-stay	Pre-stay	End-stay	Post-stay	Pre-stay	End-stay	Post-stay
M	46.5	46.3	44.3	11.4	17.5	16.1	31.5	28.8	31.1	10.7	7.4	8.5
SD	7.5	7.0	8.1	5.8	4.3	4.8	5.6	5.6	5.1	3.9	3.0	3.2
Min.	32.5	32.4	27.5	1.7	11.3	6.9	21.2	21.4	20.5	4.8	2.3	3.6
Max.	62.4	59.3	62.1	27.1	27.6	29.3	44.2	41.7	39.7	22.2	13.5	14.5

For the third step in the analysis, we present a data exploration in which we analyze the dependent variable (subject form) in relation to each fixed effect separately. Because VPs constituted less than 10 percent of the data, we removed them from subsequent examination and focus the remainder of the analysis on unexpressed subjects, LNPs, and pronouns (k = 8,563).[6] The mosaic and strip plots in Figure 1 suggest, for example, that the distribution of subject forms across categories of the number variable appear to be similar, whereas differences between categories of referent continuity are more noticeable. In the final step of the analysis, we verify the role that these independent variables play in predicting third-person subject expression by examining them jointly in one statistical model.

[6] Future research that analyzes different types of pronouns separately and that explores VPs in more detail may provide further insights into learners' variable use of third-person subjects in Spanish.

Figure 1. Data visualization of independent variables



Finally, we present the mixed-effects multinomial model. As a reminder, we selected the unexpressed subject as the reference point of the dependent variable because it was the most frequent form. Each nominal independent variable also has a reference point (see Tables 4 and 5). Four linguistic factors (referent continuity, clause type, specificity, and perseveration) and one extralinguistic factor (task) were included in the final model. Object pronoun, number, priming, time, EI score, placement type, and country did not predict use of third-person subjects. We moreover did not find any significant interactions between time and the fixed effects included in the final model. It will be recalled that the single model consists of two dependent-variable comparisons: LNP versus unexpressed subject (Table 4) and pronouns versus unexpressed subject (Table 5). Table 6 provides the statistical details of the random intercepts. We note that the nature of this model means there are numerous details; we discuss their relevance in the Discussion section.

Table 4. Fixed effects results: LNP versus unexpressed subject[7]

Effect	Estimate	*SE*	z-value	p-value	Coeffic. CI	*OR*	*OR* CI	Effect size
(Intercept)	-1.48	0.12	-12.42	<0.0001	[-1.71, -1.24]	0.23	[0.18, .29]	-0.82
Referent continuity [Same]								
Switch	1.90	0.09	21.10	<0.0001	[1.72, 2.07]	6.68	[5.60, 7.96]	1.05
Clause type [Dependent]								
Independent	0.14	0.06	2.21	0.027	[0.02, 0.26]	1.15	[1.02, 1.30]	0.08
Specificity [Specific]								
Non-spec.	-1.31	0.14	-9.62	<0.0001	[-1.57, -1.04]	0.27	[0.21, 0.35]	-0.72
Group	-0.14	0.12	-1.12	0.261	[-0.37, 0.10]	0.87	[0.69, 1.11]	-0.08
Inanimate	-0.60	0.11	-5.49	<0.0001	[-0.81, -0.38]	0.55	[0.44, 0.68]	-0.33
Perseveration [LNP]								
Unex. subj.	-1.34	0.10	-12.87	<0.0001	[-1.54, -1.14]	0.26	[0.21, 0.32]	-0.74
Pronouns	-0.43	0.19	-2.29	0.022	[-0.80, -0.06]	0.65	[0.45, 0.94]	-0.24
Not. ap.	0.56	0.09	6.24	<0.0001	[0.38, 0.73]	1.74	[1.46, 2.08]	0.31
Task [narrative]								
Interview	-0.58	0.11	-5.48	<0.0001	[-0.79, -0.37]	0.56	[0.45, 0.69]	-0.32
Essay	0.88	0.13	6.66	<0.0001	[0.62, 1.14]	2.42	[1.86, 3.13]	0.49

[7] *Note.* SE is standard error. CI is Confidence Interval. OR is Odds Ratio. The not applicable category of the perseveration variable refers to when the referent was the first mention or could not be identified.

Table 5. Fixed effects results: Pronoun vs unexpressed subject

Effect	Estimate	*SE*	z-value	p-value	Coeffic. CI	*OR*	*OR* CI	Effect size
(Intercept)	-2.78	0.18	-15.17	<0.0001	[-3.14, -2.42]	0.06	[0.04, 0.09]	-1.54
Referent continuity [Same]								
Switch	0.72	0.12	6.02	<0.0001	[0.49, 0.96]	2.06	[1.63, 2.61]	0.40
Clause type [Dependent]								
Independent	-0.89	0.08	-11.75	<0.0001	[-1.04, -0.74]	0.41	[0.35, 0.48]	-0.49
Specificity [Specific]								
Non-spec.	-0.73	0.15	-4.71	<0.0001	[-1.03, -0.43]	0.48	[0.36, 0.65]	-0.40
Group	0.11	0.14	0.77	0.439	[-0.16, 0.38]	1.11	[0.85, 1.46]	0.06
Inanimate	0.05	0.13	0.38	0.707	[-0.20, 0.30]	1.05	[0.82, 1.35]	0.03
Perseveration [LNP]								
Unex. subj.	-0.82	0.19	-4.35	<0.0001	[-1.19,-0.45]	0.44	[0.30, 0.64]	-0.45
Pronouns	1.95	0.18	10.88	<0.0001	[1.60, 2.30]	7.04	[4.95, 10.01]	1.08
Not. ap.	1.60	0.15	10.75	<0.0001	[1.31,1.89]	4.95	[3.70, 6.63]	0.88
Task [narrative]								
Interview	0.77	0.14	5.38	<0.0001	[0.49, 1.05]	2.16	[1.63, 2.86]	0.43
Essay	1.22	0.17	7.11	<0.0001	[0.89, 1.56]	3.40	[2.43, 4.76]	0.68

Table 6. Covariance matrices for the random intercepts

Participant				
	Estimate		*SE*	
	LNP vs. unex. subj.	Pronoun vs. unex. subj.	LNP vs. unex. subj.	Pronoun vs. unex. subj.
LNP vs. unex. subj.	0.0176	0.0095	2.387×10^{-6}	5.725×10^{-6}
Pronoun vs. unex. subj.	0.0095	0.0418	5.725×10^{-6}	1.585×10^{-5}
Verb				
	Estimate		*SE*	
	LNP vs. unex. subj.	Pronoun vs. unex. subj.	LNP vs. unex. subj.	Pronoun vs. unex. subj.
LNP vs. unex. subj.	0.3010	-0.0003	0.0014	0.0009
Pronoun vs. unex. subj.	-0.0002	0.2866	0.00090	0.0012

We begin with the comparison between LNPs and unexpressed subjects and the two linguistic variables that have two categories: referent continuity and clause type. We see that the log-odds of using an LNP versus unexpressed subject are greater in contexts of switch reference compared to same reference (the referent continuity variable, as indicated by the positive estimate of 1.90 in Table 4). We focus on the coefficient estimates in this paper to align with previous research on additional-language subject expression, but we note that the ORs can also be useful for interpretation. For LNP versus unexpressed subjects and referent continuity, for example, the odds of using of an LNP (compared to an unexpressed subject) in a switch-reference context are 6.68 times those in a same-reference context. For clause type, learners were more probable to use an LNP in a main clause, versus a subordinate clause (again, as shown by the positive estimate of 0.14).

Fixed effects with more than two categories require additional analysis (specificity, perseveration, and task). The mixed-effects model shows whether non-reference points are similar to or different from the reference point but does not provide statistical details on comparisons between non-reference point categories. For the latter task, we performed pairwise comparisons, which are a post-hoc analysis that compares all non-reference-point levels of a given categorical variable, two at a time, against one another (Wu & Hamada, 2011). We build Bonferroni-Holm adjusted CIs for the differences in effects (see Figures 2 and 3 with the CI plots) and observe whether they are significantly different from 0 (Holm, 1979). Point estimates for the differences, their SEs, and the resultant z-scores and *p*-values are provided in Table 7 and 8.[8]

Continuing with the LNP and unexpressed subject comparison in the mixed-effects model, we now turn to the specificity variable. When the reference point of specific is compared to the other categories (Table 4), we see that the log-odds of using an LNP are significantly lower when the referent is non-specific or inanimate. The difference between group and specific referents was not significant. For the non-reference-point categories (Figure 2 and Table 7), we see that all pairwise

[8] We provide these additional results beyond the scope of Gudmestad and Metzger (2025) to provide more transparency in the calculations and because, with a greater number of categorical effects which in turn have a greater number of levels in the current study, the post-hoc pairwise tests are more numerous than in Gudmestad and Metzger (2025).

comparisons are significant. The log-odds of using an LNP are higher when the referent is a group compared to both inanimate and non-specific referents, and when the referent is inanimate versus non-specific.

Regarding perseveration, the reference point was LNP (Table 4). We found that learners were more probable to use an LNP versus an unexpressed subject when the subject form of the preceding mention of the same referent was not applicable, meaning that it was the first mention of the referent or the referent could not be identified. In contrast, the log-odds of using an LNP were significantly lower when the subject form of the preceding mention of the same referent was a pronoun or an unexpressed subject. Similar to specificity, all pairwise non-reference-point comparisons were significant (Figure 2 and Table 7). Learners were more likely to use an LNP when the previous mention of the referent was not applicable. The log-odds of LNP use were also higher when the previous mention of the referent was a pronoun, compared to an unexpressed subject.

The final significant factor was task, for which we found that the log-odds of using an LNP were significantly higher in the essay and lower in the interview, compared to the narration task (Table 4). Learners were also less likely to use an LNP in the interview, versus the essay (Figure 2 and Table 7).

Figure 2. Pairwise difference Holm-Bonferroni adjusted CIs: LNP versus unexpressed subject

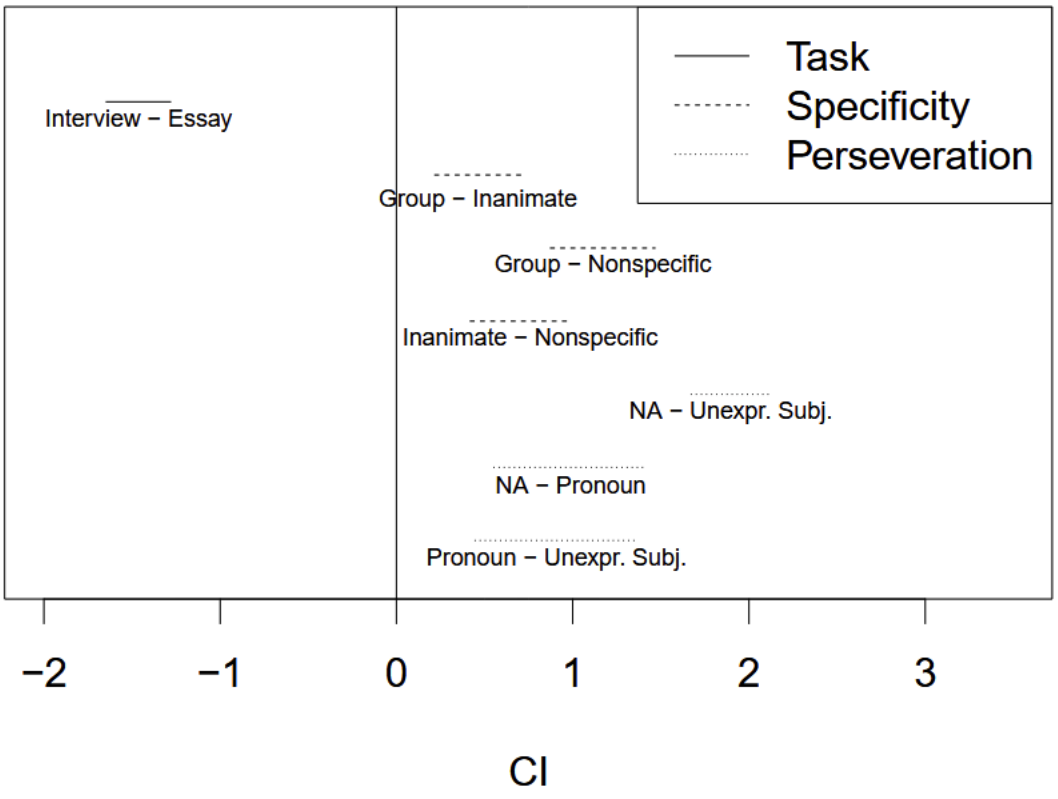


Table 7. Pairwise differences: LNP versus unexpressed subject

Variable	Comparison	Estimate	*SE*	z-value	*p*-value
Specificity	Group vs. inan.	0.46	0.10	4.52	<0.0001
	Group vs. non-spec.	1.17	0.12	9.46	<0.0001
	Inan. vs. non-spec.	0.71	0.12	5.88	<0.0001
Perseveration	N/A vs unex. subj.	1.90	0.094	20.25	<0.0001
	N/A vs. pronoun	0.99	0.18	5.44	<0.0001
	Pronoun vs. unexp. subj.	0.91	0.19	4.71	<0.0001
Task	Interview vs. essay	-1.46	0.097	-15.7	<.0001

The other comparison in the model focused on pronouns versus unexpressed subjects as dependent variables. (Tables 5 and 8 and Figure 3). We found that, for referent continuity, learners were more likely to use a pronoun in contexts of switch, compared to same, reference. The clause type variable showed that the log-odds of using a pronoun were lower in independent clauses versus dependent clauses.

Regarding specificity (Table 5), we observed that learners were less likely to use pronouns (versus unexpressed subjects) when the referent was non-specific (versus specific). The differences between both group and inanimate referents versus specific referents were not significant. Table 8 and Figure 3 illustrate that learners were more likely to use a pronoun when the referent was group or inanimate (versus non-specific). The difference between group and inanimate referents was not significant.

For perseveration, the log-odds of using a pronoun were higher when the previous mention of the same referent was also a pronoun and when the referent was mentioned for the first time or could not be identified (i.e., the not applicable category). Learners were less likely to use a pronoun when the previous mention of the same referent was an unexpressed subject (Table 5). In Table 8 and Figure 3 we see that learners were more likely to use a third-person pronoun, versus an unexpressed subject, when the previous mention of the referent was the first mention or could not be identified, compared to an unexpressed subject. In addition, the log-odds of using a pronoun were higher when the previous mention was a pronoun compared to an unexpressed subject. In contrast, learners were less likely to use a pronoun when the referent was the first mention or could not be identified versus a pronoun.

Finally, the model indicated that the log-odds of using a pronoun were greater in the interview and essay, versus the narrative task (Table 5). Learners were also less likely to use a pronoun in the interview compared to the essay (Table 8 and Figure 3). Collectively, the numerous results from this multinomial mixed-effects model suggest that referent continuity, clause type, specificity, perseveration, and task are effective at identifying many ways in which learners vary their use of third-person subject forms in language use.

Figure 3. Pairwise difference Holm-Bonferroni adjusted CIs: Pronoun versus unexpressed subject

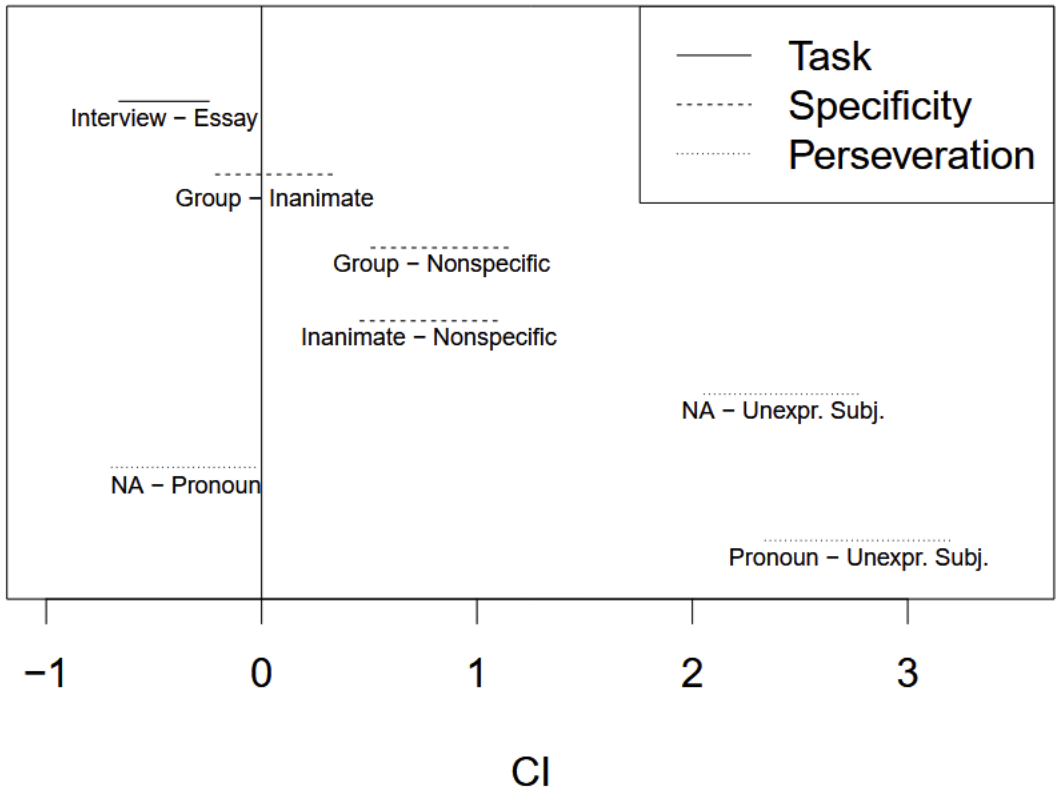


Table 8. Pairwise differences: Pronoun versus unexpressed subject

Variable	Comparison	Estimate	*SE*	z-value	*p*-value
Specificity	Group vs. inan.	0.059	0.11	0.52	0.60
	Group vs. non-spec.	0.84	0.14	6.13	<0.0001
	Inan. vs. non-spec.	0.78	0.13	5.84	<0.0001
Perseveration	N/A vs unex. subj.	2.42	0.15	15.85	<0.0001
	N/A vs. pronoun	-0.35	0.14	-2.44	0.015
	Pronoun vs. unexp. subj.	2.77	0.18	15.27	<0.0001
Task	Interview vs. essay	-0.453	0.11	-4.24	<0.0001

4. Discussion

We now reflect on the current study's findings by answering the research questions and connecting our work with previous investigations. In our discussion, we follow the

recommendations of Ortega (2014, inter alia) and investigate learner language and development independently from native-speaker norms and thus limit our connections to previous investigations of learner language and avoid making assessments of nativelikeness or targetlikeness.

The first question pertained to learners' inventory of third-person subject forms and the frequency with which they used them. The LANGSNAP participants' repertoire of third-person subjects was multifunctional; they used unexpressed subjects, pronouns (personal, demonstrative, indefinite, interrogative, and relative), LNPs, and VPs. The two most frequent forms were the unexpressed subject and the LNP (46% and 30% of third-person contexts, respectively). The remaining six forms (five types of pronouns and VPs) each constituted less than 10 percent of the data, though collectively the learners used pronouns 15 percent of the time. These findings align with other research on subject expression in Spanish in a study-abroad context that has shown that unexpressed subjects are the dominant form among learners (Brown et al., 2025; Gudmestad & Edmonds, 2023; Linford, 2016). For third-person subjects and advanced-level learners living in the United States, Geeslin and Gudmestad (2008) found a similar hierarchy of use as the present investigation: Unexpressed subjects were most frequent; LNPs were used about a third of the time; and pronouns were the least frequent group.[9] The relatively high frequency of LNPs in additional-language use also demonstrate that in order to fully understand how learners develop the ability to use third-person subjects, LNPs must be part of the scope of inquiry. In other words, research that is restricted to unexpressed subjects and personal pronouns misses an important form at learners' disposal and therefore is limited in how it can advance knowledge about the additional-language subject expression in Spanish.

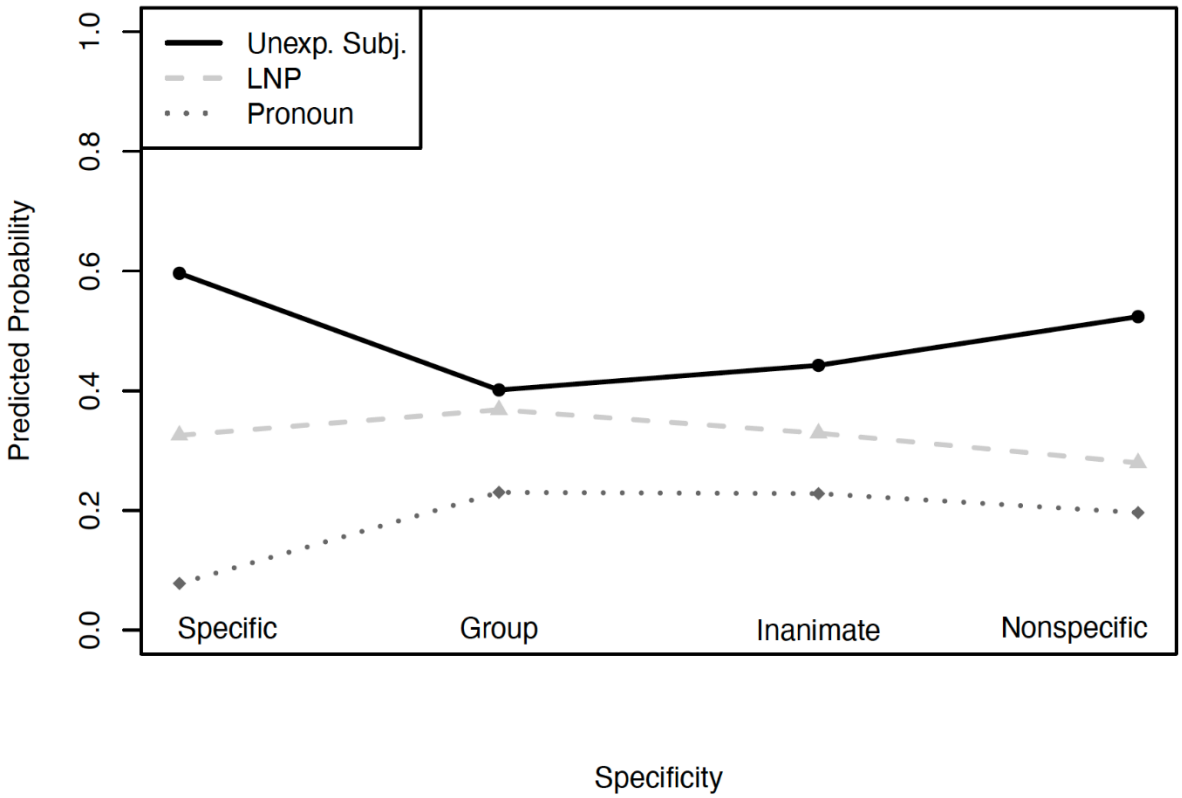
The second question pertained to the linguistic and extralinguistic factors that predicted the variable use of third-person subjects. For this portion of the analysis, we focused on unexpressed subjects, pronouns, and LNPs, and we explored the role that linguistic factors shown to be important in previous research and extralinguistic factors that can shed light on individual characteristics may play on learners' variable use of third-person subject forms.[10] We found five factors to be influential: referent continuity, clause type, specificity, perseveration, task. One important observation is that not only did learners produce a range of forms to express third-person subjects but they also used these forms in systematic and nuanced ways (i.e., their use cannot be attributed to just one factor). We discuss each independent variable in relation to other additional-language scholarship on subject expression. Our results align with prior studies on first-person-singular and third-person subject expression and referent continuity, which has shown that unexpressed subjects are more likely when there is no change in referent from that of the preceding tensed verb (e.g., Brown et al., 2025; Gudmestad & Edmonds, 2023; Linford et al., 2016). Regarding clause type, Gudmestad and Edmonds (2023) found that the LANGSNAP learners were more likely to use *yo* in an independent clause. While this result might suggest a connection between overt subject forms and main clauses, the current study further nuances the role that clause type plays, as LNPs were more likely and pronouns were less likely to be used in independent clauses (compared to unexpressed subjects).

The third significant independent variable was specificity; the mixed-effects model and pairwise comparisons in the Results section revealed that specificity influenced the use of third-person subjects in various ways. To visualize this effect more easily, Figure 4 illustrates the predicted probabilities for the four subject forms and each category of specificity. In terms of third-person subjects and previous research, the important difference that Linford et al. (2016) found

[9] Geeslin and Gudmestad (2008) did not code for VPs.
[10] We leave the extralinguistic factor of time to the third research question.

was that participants were more likely to use pronouns (compared to unexpressed subjects) when the referent was specific. Our findings differ in that the LANGSNAP participants exhibited the lowest probability of pronoun use with specific referents. We highlight four other findings related to specificity. One is that although unexpressed subjects were the most likely form overall (solid line), their highest probability of use was with specific and non-specific referents. Second, we see that LNPs (dashed line) showed a similar patterning with specific, group, and inanimate referents and their predicted probability was lowest with non-specific referents. Third, learners were more likely to use pronouns (dotted line) with group, inanimate, and non-specific referents, compared to specific referents. Lastly, when we compare the four categories of specificity, we see that specific referents are those that are connected to more divergent patterning. In other words, the tendency toward unexpressed subjects is strong and pronoun use is low. While a similar pattern is seen for non-specific referents, it is less divergent, and we observe that for group and inanimate referents, patterning is more similar across the subject forms, with only a slight preference for unexpressed subjects. Additional research is needed in order to better understand whether the patterns observed in the LANGSNAP data are consistent with other learner populations.

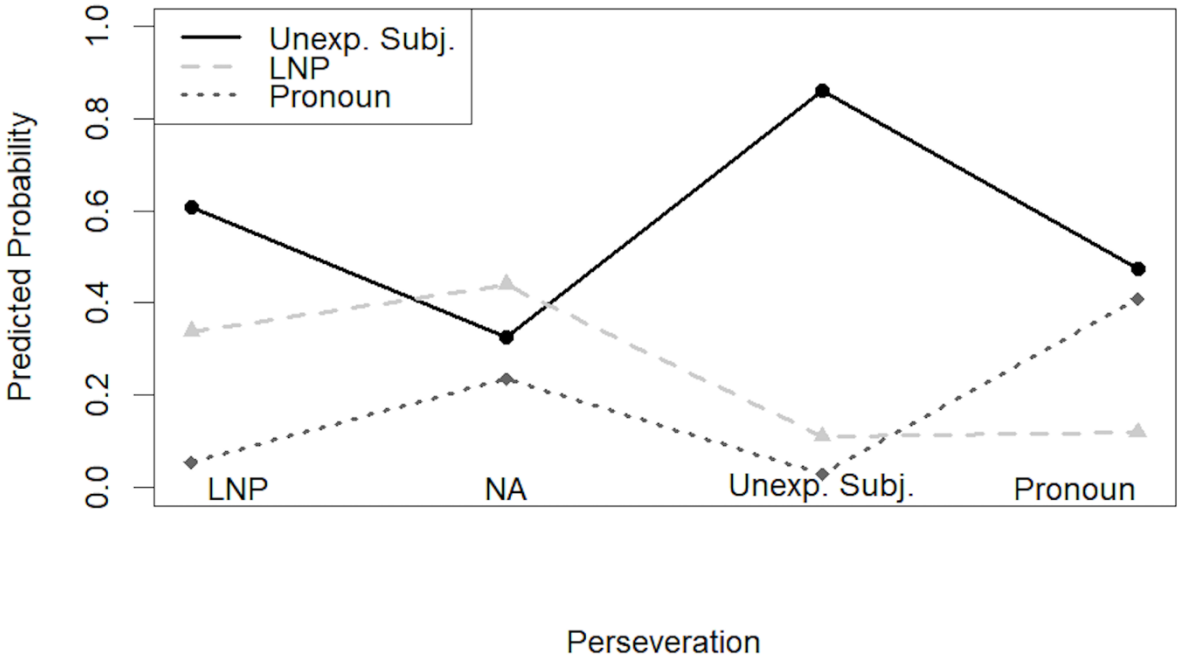
Figure 4. Average predicted subject-form use according to specificity



The fourth significant linguistic factor was perseveration. As a reminder, this variable examines the subject form of the preceding mention of the same referent, with the hypothesis being that learners will be more likely to produce the same form that they used in the previous mention of the same referent (i.e., the "birds of a feather" effect, Pereira Scherre & Naro, 1991). The numerous statistical details presented in the Results section suggest that the perseveration does not have a

uniform influence on all subject forms. As we did for specificity, we visualize the predicted probabilities for the four subject forms and each category of perseveration in order to more easily see patterns of use (Figure 5). We see that, while it is true that learners showed a strong tendency to use an unexpressed subject when the previous mention of the referent is also an unexpressed subject (i.e., evidence in support of the hypothesis), unexpressed subjects are also more likely to be used when the previous mention is an LNP or pronoun, which does not align with the perseveration assumption. Figure 5 moreover demonstrates that the LNP is the most probable form when the previous mention of the same referent was not identifiable or when the referent was not mentioned previously (i.e., the not applicable category). This finding aligns with previous research that has shown that LNPs occur most often in contexts of first mention (Dumont, 2006). More generally, although Gudmestad and Edmonds (2023) found evidence consistent with the "birds of a feather" phenomenon in the use of first-person-singular subjects, Gudmestad et al. (2013)'s results for third-person subjects aligned with those in the current study in that they did not uniformly support the hypothesis. Thus, it may be that multifunctional linguistic phenomena in additional languages, such as third-person subjects, are less susceptible to the tendency to repeat forms, compared to binary linguistic structures, such as first-person-singular subjects.

Figure 5. Average predicted subject-form use according to perseveration



Lastly, although learners' third-person subject expression did not vary according to the individual characteristic of EI score, placement type, and country of abroad stay, the extralinguistic factor of task was influential, which aligns with variationist SLA research that has shown ample evidence of task as a source of learner variation (e.g., Geeslin, 2006; Tarone & Parrish, 1988). In this regard, the current study has extended the investigation of task variation to third-person subject expression in Spanish and has demonstrated that there are important differences between both LNPs and pronouns (versus unexpressed subjects) and each of the three tasks. In order to understand whether this variation can be attributed to features of the task (e.g., modality) or to the

linguistic contexts that emerge from these tasks (see Geeslin, 2006; Tarone & Parrish, 1988), future research is needed.

The final research question addressed the possibility of language development over the 21-month period, which included an academic year abroad. No strong evidence of language development in the variable use of third-person subjects was observed. An examination of the frequency of use of subject forms across the three data-collection points revealed that the largest rate of change over the 21-month period (i.e., between pre-stay and post-stay) occurred with pronouns (4.7% change). The fluctuations in rate of use for unexpressed subjects, LNPs, and VPs were 3.3 percent or less. Additional evidence for the stability of third-person subject forms among this group of learners is seen in the mixed-effects multinomial model. First, our model selection procedure did not identify time as important in the model. Second, time did not significantly interact with any of the fixed effects included in the final model. Although previous research on study abroad has not uniformly shown that learners' linguistic systems develop during a sojourn in a target-language community (Churchill & DuFon, 2006; Llanes, 2011), research on variable structures tends to show evidence of language change (Zahler, Long, & Linford, 2023). The latter raises the question as to why third-person subject expression remained stable for these learners. We suggest that the reason may be due to the multifunctionality of third-person subject expression. Previous research on first-person-singular subject expression and the LANGSNAP corpus support this observation. The findings in Gudmestad and Edmonds (2023) and Brown et al. (2025) demonstrated that learners changed the ways in which they used first-person singular subject expression between pre-stay and end-stay (Gudmestad & Edmonds, 2023) and between pre-stay and post-stay (Brown et al., 2025). With the focus on first-person-singular subject expression, both studies examined variation between two forms: unexpressed subjects and personal pronouns. In the present investigation, we found that learners used unexpressed subjects, various types of pronouns, LNPs, and VPs to express third-person subjects. It may be that the multiplicity of forms, in conjunction with the diverse ways in which they are used (as evidenced by the several factors that predict their use), constitutes a greater acquisitional challenge than a variable structure that consists of just two forms. Even though a sojourn abroad can provide learners with rich sources of input, it appears that the experiences that the LANGSNAP participants had over the 21-month period under investigation were insufficient to promote changes in how they used third-person subject expression.

5. Conclusion

The current study constitutes the first longitudinal investigation of the additional-language use of third-person subject expression in Spanish. Our analysis demonstrated that by the first data-collection period in the LANGSNAP corpus, learners had already attended to the multifunctional nature of third-person subjects in the input and incorporated a range of forms (i.e., unexpressed subjects, pronouns, LNPs, and VPs) into their own language use. Additionally, adopting a variationist SLA approach enabled us to make two important observations about learners' use of third-person subjects. The first observation is that learners varied their use of unexpressed subjects, LNPs, and pronouns in intricate ways. Linguistic (referent continuity, clause type, specificity, and perseveration) and extralinguistic (task) factors worked together to influence variable use of third-person subjects. The second notable finding that emerged from the variationist analysis was that the LANGSNAP participants' use of third-person subjects was stable during and after their sojourn in Spain and Mexico. In order to learn more about learners' development of multifunctionality in third-person subject expression (e.g., when certain forms emerge) and to better understand how

learners develop (i.e., change) the ability to systematically vary their use of third-person subjects, future longitudinal research that examines learners at earlier points along the developmental trajectory is needed. Altogether, we hope to have contributed to the growing body of research on third-person subject expression in Spanish by demonstrating the valuable insights that emerge when the analysis includes a wider inventory of forms than animate unexpressed subjects and personal pronouns and by shedding light on the complexity and stability that is present in learners' variable usage patterns of third-person subjects.

Aarnes Gudmestad
Virginia Polytechnic Institute and State University
Department of Modern & Classical Languages & Literatures
323 Major Williams Hall
220 Stanger St.
Blacksburg, VA 24061
USA
ORCiD: 0000-0003-2303-7802

Thomas A. Metzger
The Ohio State University
Department of Statistics
1958 Neil Avenue, Room 340E
Columbus, Ohio 43210-1247
USA
ORCiD: 0000-0003-3620-1405

References

Abreu, L. (2009). *Spanish subject personal pronoun use by monolinguals, bilinguals, and second language learners*. Doctoral dissertation, University of Florida.

Andersen, R. W. (1990). Models, processes, principles and strategies: Second language acquisition inside and outside the classroom. In B. VanPatten & J. F. Lee (eds.), *Second language acquisition—Foreign language learning*, pp. 45–78. Clevedon, UK: Multilingual Matters.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), pp. 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bayley, R., & Pease-Álvarez, L. (1997). Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change*, 9, pp. 349–371. https://doi.org/10.1017/S0954394500001964

Blackwell, S. E., & Quesada, M. L. (2012). Third-person subjects in native speakers' and L2 learners' narratives: Testing (and revising) the Givenness Hierarchy for Spanish. In K. Geeslin & M. Díaz-Campos (eds.), *Selected proceedings of the 14th Hispanic Linguistics Symposium*, pp. 142-164. Malden, MA: Cascadilla Proceedings Project.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, 24(3), pp. 127-135. https://doi.org/10.1016/j.tree.2008.10.008

Brown, E., Quan, T., & Rivas, J. (2025). Testing cumulative lexicalized effects in study abroad: Variable subject pronoun expression in Spanish as an additional language. *Languages*, 10(5), 110. https://doi.org/10.3390/languages10050110

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, 33(2), pp. 261-304. https://doi.org/10.1177/0049124104268644

Cameron, R. (1993). Ambiguous agreement, functional compensation, and non-specific tú in the Spanish of San Juan, Puerto Rico and Madrid, Spain. *Language Variation and Change*, 5, pp. 305-334. https://doi.org/10.1017/S0954394500001526

Cameron, R. (1995). The scope and limits of switch reference as a constraint on pronominal subject expression. *Hispanic Linguistics*, 6-7, pp. 1-27.

Cameron, R., & Flores-Ferrán, N. (2004). Perseveration of subject expression across regional dialects of Spanish. *Spanish in Co*ntext, 1(1), pp. 41-65. https://doi.org/10.1075/sic.1.1.05cam

Carvalho, A. M., Orozco, R., Shin, N. L. (eds.). (2015). *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Washington, DC: Georgetown University Press.

Collentine, J., & Freed, B. F. (2004). Learning context and its effects on second language acquisition: An introduction. *Studies in Second Language Acquisition*, 26, pp. 153-171. https://doi.org/10.1017/S0272263104262015

Churchill, E., & DuFon, M. A. (2006). Evolving threads in study abroad research. In M. A. DuFon & E. Churchill (eds.), *Language learners in study abroad contexts*, pp. 1-27. Clevedon, UK: Multilingual Matters. https://doi.org/10.2307/jj.27195488.7

de Prada Pérez, A. (2018). Subject pronoun expression and language mode in bilingual Spanish. *Studies in Hispanic and Lusophone Linguistics*, 11, pp. 303-336. https://doi.org/10.1515/shll-2018-0010

Denbaum, N. (2020). Role of social interaction abroad in the L2 acquisition of sociolinguistic variation: The case of subject expression in the Dominican Republic. In D. Pascual y Cabo & I. Elola (eds.), *Current theoretical and applied perspectives on Hispanic and Lusophone linguistics*, pp. 63-84. Amsterdam: John Benjamins. https://doi.org/10.1075/ihll.27.03den

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

Elff, M. (2022). *mclogit: Multinomial logit models, with or without random effects or overdispersion. R package version 0.9.6,* https://CRAN.R-project.org/package=mclogit.

Flores-Ferrán, N. (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, pp. 49–73. https://doi.org/10.1017/S0954394504161048

Geeslin, K. L. (2006). Task design, discourse context and variation in second language data elicitation. In C. A. Klee & T. L. Face (eds.), *Selected proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, pp. 74-85. Malden, MA: Cascadilla Proceedings Project.

Geeslin, K. L. (2023). Epilogue. In S. L. Zahler, A. Y. Long, & B. Linford (Eds.). *Study abroad and the second language acquisition of sociolinguistic variation in Spanish*, pp. 350-368. Amsterdam: John Benjamins. https://doi.org/10.1075/ihll.37.12gee

Geeslin, K. L., & Gudmestad, A. (2008). Variable subject expression in second-language Spanish: A comparison of native and non-native speakers. In M. Bowles, R. Foote, & S. Perpiñán

(eds.), *Selected proceedings of the 2007 Second Language Research Forum*, pp. 69-85. Malden, MA: Cascadilla Proceedings Project. https://doi.org/10.1075/sic.13.1.03gee

Geeslin, K. L., & 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), pp. 53-79. https://doi.org/10.1075/sic.13.1.03gee

Geeslin, K., Linford, B., & Fafulas, S. (2015). Variable subject expression in second language Spanish: Uncovering the developmental sequence and predictive linguistic factors. In A. M. Carvahlo, R. Orozco, N. L. Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective*, pp. 191-209. Washington, DC: Georgetown University Press.

Geeslin, K. L. with Long, A. Y. (2014). *Sociolinguistics and second language acquisition: Learning to use language in context*. New York, NY: Routledge. https://doi.org/10.4324/9780203117835

Grammon, D. (2022). Es un mal castellano cuando decimos 'su': Language instruction, raciolinguistics ideologies and study abroad in Peru. *Linguistics and Education*, 71, pp. 1-10. https://doi.org/10.1016/j.linged.2022.101078

Grammon, D. (2025). Inappropriate identities: Racialized language ideologies and sociolinguistic competence in a study abroad context. *Applied Linguistics*, 46, pp. 193-210. https://doi.org/10.1093/applin/amae003

Gudmestad, A., & Edmonds, A. (2023). The variable use of first-person-singular subject forms during an academic year abroad. In S. L. Zahler, A. Y. Long, & B. Linford (eds.), *Study abroad and the second language acquisition of sociolinguistic variation in Spanish*, pp. 266-290. Amsterdam: John Benjamins. https://doi.org/10.1075/ihll.37.09gud

Gudmestad, A., & Geeslin, K. L. (2021). Overlapping envelopes of variation: The case of lexical noun phrases and subject expression in Spanish. In M. Díaz-Campos (ed.), *The Routledge handbook of variationist approaches to Spanish,* pp. 437-449. New York, NY: Routledge. https://doi.org/10.4324/9780429200267-39

Gudmestad, A., House, L., & Geeslin, K. L. (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), pp. 371-399. https://doi.org/10.1111/lang.12006

Gudmestad, A., & Metzger, T. A. (2025). Mixed-effects modeling with a multinomial dependent variable. *Language Learning*, 75(1), pp. 212-249. https://doi.org/10.1111/lang.12667

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), pp. 65-70.

Kanwit, M. (2017). What we gain by combining variationist and concept-oriented approaches: The case of acquiring Spanish future-time expression. *Language Learning*, 67, pp. 461-498. https://doi.org/10.1111/lang.12234

Kinginger, C. (2012). Enhancing language learning in study abroad. *Annual Review of Applied Linguistics*, 31, pp. 58-73. https://doi.org/10.1017/S0267190511000031

Lapidus, N., & Otheguy, R. (2005). Overt nonspecific ellos in the Spanish of New York. *Spanish in Context*, 2, 157-176. https://doi.org/10.1075/sic.2.2.03lap

Limerick, P. P. (2021). Establishing a variable context for lexical subjects in Spanish. *Forma y Función*, 34(I). https://doi.org/10.15446/fyf.v34n1.82653

Linford, B. G. (2016). *The second-language development of dialect-specific morpho-syntactic variation in Spanish during study abroad*. (Unpublished doctoral dissertation). Indiana University.

Linford, B., Long, A. Y., Solon, M., Whatley, M., & Geeslin, K. (2016). Lexical frequency and subject expression in native and non-native Spanish: A closer look at independent and mediating effects. In S. Sessarego & F. Tejedo-Herrero (eds.), *Spanish language and sociolinguistic analysis*, pp. 197-215. Amsterdam: John Benjamins. https://doi.org/10.1075/ihll.8.08lin

Linford, B. G., Zahler, S., & Whatley, M. (2018). Acquisition, study abroad and individual differences: The case of subject pronoun variation in L2 Spanish. *Study Abroad Research in Second Language Acquisition and International Education*, 3(2), pp. 243-274. https://doi.org/10.1075/sar.17003.lin

Llanes, À. (2011). The many faces of study abroad: An update on the research on L2 gains emerged during a study abroad experience. *International Journal of Multilingualism*, 8(3), pp. 189-215. https://doi.org/10.1080/14790718.2010.550297

Long, A. Y. (2021). Korean learners' acquisition and use of variable first-person subject forms in Spanish. *Languages*, 6, 208. https://doi.org/10.3390/languages6040208

Lozano, C. (2002). Knowledge of expletive and pronominal subjects by learners of Spanish. *ITL Review of Applied Linguistics*, 135, pp. 37-60. https://doi.org/10.1075/itl.135-136.02loz

Lozano, C. (2016). Pragmatic principles in anaphora resolution at the syntax-discourse interface: Advanced English learners of Spanish in the CEDEL2 corpus. In M. Alonso-Ramos (ed.), *Spanish learner corpus research: Current trends and future perspectives*, pp. 235-265. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.78.09loz

Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017). *Anglophone students abroad: Identity, social relationships, and language learning*. New York, NY: Routledge. https://doi.org/10.4324/9781315194851-2

Ortega, L. (2000). *Understanding syntactic complexity: The measurement of change in the syntax of instructed L2 Spanish learners* (Unpublished doctoral dissertation). University of Hawai'i at Manoa.

Ortega, L. (2014). Ways forward for a bi/multilingual turn in SLA. In S. May (ed.), *The multilingual turn: Implications for SLA, TESOL, and bilingual education*, pp. 32-53. New York: Routledge.

Otheguy, R., & Zentella, A. C. (2012). *Spanish in New York: Language contact, dialect leveling, and structural continuity*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199737406.001.0001

Otheguy, R., Zentella, A. C., & Livert, D. (2007). Language and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language*, 83, pp. 770-802. https://doi.org/10.1353/lan.2008.0019

Pasta, D. J. (2009). *Learning when to be discrete: Continuous vs. categorical predictors*. SAS. Global Forum (Vol. 248). SAS.

Pereira Scherre, M. M., & Naro, A. J. (1991). Marking in discourse: "Birds of a feather". *Language Variation in Change*, 3, pp. 23-32. https://doi.org/10.1017/S0954394500000430

Pozzi, R., & Bayley, R. (2020). The development of a regional phonological feature during a semester abroad in Argentina. *Studies in Second Language Acquisition*, 43, pp. 109-132. https://doi.org/10.1017/S0272263120000303

R Core Team (2021). *R: A language and environment for statisticalcomputing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rothman, J., & Iverson, M. (2007). On parameter clustering and resetting the null-subject parameter in L2 Spanish: Implications and observations. *Hispania*, 90, pp. 328-341.

Sanz, C., & Morales-Front, A. (Eds.). (2018). *The Routledge handbook of study abroad research and practice*. New York, NY: Routledge. https://doi.org/10.4324/9781315639970

Shin, N. L., & Otheguy, R. (2009). Diminishing sensitivity to continuity of reference: Subject pronoun use in New York City. In M. Lacorte & J. Leeman (eds.), *Español en Estados Unidos y en otros contextos: Cuestiones sociolingüísticas, políticas, y pedagígicas*, pp. 111–136. Madrid/Frankfurt: Iberoamericana/Vervuert Verlag. https://doi.org/10.31819/9783865279033-007

Silva-Corvalán, C. (2015). The acquisition of grammatical subjects by Spanish-English bilinguals. In A. Carvalho, R. Orozco, & N. Lapidus Shin (eds.), *Subject expression in Spanish: A cross-dialectal perspective*, pp. 211–230. Georgetown and Washington, DC: Georgetown University Press.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). *Bayesian measures of model complexity and fit. Journal of the royal statistical society: Series b (statistical methodology)*, 64(4), pp. 583-639. https://doi.org/10.1111/1467-9868.00353

Tarone, E., & Parrish, B. (1988). Task-related variation in interlanguage: The case of articles. *Language Learning*, 28, pp. 21-44. https://doi.org/10.1111/j.1467-1770.1988.tb00400.x

Torres Cacoullos, R., & Travis, C. (2010). Variable yo expression in New Mexico: English influence? In S. Rivera-Mills & D. Villa Crésap (eds.), *Spanish of the U.S. southwest: A language in transition*, pp. 189–210. Madrid, Spain: Iberoamericana/Vervuert. https://doi.org/10.31819/9783865278692-012

Travis, C. (2007). Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change*, 19, pp. 101–135. https://doi.org/10.1017/S0954394507070081

Wu, C. J., & Hamada, M. S. (2011). Experiments: planning, analysis, and optimization. Hoboken, NJ: John Wiley & Sons.

Zahler, S. L., Long, A. Y., & Linford, B. (eds.). (2023). *Study abroad and the second language acquisition of sociolinguistic variation in Spanish*. Amsterdam: John Benjamins. https://doi.org/10.1075/ihll.37

Appendix

Model selection using best subset selection

Fixed effects	BIC	BIC rank
Ref. cont., clause type, specificity, perseveration, task	13625.5	1
Ref. cont., clause type, specificity, perseveration, object pronoun task	13627.77	2
Number	16601.67	510
Number, time	16608.76	511