WRITTEN SUBJECT PRONOUN EXPRESSION AMONG SPANISH HERITAGE LANGUAGE LEARNERS[1]

LA EXPRESIÓN ESCRITA DEL PRONOMBRE SUJETO ENTRE HABLANTES DE HERENCIA DE LENGUA ESPAÑOLA

CARMEN FERNÁNDEZ-FLÓREZ

Grand Valley State University

Resumen

Este artículo revisa la expresión del pronombre sujeto (PS) en español desde una perspectiva poco estudiada: las narrativas escritas. A través del análisis de 90 ensayos escritos por estudiantes de español como lengua de herencia, el presente estudio tiene como objetivo mejorar nuestra comprensión del comportamiento del pronombre sujeto en la lengua escrita de esta comunidad de hablantes. Si bien el uso del pronombre sujeto en la lengua oral varía en términos de tasas generales de expresión entre los diferentes dialectos del español, se ha descubierto que una gramática subyacente similar predice su comportamiento (Carvalho et al., 2015). El presente estudio explora si estos patrones subyacentes se mantienen en las narrativas escritas.
Palabras clave: variación lingüística, sociolingüística, expresión del pronombre sujeto, lengua de herencia

Abstract

This paper revisits variable subject pronoun expression (SPE) in Spanish from an understudied perspective: written narratives. Using a total of 90 essays written by heritage language learners (HL) of Spanish, the present study aims to further our understanding of SPE's behavior in written language. While spoken SPE varies in terms of overall rates of expression across dialects, a similar underlying grammar has been found to predict its behavior (Carvalho et al., 2015). The current study explores if these underlying patterns are maintained in written narratives.
Key Words: language variation; sociolinguistics; subject pronoun expression; heritage languages

1.   INTRODUCTION

The Spanish language, with its rich cultural and linguistic heritage, has spread its influence across the globe, creating a diverse community of speakers, including heritage speakers—individuals raised in Spanish-speaking environments but exposed to varying degrees of formal instruction in the language. While much research has focused on second language acquisition, the study of Spanish heritage speakers offers a unique lens through which to examine the intricate dynamics of language development. A central aspect of this exploration is the expression of subject pronouns in Spanish—a topic that unveils an intriguing interplay of linguistic variation, identity, and cultural connection. It is not uncommon for teachers of Spanish as a heritage language to hear their students make claims like the following:

"He crossed out all my subject pronouns except these two; I guess these were important." This comment was made by a student in an advanced Spanish for heritage speakers course while revising a class essay. Clearly, he was aware of the variability of subject expression in Spanish but not of the patterns underlying that usage. This gap is fairly common, since the rules for subject pronoun expression (SPE) are rarely taught in university Spanish courses (Dracos, 2018). The issue of SPE is complicated by the fact that teachers and textbooks alike usually overproduce subject pronouns during early phases of instruction to reinforce verb conjugation. Consequently, learners often acquire this grammatical feature without any explicit instruction on when to produce or omit the subject pronoun.

In this study, we analyze patterns of SPE in a context barely studied before, written language, to discover how Spanish heritage language (SHL) learners use a variable feature that has been abundantly investigated in spoken language but almost never in writing. It is well established that "language must be studied as both a spoken and written phenomenon. Written language has its own norms, it produces for unique communicative needs, and variation and change in written language merits study just as much as that of spoken language" (Romaine, 1982, p. 24). Therefore, the current study investigates variation of SPE in written narratives of SHL learners.

Biber (1995) and Biber and Conrad (2001) compared spoken to written language. They concluded that written language differs from spoken mainly in being more structurally elaborate, complex, and explicit. These characteristics are especially important for SPE, since its usage entails control of syntactic, morphosyntactic, and pragmatic features. In this regard, evidence from corpus-based register studies show spoken and written modalities differ mainly in their potential for linguistic variation, probably because the inherent linguistic characteristics of spoken language are highly constrained whereas writing allows for a wider range of linguistic expression (Biber & Conrad, 2009, p. 261). Written language may resemble spoken utterances or be structurally more complex, a fact that is especially relevant for this study.

Due to the different linguistic backgrounds of the participant groups, the writing of SHL participants with lower proficiency levels is likely to resemble their spoken communication more closely than is the case for advanced SHL learners or monolingual Spanish speakers. Therefore, this study will focus narrowly on analysis of

a specific grammatical feature, SPE, in semi-formal written texts and will not address formal written texts, such as academic papers.

2. CHARACTERISTICS OF SHL LEARNERS

Many definitions for "heritage language learner" have been proposed due to the diversity of the population. Valdés describes an HL learner as "a language student who is raised in a home where a non-English language is spoken, who speaks or at least understands the language, and who is to some degree bilingual in that language and in English" (2001, p. 38). Meanwhile, Carreira defines HL learners based on three characteristics: "A) the learner's place in the HL community, B) the learner's personal connection to the HL and heritage culture through his/her family background, and C) the learner's proficiency in the HL" (2004, p. 2).

The varying backgrounds of the U.S. Latina/o/x[2] SHL population create very different types of Spanish speakers who have specific educational needs. The present study aims to contribute to the SHL field through a better understanding of the linguistic traits that differentiate HL learners' oral and written SPE. Because no SPE data has been collected for written texts, data on L1 written SP usage (see Appendix A) was also collected as a baseline comparison with HL, although establishing a direct comparison between both groups is not the focus of this paper.

2.1. Writing Production of Heritage Learners

Investigations of the writing performance of SHL students have been relatively prolific (Martínez, 2007; Mikulski & Elola, 2011; Gevers, 2018). Heritage speakers tend to use informal speech, and many have low writing skills because they have not received formal instruction in their heritage language. Writing requires much more than oral proficiency in the target language (Beaudrie et al., 2014). In fact, in a survey of 1,700 U.S. heritage learners of 22 languages, participants rated themselves lowest on writing skills (Carreira & Kagan, 2011).

Although writing is an important concern in the language development of heritage learners, little attention has been given to this modality, and few methodological advances in teaching writing have been made (Roca, 1997; Spicer-Escalante, 2006). A major difficulty HL teachers face is that heritage learners in the same classroom may have very different proficiency levels. At the same time, most HL curricula emphasize academic writing. Therefore, the particular profile of heritage learners makes analysis of their writing an interesting research topic.

Regarding stylistic differences, sentence-length structural differences between English and Spanish are particularly relevant to the present study. Spicer-Escalante

2   Latina/o/x is a term that aims to be gender inclusive for those who do not identify as binary (either female or male). This term is most widely used in the U.S., where the participants in our study are from.

(2009) compared the writing skills of SHL college students to those of English L1 speakers, Spanish L2 learners, and Spanish monolinguals. Regarding sentence length (measured by t-units[3]) the written texts of heritage learners were more similar to those of L2 native English speakers than to those written by Spanish monolinguals. However, both Spanish monolinguals and heritage learners used more subordinate clauses than did L2 learners of Spanish. These complexities make SHL students' written production a unique style that "borrows" from both of their languages when they write in either Spanish or English (Spicer-Escalante, 2009; see also Fairclough, 2006).

SHL speakers develop their own expressive writing space by shifting between both of their writing systems. They take elements from both English and Spanish to create their own discourse that reflects their everyday reality of living between two languages and two cultures in the United States. Therefore, their writing represents not only a translingual practice, but also a cultural one. HL instructors should recognize the value and richness of this written discourse to make writing instruction more effective in both Spanish and English.

2.2. Subject Pronoun Oral Expression among SHL Learners

Overall rates of SPE are variable across different varieties of Spanish. However, researchers have explained higher than average SPE production in certain Spanish dialects as based on contact with English (Abreu, 2009). Since subject pronouns are almost always overtly expressed in English, an important thread in the literature has investigated whether higher rates of SPE in bilingual or English-contact communities are indeed due to crosslinguistic influence.

Of special interest for this study is Silva-Corvalán's (2015) article analyzing the acquisition of subject pronouns by two Spanish-English bilingual siblings who had learned both languages from birth and had attained different levels of proficiency in Spanish. Her main research question was whether the young siblings would overproduce null subjects in English and/or overproduce overt subjects in Spanish compared to adult input, to their monolingual counterparts, and also to each other. Silva-Corvalán's results show that, for the acquisition of SPE, the dominant language, English, does not evidence any influence from contact with Spanish. A lesser degree of exposure to Spanish, however, results in higher frequency of overt subjects. The sibling with less exposure to Spanish expressed subject pronouns much more frequently, "suggesting that his Spanish is experiencing some degree of stress as English patterns become more entrenched" (p. 346).

---

3   *Minimal terminable unit* or *Total unit*, a linguistic concept used to measure the length and complexity of sentences. A t-unit consists of a main clause and any associated subordinate clauses or phrases that are attached to it. It is a way to break down complex sentences into smaller, more manageable units for analysis. Each t-unit represents a complete thought or idea that can stand alone as a sentence or be part of a larger sentence.

However, other studies of SPE in bilinguals have not found evidence of crosslinguistic interaction between null and overt pronoun languages. Ezeizabarrena (2012) found that for heritage bilinguals, the production of overt subjects increases developmentally in the overt pronoun language (English), while in the null language, Spanish, overt subjects are maintained at approximately the same rate over time. The type of early bilingualism is also an important factor. Montrul and Sánchez-Walker (2015) go beyond developmental vulnerabilities, erosions, or incomplete acquisition of the pragmatic features of Spanish overt subjects or a tendency toward structural convergence with English to find an alternative explanation for the SPE behavior. They emphasize the complex interplay of syntax and pragmatics in null-subject languages, "where pragmatic features of overt subjects play an important role" (p. 356). According to the authors, bilingual speakers tend to opt for SPE not because of English influence but because overt subjects are linguistically less complex than unmarked ones (see also Sorace, 2011).

In the case of oral narratives, Montrul (2004) found that young adult SHL speakers usually overproduce overt pronouns when maintaining reference across two consecutive grammatical subjects and overproduce null ones when switching reference. In spoken language, simultaneous Spanish-English bilingual children "exhibit the highest rates of overt subjects in pragmatically illicit contexts" (Montrul & Sánchez-Walker, 2015, p. 357), or, in variationist terms, in contexts normally unlikely to trigger the realization of a pronoun. The differences between children and young adults in regard to producing a subject pronoun in these different contexts are proof of distinct developmental paths in the SPE of heritage speakers.

In general, oral narratives tend to be more spontaneous than written texts; therefore, one could expect written production to show less SPE in pragmatically illicit contexts. The written style tends to be more formal and perhaps might be less redundant than oral discourse. Martínez Mira (2009) analyzed the subjunctive expression of SHL speakers in New Mexico. Analyzing written sentences and oral samples from Spanish monolinguals and heritage speakers, she found the two groups behaved very similarly in oral but not in written production. While both groups used subjunctive with concessive clauses in oral data, in the written samples, the monolinguals maintained the subjunctive usage but heritage speakers produced lower rates of subjunctive than in the oral task. The author concedes that differences in the tasks could have contributed to eliciting different results, but she also concludes that more attention to the written production of heritage speakers is needed in order to better understand their writing processes.

Regarding written variation, SHL speakers seem to borrow from both languages in their writing, so written texts may be an ideal context in which to analyze possible effects of contact with English, as well as developmental patterns of acquisition of SPE.

3.   METHODOLOGY

By combining variationist methodologies from previous studies (Martínez, 2007; Travis, 2007; Torres Cacoullos & Travis, 2010, 2011), this study seeks to identify the predictors for written Subject Pronoun Expression (SPE) performance among heritage language (HL) learners. By addressing this primary question, we aim to delve into a comprehensive analysis of the following sub-questions:

Firstly, we intend to examine how the written production of subject pronoun expression in HL learners compares with their oral production. This comparison will offer valuable insights into the dynamics of language production in the heritage language context and shed light on potential disparities between written and oral proficiency.

Secondly, our study will explore whether a comparison of subject pronoun expression at three distinct proficiency levels among HL learners reveals discernible stages of acquisition. This sub-question aims to discern whether there is a developmental trajectory in the acquisition of SPE in the context of heritage language learners, thereby providing a more nuanced understanding of their language patterns.

3.1. Participants

Participants were 36 L1 Spanish-speaking students at the Instituto Tecnológico de Nogales (Sonora, Mexico) and 90 SHL learners attending a university in the U.S. Southwest. The L1 speakers were between the ages of 19 and 23 (M= 19.89, SD = 1.62). They had never lived in the U.S. and had had no contact with English other than the courses they took in high school. The HL learners ranged in age from 18 to 23 (M= 19.78, SD = 1.39) and were enrolled in either Intermediate Spanish I or Intermediate Spanish II for Heritage Learners. They completed a language background questionnaire adapted from the Corpus de español en el sur de Arizona (CESA, Carvalho, 2012) to ensure they had similar backgrounds, had never lived abroad, and spoke the same variety of Sonoran Spanish as the students from the Instituto Tecnológico. Thus, only HL speakers with Northern Mexican ascendancy were considered.

3.2. Data Collection

Student participants were recruited by means of classroom visits. They were asked to write an essay of around 500 words (in Microsoft Word) based on a prompt shown on the classroom projector:

Por favor, escriba sobre una experiencia interesante que haya tenido en el pasado o relate las aventuras del mejor viaje que haya hecho.

[Please, write about an interesting experience that you have had in the past, or narrate the adventures of the best trip you have ever had.][4]

_____

4   This and all translations hereafter are own translations

This topic was chosen based on Martínez (2007), who also asked his students to write about a personal experience. He argued that such informal writing would show more similarities to speech routines. For purposes of comparing oral and written SPE production, it was determined that informal texts would produce more relevant results. Formal texts would contain more differences due to both literacy skill transfer and differences in the nature of the oral and written tasks. Students completed these compositions as in-class practice assignments that did not affect their course grade. Assigning an ungraded story about students' own lives was intended to reduce "the pronounced relational asymmetry with the instructor and [ensure] that they did not write with the specific intent of living up to the instructor's expectations" (Martínez, 2007: 34). The participants uploaded their completed compositions to a designated folder during the same week (week 13 in a 16-week semester). This measure assured that all students had the same amount of in-class input and experience with their corresponding instructors at the time of data collection. The course instructors varied in country of origin (most were Mexican, but a few were Spaniards); however, none were from the Caribbean, which is home to a Spanish variety known for its distinct use of SPE (Abreu, 2009).

3.3. Analysis

This section explains why only first-person SPE was analyzed, followed by a description of the coding criteria, the statistical analyses, and the variationist analysis of this variable.

3.3.1 Why Only Analyze First-Person Subject Pronouns?

There were several reasons to limit analysis to first-person subject pronouns. First, previous research has uncovered differences in the way linguistic variables function across different verbal conjugations (Travis 2007; Torres Cacoullos & Travis 2010; Geeslin & Gudmestad, 2011; Shin 2014). Second, a personal narrative, by its nature, tends to elicit first-person voice and could contain insufficient references to other grammatical persons for analysis. In addition, research has shown that SHL learners usually demonstrate increased use of overt third-person pronominal subjects, mainly attributed to contact with English (Otheguy & Zentella, 2012), but results of studies on first-person pronouns are less conclusive (Travis & Torres Cacoullos, 2012).

3.3.2 Coding Criteria

The data for this study included all first-person singular and plural verbs with which a human pronominal subject could be expressed or omitted. Therefore, any contexts where use of a first-person pronoun was mandatory were omitted from the data. Examples are *yo que sé* or *que sé yo* (what do I know), which are fixed expressions containing the pronoun. Other contexts that are usually excluded in analyses of oral SPE, such as abandoned utterances (e.g., *ella iba a…*; 'she was going to…'), did not occur

in this dataset since participants had the option to revise their texts. This makes the analysis of written data even more interesting, because the two discourse modalities can produce and follow different norms to some extent. Another difference between the data collected in sociolinguistic interviews versus the written data in this study is that compositions are a unidirectional discourse where the presence of the researcher is less direct and the observer's paradox is less likely (Labov, 1984). Because participants are not influenced by the SPE of the interviewer, they are more likely to demonstrate their own command of SPE.

3.4. Coding of SPE

The following paragraphs describe in detail each independent factor accounted for in codification of the data. Table 1 lists the factors, with examples for each.

Table 1. Independent factors coded in the data.

Factor	Example	Gloss
Number		
First-person singular	*Yo viajo a México* (participant HL-15)	I travel to Mexico
First-person plural	*Nosotros fuimos a esquiar* (participant L1-12)	We went skiing
Tense, mood, and aspect (TMA)		
Present simple indicative	Ø *Pienso que ella era muy bonita* (participant LH-85)	I think she was very pretty
Present progressive	*Yo estoy yendo al gimnasio todos los días* (participant LH-65)	I'm going to the gym every day
Present perfect	*Nunca Ø he ido a Nueva York* (participant LH-66)	I have never been to NYC
Preterit	*Yo fui a México* (participant LH-30)	I went to Mexico
Imperfect indicative	*Cuando Ø iba a mi pueblo* (participant LH-33)	When I used to go to my village
Pluperfect indicative	*Yo había viajado la semana anterior* (participant LH-35)	I had traveled the previous week
Imperfect subjunctive	*Ojalá Ø estuviera de vacaciones* (participant L1-12)	I wish I were on vacation
Pluperfect subjunctive	*Si yo hubiera sabido lo que pasaba* (participant L1-07)	If I had known what was going on

Conditional	*Ø iría allí de nuevo* (participant L1-11)	I would go back there again
Other (rare verb tenses occurring in less than 3% of tokens)	*Yo no querré volver* (participant LH-66)	I won't want to come back

Clause type		
Main	*Yo tenía mucho sueño* (participant LH-31)	I was very sleepy
Subordinate	*Lo mejor fue cuando Ø fui con mis amigas* (participant L1-08)	The best part was when I went with my friends
Coordinate	*Yo fui y yo compré los boletos* (participant L1-20)	I went and I bought the tickets

Switch reference		
First token	*Yo pienso que es interesante* (participant LH-65)	I think it is interesting
Same reference	*Yo siempre voy y Ø compro dulces* (participant LH-33)	I always go and I buy candies
Switch reference	*Ø Siempre voy pero ella nunca va* (participant L1-08)	I always go but she never does

Discourse connectedness		
Degree 1: Subject, verb tense, and verbal mood remain the same throughout the clauses.	*El jueves fui a la tienda. Ø Fui con unas amigas* (participant LH-45)	Last Thursday I went to the store. I went with my friends.
Degree 2: Subject stays the same, but either the tense, mood, or aspect of the verb changes.	*Hace tiempo que compré el carro y Ø estoy muy contenta* (participant LH-52)	I bought the car a while ago and I am very happy with it.
Degree 3: Subject changes, but the clause between the subject pronoun and the previous token is not a rival candidate for subject.	*En verano yo fui de vacaciones a México. Hacía mucho calor y entonces **yo** sufrí mucho* (participant LH-63)	Last summer I went to Mexico. It was very hot and I suffered a lot.
Degree 4: Subject changes and it last appeared in a different syntactic function (e.g., direct object, possessive).	*Ellos me llamaron a las 6am. Ø estaba muy cansada* (participant LH-57)	They called me at 6am. I was very tired.

173

Degree 5: Subject continuity is not maintained in the adjacent or in the following clause.	Ø Fuimos a cenar a un italiano. Mi amiga pidió lasaña y Ø *le gustó mucho. Yo pedí pizza carbonara* (participant L1-20)	We went to an Italian for dinner. My friend ordered lasagna and she liked it a lot. I ordered pizza carbonara.
Degree 6: Discourse topic entirely changes.	*La cena duró dos horas y Ø comí mucho. En cuanto a mi viaje a Florida, Ø fui en marzo* (participant L1-20)	Dinner lasted two hours and I ate a lot. In regards to my trip to Florida, I went in March.
Reflexivity		
Reflexive verb	*Mis amigos y yo nos conocimos hace mucho* (participant LH-33)	My friends and I met long ago
Non-reflexive verb	*No Ø voy a volver* (participant L1-09)	I am not going to come back
Accuracy: Beginner; Intermediate; Advanced[5]		

3.4.1 Number

In previous analyses, both monolingual and bilingual speakers tended to omit pronouns with a plural subject (Flores-Ferrán, 2004; Hurtado, 2005; Abreu, 2009); therefore controlling for the effect of number on SPE was important (LaCasse, 2019), especially given that this is the first time SPE has been analyzed in naturalistic written texts.

Analyzing the difference between first-person singular and plural pronouns is relevant since previous research indicates speakers exhibit different trends based on person and number. It was hypothesized that the monolingual group would follow the trend found for monolingual oral SPE, with higher expression of the singular pronoun than the plural. For the SHL learners, it was expected that those from the lower-level group would produce higher rates of overt singular pronouns but similar rates of plural pronouns compared to the monolinguals. Heritage speakers from the more advanced group were expected to present a similar pattern of SPE as monolingual speakers for both singular and plural pronouns.

3.4.2 Tense, Mood, and Aspect (TMA)

TMA is one of the most common factors considered when coding for SPE because it helps determine whether verb morphology influences the selection of one type of pronoun over another. In Spanish, verb ambiguity exists in first- and third-person

---

5   No language example presented here since only a comprehensive analysis of the entire writing sample could be proof of a certain accuracy level. More details of this comprehensive analysis in section 3.4.8.

singular forms in the conditional, subjunctive, and imperfect tenses. Some previous research indicates this ambiguity triggers an increase in the selection of overt pronouns (Bayley & Pease-Álvarez, 1997; Shin & Montes-Alcalá, 2014); other studies, however, found no effects due to ambiguity in oral discourse (Bentivoglio, 1987; Travis, 2007).

To clarify the effects of morphological ambiguity on SPE, Linford & Shin (2013) conducted a study that combined TMA with lexical frequency. Their results showed that, among fourth-semester L2 Spanish students, TMA does not play a role in SPE unless the verbs are coded by frequency: imperfect morphology shows a bigger effect on SPE for frequent verbs. Shin (2016) found a similar pattern for monolingual children (6-7-year-olds), who were not sensitive to the effect of the imperfect, except when frequent verb lexemes were isolated. For this reason, frequency is also considered in the present analysis. It is hypothesized that because of the less ambiguous nature of written discourse, the frequency of imperfect verbs will not affect SPE.

3.4.3 Clause type

Previous research has found pronouns are most likely to be expressed in main clauses, are less likely in subordinate clauses, and are least likely in coordinate ones (e.g., Carvalho et al., 2015). This general tendency has been found in the oral discourse of monolinguals (Abreu, 2009; Orozco & Guy, 2008) and bilinguals (Flores-Ferrán, 2004; Otheguy & Zentella, 2012; Shin & Montes-Alcalá, 2014; Bessett, 2017). However, a few studies with bilingual speakers did not find an effect for clause type (Prada Pérez, 2009; Carvalho & Child, 2011; Torres Cacoullos & Travis, 2011). It is hypothesized that, if English is influencing participants' SPE behavior, they will not be sensitive to clause type (especially for main and subordinate clauses, since pronoun dropping may occur in English with coordinate clauses). Moreover, the written modality may influence the type-of-clause factor. Since writing favors the retrieval of information by returning to the text. participants might not need to express pronouns as often, not even in main clauses. If this is the case, lower rates of SPE would be expected in all types of clauses.

3.4.4 Switch Reference

This category codes the relationship between subjects in consecutive sentences. Topic continuity in important when analyzing discourse because "what is continuing is more predictable, and what is predictable is easier to process" (Givón, 1983: 12). In general, pronouns are expressed more often when there is a switch in reference than when the reference is maintained across clauses (Carvalho et al., 2015).

3.4.5 Discourse Connectedness

Paredes Silva (1993) introduced the concept of discourse connectedness in an analysis of personal letters written in Brazilian Portuguese. Because this linguistic factor was specifically designed for written data, it suits the purpose of the current

study. Discourse connectedness goes beyond a switch in reference to consider the relationship of a given reference with references in the previous clauses. This linguistic factor, adapted from Paredes Silva (1993), consists of six degrees of connectedness (see Table 1). Paredes Silva (1993) found that the lower the degree of discourse connectedness, the higher the probability of finding explicit pronouns. Similar results are expected in the present data, with a clearer tendency toward connectedness among the L1 speakers, who have formal training in Spanish writing.

3.4.6 Reflexivity

Previous literature found that reflexive verbs triggered the omission of subject pronouns (Abreu, 2009, 2012; Bayley & Pease-Álvarez, 1996; Carvalho et al., 2015; Otheguy & Zentella 2012; Shin & Montes-Alcalá, 2014); therefore, both participant groups in this study are expected to show some degree of effect for verb reflexivity.

3.4.7 Accuracy

Recall that SHL participants were recruited from two consecutive intermediate-level courses. In reading the compositions, it became obvious that participants enrolled in the same course had very different Spanish linguistic proficiency levels and SPE behavior. Therefore, an "accuracy" coding criterion was added to group participants based on linguistic ability rather than course enrollment. Because language proficiency tests can be extremely inaccurate for heritage learners (Thompson, 2015), particularly as most participants in the present study had never taken a Spanish course before and were unfamiliar with language testing practices, their compositions were individually analyzed to assign an accuracy level. A cumulative rating scale commonly employed in sociolinguistics to interpret the social categories of participants was used to derive a score corresponding to beginner, intermediate, or advance accuracy (Blommaert et al., 2025). Table 2 shows how the analysis was operationalized.

Table 2. Operationalization of linguistic accuracy scores

# Non-Target-Like Forms	0-3	4-7	+8
Points Assigned	1	2	3
Multiplier for Error Type	Syntactical = 0.5	Morphological = 0.3	Lexical = 0.2
Accuracy Level by Score	Beginner = 2.01-3.0	Intermediate = 1.01-2.0	Advanced = 0-1.0

The three categories of syntactical, morphological, and lexical errors had a weighting multiplier assigned to them, from 0.5 for syntactical errors to 0.2 for lexical ones. Participants' compositions received a point value from 1 to 3 for each category based on the number of non-target-like forms. A composition with more than 8 non-target-like syntactic forms, 3 non-target-like morphological forms, and 5 non-target-like lexical ones would receive a total score of 2.2 by the following equation: (syntax: 3 x 0.5) + (morphology: 1 x 0.3) + (lexicon: 2 x 0.2) = 2.2. The higher the score, the greater the number of non-target forms and hence the lower the accuracy level. As Table 2 shows, a score of 2.2 corresponds to beginner level. Recall that all the participants were in intermediate Spanish courses, so "beginner," "intermediate," and "advanced" merely ranks them within an overall intermediate level. Therefore, an "advanced" speaker in this context does not necessarily correspond to the characteristics of an advanced speaker in other proficiency scales (e.g., ACTFL). This classification system offers a holistic analysis of participants' language skills, based on their own written production and, therefore, is expected to well describe their language performance. It is hypothesized that accuracy will have a significant effect on SPE behavior and that the statistical analyses will support this method of grouping participants.

4. STATISTICAL ANALYSES AND RESULTS

Initially, overall rates of pronominal subject expression were examined, followed by a multivariate logistic regression analysis to investigate the influence of the various linguistic factors on SPE across groups. Rbrul software was used for the mixed-effects model analyses to determine which linguistic factors were and were not significant "over and above the effect of individual" (Tagliamonte, 2012: 143).

4.1. Percentages of Overt Pronoun Expression by Accuracy Level

The overall percentage of overt SPE is similar across the three accuracy groups (see Table 3). However, beginner-level participants show a significantly higher rate of overt expression than the other groups (p = <.0001). This is not surprising given that beginning learners of Spanish and English-dominant bilinguals usually show a higher rate of overt SPE due to cross-linguistic influence from English. Regardless, the overall rates clearly progress from the highest overt expression rate at the beginner level to the lowest rate at the advanced level. Overall rates of SPE are markedly lower than found in previous studies of bilingual speakers (e.g., the Mexican-descendent immigrant children in Bayley & Pease-Álvarez, 1997, expressed subject pronouns 24% of the time), and also lower than rates reported for monolingual Mexican adults, usually around 20% (Lastra & Martín Butragueño, 2015; Michnowicz, 2015; Otheguy & Zentella, 2012; Silva-Corvalán, 1994).

Table 3. Overall rates of SPE among HL students by accuracy level

	Beginner		Intermediate		Advanced	
	#	%	#	%	#	%
Expressed SP/overall SP	58/413	14	59/740	8	44/875	5
Unexpressed SP/overall SP	355/413	68	681/740	92	831/875	95

Research on SPE among heritage learners is scarce, and most of what exists was conducted with balanced bilinguals or involved grammaticality judgment tests rather than spontaneous production (e.g., Montrul & Rodríguez-Louro, 2006). Though overall rates of SPE in this study differ from findings of previous studies, all the previous studies addressed spoken language, whereas this study examined written texts (see Table 4).

Table 4. Previous studies of HL speakers' oral SPE compared to present study.

Community (Study)	Overall %
Mexican Spanish in Los Angeles (Silva-Corvalán, 1994)	30%
Mexican Spanish in California (Bayley & Pease-Álvarez, 1997)	23%
Puerto Rican Spanish in Florida (Abreu, 2009)	33.5%
Mexican Spanish in Tucson, AZ (Bessett, 2017)	16.7%*
Mexican Spanish in Arizona (current study with written data)	8%

*only first-person singular

Because most studies on HL SPE production have been conducted with spoken language, it is difficult to compare rates of production across bilingual varieties due to differences in the type of task. However, the data in Table 4 do show two clear tendencies. First, Caribbean Spanish demonstrates significantly higher rates of pronoun expression in comparison to mainland varieties such as Mexican Spanish; second, overall SPE production is clearly different between oral and written narratives, with written narratives showing a greater tendency to avoid the expression of pronouns. In the table we can also see how in Bessett (2027), who also only analyzed first person pronouns but in spoken language, the overall production rate was double the production in the current study with written data (16.7% vs. 8%).

4.2. Multivariate Rule Analysis for Beginner Accuracy Level

After reviewing the overall tendencies in subject pronoun expression, it is turn now to describe the specific linguistic factors that influence SPE among beginner heritage learners. The factor weights in Table 5 show the direction of the effect; weights higher than .50 indicate that overt pronouns are favored in that context. The range indicates the effect size of each linguistic factor group, and factors are listed in order from largest to smallest magnitude effect.

Table 5. Multivariate rule analysis of linguistic factors influencing overt SPE among beginner heritage learners.

Factor	Weight	%	# SPE Tokens/ Total Tokens	*p* Value
Clause type				<0.0001
Main	.75	18%	36/217	
Subordinate	.70	15%	20/133	
Coordinate	.13	2%	2/63	
Range 62				
Grammatical number				<0.001
1st-person singular	.60	18%	41/227	
1st-person plural	.40	11%	21/186	
Range 20				

Total N 413
Nagelkerke R2 .5
Intercept -2.15
Log.likelihood -143.428

As Table 5 shows, only two linguistic factors were significant for beginner heritage learners. Clause type, as the factor with the biggest range, had the greatest effect on the realization of overt pronouns for these speakers. Within this factor, main, subordinate, and coordinate clauses are ranked as expected, with the last being the least likely to trigger an overt subject pronoun. With a range of 20, grammatical number behaves very similarly to the results obtained for this factor group in previous literature: first-person singular triggers more overt pronouns than first-person plural does (e.g., Abreu, 2009; Bayley & Pease-Álvarez, 1996).

4.3. Multivariate Rule Analysis for Intermediate Accuracy Level

Turning now to the SPE behaviors among intermediate-level learners, Table 6 presents the linguistic constraints that significantly influence them.

Table 6. Multivariate rule analysis of linguistic factors influencing overt SPE among intermediate heritage learners.

Factor	Weight	%	# SPE Tokens/ Total Tokens	*p* Value
Grammatical number				<0.0001
1st-person singular	.75	13%	47/361	
1st-person plural	.25	1%	4/379	
Range 48				
Switch reference				<0.001
First token	.68	47%	7/15	
Switch reference	.57	10%	37/374	
Same reference	.31	5%	18/351	
Range 37				
Clause type				<0.01
Main	.63	10%	39/395	
Coordinate	.50	5%	6/116	
Subordinate	.37	7%	16/229	
Range 26				

Total N 740

Nagelkerke R2 .63

Intercept -3.96

Log.likelihood -154.253

For intermediate learners, three linguistic group factors were important: grammatical number had the greatest impact, followed by switch reference. Like Abreu's (2009) results for bilingual speakers of Puerto Rican Spanish, this finding implies that a change in reference triggers more SPE. The fact that this effect was found in the intermediate heritage learners but not in the beginners group is also significant. The final constraint for the intermediate group, clause type, followed the expected pattern: main clauses triggered more SPE than the other types of clauses. Specifically,

in the current data, overt expression was low in coordinate clauses because they mainly occur in same-reference contexts and, thus, favor null expression.

4.4. Multivariate Rule Analysis for Advanced Accuracy Level

Continuing on to the advanced group of heritage learners, Table 7 shows the hierarchy of linguistic constraints influencing their selection of overt SPE.

Table 7. Multivariate rule analysis of linguistic factors influencing the appearance of overt SP among advanced heritage learners.

Factor	Weight	%	# Tokens	*p* Value
Grammatical number				<0.0001
1st-person singular	.83	9%	40/441	
1st-person plural	.17	1%	4/434	
Range 66				
Switch reference				<0.001
First token	.75	7%	30/426	
Switch reference	.53	3%	13/428	
Same reference	.23	8%	9/21	
Range 52				
TMA				
Subjunctive	.72	19%	3/16	
Imperfect indicative	.64	10%	14/144	
Pluperfect indicative	.63	13%	2/15	<0.001
Preterit indicative	.42	4%	22/532	
Present indicative	.27	2%	2/108	
Range 45				
Clause type				<0.01
Main	.60	6%	27/455	
Coordinate	.59	5%	6/125	
Subordinate	.33	3%	9/295	
Range 27				

Total N 875

Nagelkerke R2 .76

Intercept -7.493

Log.likelihood -132.985

181

This statistical analysis yielded four significant linguistic factors. The first, grammatical number, paralleled the results from the previous groups as well as generalized behavior in oral Spanish (Carvalho et al., 2015, p. xiii). First-person singular clearly triggered more SPE than first-person plural did (weights of .83 versus .17, respectively). Ranked next and following previous patterns for this factor, switch reference triggered more SPE than did same-referent contexts. Third in the hierarchy, with regard to tense, mood, and aspect, subjunctive and imperfect, both ambiguous forms, showed the biggest effect on the selection of SPE. The last influential factor for advanced heritage learners was clause type, where like in the intermediate group, main and subordinate clauses had stronger tendencies toward selection of overt SPE than subordinate clauses did (weights of .60, .59, and .33, respectively).

5. DISCUSSION

Based on the results in Tables 5, 6, and 7, various patterns of linguistic constraints affected SPE by students at all three accuracy levels. In this section, the results for each accuracy group are compared against each other and with previous findings regarding Spanish SPE by heritage learners and bilinguals.

5.1. *Comparison of Significant Linguistic Factors across Accuracy Levels*

Table 8 shows which linguistic factors were significant for each accuracy group.

Table 8. Significant linguistic constraints for SPE across HL beginner, intermediate, and advanced accuracy levels

	Beginner	**Intermediate**	**Advanced**
Grammatical number	✔	✔	✔
Switch reference		✔	✔
Clause type	✔	✔	✔
TMA			✔
Reflexivity			
Discourse connectedness			

Only four factors out of the six analyzed produced significant results. However, Table 8 also presents a clear progression toward the acquisition of factors in that learners increase the complexity of the variable by controlling more linguistic constraints as their language skills advance. To compare the similarities and dissimilarities of these three accuracy groups, the linguistic constraints that all three groups share are first discussed to then turn individually to those that affect only some groups.

5.2. Grammatical Number

Grammatical number was significant for all the accuracy groups. Since the current study only examined first-person pronouns, the results are difficult to compare with previous studies that generally looked at various grammatical persons. However, one clear tendency echoes what other studies have found: in terms of number, singular persons trigger more overt expressions than plural ones do (Abreu, 2012; Carvalho et al., 2015; Linford, 2016). Nonetheless, all three groups of heritage learners still show a lower overall rate than bilinguals in other communities used in spoken language (e.g., Otheguy & Zentella, 2012; Shin & Otheguy, 2013). These results highlight the differences across communication modes, suggesting that written and spoken language have different overall rates of SPE.

5.3. Switch Reference

Switch reference was significant for both intermediate and advanced heritage learners, but not for beginners. The tendencies among intermediate and advanced learners were expected and correspond to those in previous research on spoken SPE (Bessett, 2017): a switch in reference triggers production of more overt pronouns. However, the effect was lower in the written data analyzed in the current study compared to the previous findings for spoken language. Shin & Otheguy (2009), in their study of reference continuity effects in spoken SPE among first- and second-generation bilingual speakers in New York City, claimed a desensitization effect on continuity of reference as a predictor of the overt-versus-null variation of subject pronouns. They claimed that second-generation bilinguals had a weaker tendency to favor overt pronouns more in switch-reference than same-reference contexts. This desensitization occurred with first- and second-person verbs, where this variable is less important for communicative purposes. When reference tracking was more difficult, however—that is, when referring to third-person singular subjects—sensitivity to continuity remained essentially the same as among first-generation bilinguals. These results could help explain the behavior of heritage learners, for whom the effect of switch reference was lower than for L1 speakers. In the current study, however, heritage learners displayed an increasing sensitivity to switch reference as their accuracy level increased. This tendency is especially interesting if we consider that (due to the topic of the assigned composition) half of the data analyzed for this study consisted of first-person singular pronouns expressed in past tenses, such as the ambiguous imperfect tense. Thus, we could expect switch reference to have had a greater effect, since first person can create some ambiguous references if the pronoun is null (e.g., *yo iba* versus *él iba*). Yet, despite these characteristics of the data, switch reference affected written SPE in this study less robustly than it did in previous studies of spoken language (Abreu, 2009; Linford, 2016). One might postulate a slight loss of sensitivity to switch-reference. Yet, if we consider the data presented in Appendix A for L1 speakers, and how the direction and size of the switch-reference effect was similar in the writing of the HL learners and of the L1 speakers, it seems more plausible that the lesser impact

of switch reference on SPE was due to differences between oral and written language use, rather than desensitization.

5.4. Clause Type

Clause type was a significant factor at all three HL accuracy levels, with the largest size effect range (62) among beginners. These results align with previous research and are expected since, as Bybee (2002) mentioned, main clauses are the type most likely to bring new information into the discourse and, thus, trigger the selection of overt pronouns, while subordinate clauses usually contain previously presented information. Given this pattern, it makes sense that the beginner group relies most heavily on this factor to navigate the variability of subject pronouns in Spanish. Clause type is relatively easy to notice and control since, for example, coordinate sentences function similarly in English and Spanish. If the reference is maintained, null pronouns are favored, but when the coordinate clause introduces a change in reference, pronoun expression is mandatory in English and more likely in Spanish. Thus, this is an easier linguistic constraint for beginners to master since it does not rely on metalinguistic knowledge the way TMA does, but rather on information the writer possesses about the referents in the discourse. If a new person is not being introduced in a subordinate clause, the writer is able to realize that the pronoun likely already expressed in the main clause need not be repeated.

Beginner HL learners showed the pronominal behavior most similar to Spanish L1 speakers and Spanish-English bilinguals regarding clause type since the frequency of SPE follows the same hierarchy across groups: from main (most SP expression) to subordinate to coordinate (least SP expression). The intermediate and advanced HL students also showed an effect for clause type, but in their case, the order of the effect was main, then coordinate, then subordinate. Note, however, that the beginner group had around half the number of coordinate clauses (63) as the intermediate and advanced groups (116 and 125, respectively). This imbalance might have affected the order of the effect for the different clause types.

5.5. Tense, Mood, and Aspect

The last factor influencing SPE was TMA, which was significant only for the advanced group. Although most studies have claimed a functional effect of TMA on the oral SPE of monolinguals (Carvalho et al., 2015; Hochberg, 1986), other studies have found mixed results (Orozco & Guy, 2008). Among bilingual speakers, however, the common finding is that ambiguous forms trigger more overt pronouns. For example, in the case of Maya-Spanish bilinguals in Yucatán, 36% of overt pronouns were expressed with ambiguous tenses (Michnowicz, 2015). This 36% occurrence rate contrasts with the average of 9% SPE with ambiguous forms found in the writing of advanced heritage learners in the present study. Thus, while the current study aligns with previous findings in respect to ambiguous forms showing higher overt subject

pronouns, the overall rates of SPE are still significantly lower than those obtained in research based on oral data.

6. CONCLUSIONS

This study analyzed the different behaviors that heritage learners show in regard to patterns of SPE. Results show that they do acquire variability in this morphosyntactic structure. Their overall rates of production are very low (less than 20% for all three groups), clearly indicating that these participants are not drawing on their English grammar competence, since transfer from English would imply higher rates of SPE. Furthermore, each accuracy group shows clear behavioral patterns: beginning learners show command of two linguistic constraints (grammatical number and clause type), intermediate learners are sensitive to three constraints (grammatical number, clause type, and switch reference), while advanced learners have mastered four linguistic factors (grammatical number, switch reference, clause type, and TMA).

After confirming that heritage learners do indeed acquire this variable grammar and produce it in their writing, this study also demonstrated how using a bottom-up analysis of written data results in a more accurate description of their SPE behavior. This analytical approach offers better insights into the sequence of acquisition for these learners since it is able to capture how beginner versus advanced learners differ in terms of both overall rates of SPE and control of linguistic constraints. Whereas beginners produce overt SP 14% of the time, advanced learners do so 5% of the time, a distinction that would have been lost had participants been grouped based only on their course enrollment. This point is even more relevant if we consider that these students had taken a placement exam that had identified them as having similar levels of proficiency and had placed them in the same SHL course.

Finally, the present analysis of HL writing has shown that the linguistic constraints on SPE are consistent across spoken and written Spanish, showing that the grammar underlying SPE is basically the same for SHL learners (Michnowicz, 2015; Travis, 2007; Torres Cacoullos & Travis, 2010). However, given that overall rates of SPE in the present study are lower than was found in previous studies, we can conclude that heritage learners employ this variable structure differently in spoken language than in writing. Further research on written SPE is needed to determine whether HL learners' infrequent use of SPE holds across different types of writing or is in fact the result of writing a narrative.

This study provides further evidence of the importance of investigating variation in written versus oral discourses as separate phenomena. Doing so not only acknowledges the differences between these two modalities of communication, but might also establish a connection between these differences and the type of input SHL learners receive in their courses. Whereas recent approaches to the HL curriculum embrace code-switching practices during classroom discussions (Valdés et al., 2016), instruction in writing is still very much dominated by more formal academic language practices (Mikulski & Elola, 2011; Gevers, 2018). An analysis of these different

instructional approaches and how they affect the SPE of heritage learners would be a fruitful next step. Such a study could help inform writing instruction in the SHL classroom. If as language teachers we provide different types of input to foster the oral and written skills of our HL learners, as language researchers we should also employ distinct methodological tools to describe and better understand the differences between the oral and written data we collect for our research.

REFERENCES

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

Abreu L. (2012). Subject pronoun expression and priming effects among bilingual speakers of Puerto Rican Spanish. In Geeslin & Díaz-Campos (Eds.), *Selected proceedings of the 14th Hispanic Linguistics Symposium* (pp. 1-8). Cascadilla Proceedings Project.

Bayley, R., & Pease-Álvarez, L. (1996). Null and expressed pronoun variation in Mexican-descent children's Spanish. In J. Arnold, R. Blake & B. Davidson (Eds.), *Sociolinguistic variation: Data, theory, and analysis* (pp. 85-99). Stanford: Center for the Study of Language and Information.

Bayley, R., & Pease-Álvarez, L. (1997). Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change, 9*(3), 349-371.

Beaudrie, S., Ducar, C., & Potowski, K. (2014). *Heritage language teaching: Research and practice*. McGraw-Hill Education.

Bentivoglio, P. (1987). La posición del sujeto en el español de Caracas: un análisis de los factores lingüísticos y extralingüísticos. In R. Hammond & M. Resnick (Eds.), *Studies in Caribbean Spanish dialectology* (pp. 13-23). Georgetown University Press.

Bessett, R. M. (2017). Testing English influence on first-person singular 'yo' subject pronoun expression in Sonoran Spanish. *Contemporary Trends in Hispanic and Lusophone Linguistics*, *15*, 355-372.

Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge University Press.

Biber, D., & Conrad, S. (2001). Quantitative corpus-based research: Much more than bean counting. *TESOL Quarterly, 35,* 331-336.

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.

Blommaert, J., Westinen, E., & Leppänen, S. (2015). Further notes on sociolinguistic scales. *Intercultural Pragmatics*, 12(1), 119-127

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*, 261-290.

Carreira, M. (2004). Seeking explanatory adequacy: A dual approach to understanding the term "heritage language learner." *Heritage Language Journal, 2*(1), 1-25.

Carreira, M., & Kagan, O. (2011). The results of the National Heritage Language Survey: Implications for teaching, curriculum design, and professional development. *Foreign Language Annals*, *44*(1), 40-64.

Carvalho, A. M. (2012). *Corpus del español en el sur de Arizona (CESA)*. University of Arizona.

Carvalho, A. M., & Child, M. (2011). Subject pronoun expression in a variety of Spanish in contact with Portuguese. In *Selected Proceedings of the 5th Workshop on Spanish Sociolinguistics* (pp. 14-25).

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

Dracos, M. (2018). Teacher talk and Spanish subject personal pronouns. *Journal of Spanish Language Teaching*, *5*(1), 1-15.

Ezeizabarrena, M. J. (2012). Overt subjects in early Basque and other null-subject languages. *International Journal of Bilingualism, 17*(3). https://doi.org/10.1177/1367006912438997

Fairclough, M. (2006). Spanish/English interaction in U.S. Hispanic heritage learners' writing. In *Globalization and language in the Spanish-speaking world* (pp. 76-93). Palgrave Macmillan.

Flores-Ferrán, N. (2004). La expresión del sujeto en el español de Nueva York: el factor de la perseverancia. ASJU *International Journal of Basque Linguistics and Philology, 38*(1): 349-362*.*

Geeslin, K. L., & Gudmestad, A. (2011). Using sociolinguistic analyses of discourse-level features to expand research on L2 variation: Native and non-native contrasts in forms of Spanish subject expression. In L. Plonsky & M. Schierloh (Eds.), *Selected proceedings of the 2009 Second Language Research Forum: Diverse contributions to SLA* (pp. 16-30). Cascadilla Proceedings Project.

Gevers, J. (2018). Translingualism revisited: Language difference and hybridity in L2 writing. *Journal of Second Language Writing*, *40*, 73-83.

Givón, T. (1983). Topic continuity in discourse: The functional domain of switch-reference. In J. Haiman & P. Munro (Eds.), *Switch-reference and universal grammar* (pp. 51-81). John Benjamins.

Hochberg, J. (1986). Compensation for /s/ deletion in Puerto Rican Spanish. *Language*, *62*, 609-621.

Hurtado, L. M. (2005). Syntactic-semantic conditioning of subject expression in Colombian Spanish. *Hispania*, *88*(2), 335-348.

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.

Johnson, D. E. (2022). Rbrul, Package version 3.1.6.

Labov, William. (1984). *The intersection of sex and social factors in the course of language change*. Paper presented at NWAVE, Philadelphia.

LaCasse, D. (2019) The persistence of expression: Clusivity, partial co-reference, and socioeconomic differentiation of first-person plural subject pronoun expression in Spanish. *Studies in Hispanic and Lusophone Linguistics, 12*(1). DOI:10.1515/shll-2019-2002

Lastra, Y., & Martín Butragueño, P. (2015). Subject pronoun expression in oral Mexican Spanish. In A. M. Carvalho, R. Orozco. & N. L. Shin (Eds.), *Subject pronoun expression in Spanish. A cross-dialectal perspective* (pp. 39-57)*.* Georgetown University Press.

Linford, B. (2016). *The second-language development of dialect-specific morpho-syntactic variation in Spanish during study abroad.* ProQuest Dissertations and Theses.

Linford, B., & Shin, N. (2013). Lexical frequency effects on L2 Spanish subject pronoun expression. In J. Aaron, J. C. Amaro, G. Lord, & A. de Prada Pérez (Eds.), *Selected proceedings of the 16th Hispanic Linguistics Symposium.*

Martínez, G. (2007). Writing back and forth: The interplay of form and situation in heritage language composition. *Language Teaching Research*, *11*(1), 31-41.

Martínez Mira, M. (2009). Spanish heritage speakers in the Southwest: Factors contributing to the maintenance of the subjunctive in concessive clauses. *Spanish in Context, 6*(1), 105-126.

Michnowicz, J. (2015). Subject pronoun expression in contact with Maya in Yucatan Spanish. In *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. 101-119).

Mikulski, A., & Elola, I. (2011). Spanish heritage language learners' allocation of time to writing processes in English and Spanish. *Hispania*, 715-733.

Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: A case of morpho-syntactic convergence. *Bilingualism: Language and Cognition 7,* 125-142.

Montrul, S., & Rodríguez-Louro, C. (2006). Beyond the syntax of the null subject parameter: A look at the discourse-pragmatic distribution of null and overt subjects by L2 learners of Spanish. In L. Escobar and V. Torrens (Eds.). *The acquisition of syntax in romance languages* (pp. 401-418). John Benjamins.

Montrul, S., & Sánchez-Walker, N. (2015). Subject expression in bilingual school-age children in the United States. In *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. 231-247).

Orozco, R., & Guy, G. R. (2008). El uso variable de los pronombres sujetos: ¿Qué pasa en la costa Caribe colombiana? In M. Westmoreland & J. A. Thomas (Eds.), *Selected proceedings of the 4th Workshop on Spanish Sociolinguistics* (pp. 70-80). Cascadilla Proceedings Project.

Otheguy, R., & Zentella, A. C. (2012). *Spanish in New York: Language contact, dialectal leveling, and structural continuity*. Oxford University Press.

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

Paredes Silva, V. L. (1993). Subject omission and functional compensation: Evidence from written Brazilian Portuguese. *Language Variation and Change*, *5*, 35-49.

Roca, A. (1997). La realidad en el aula: Logros y expectativas en la enseñanza del español para estudiantes bilingües. In C. Colombi & F. Alarcón (Eds.), *La enseñanza del español a hispanohablantes. Praxis y teoría* (pp. 55-64)*.* Houghton Mifflin.

Romaine, S. (1982). *Socio-historical linguistics: Its status and methodology.* Cambridge University Press.

Shin, N. L. (2014). Grammatical complexification in Spanish in New York: 3sg pronoun expression and verbal ambiguity. *Language Variation and Change*, *26*(3), 303-330.

Shin, N. L. (2016). Acquiring constraints on morphosyntactic variation: Children's Spanish subject pronoun expression. *Journal of Child Language*, *43*(4), 914-947.

Shin, N. L., & Montes-Alcalá, C. (2014). El uso contextual del pronombre sujeto como factor predictivo de la influencia del inglés en el español en Nueva York. *Sociolinguistic Studies*, *8*(1), 85-110.

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 en otros contextos: cuestiones sociolingüísticas, políticas y pedagógicas* (pp. 111-136). Iberoamericana.

Shin, N. L., & Otheguy, R. (2013). Social class and gender impacting change in bilingual settings: Spanish subject pronoun use in New York. *Language in Society 42*, 429-452.

Silva-Corvalán, C. (1994). *Language contact and change.* Oxford University Press.

Silva-Corvalán, C. (2015). The acquisition of grammatical subjects by Spanish-English bilinguals. In *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. 211-230).

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism*, *1*(1), 1-33.

Spicer-Escalante, M. L. (2006). The use of verbal forms in the Spanish and English writing discourse of Spanish heritage speakers in the United States: A contrastive linguistic analysis. *Río Bravo, 2*(1): 13-24.

Spicer-Escalante, M. L. (2009). Análisis lingüístico de la escritura bilingüe (español-inglés) de los hablantes de español como lengua hereditaria en los Estados Unidos. *Estudios de Lingüística Aplicada*, *27*(50), 477-494.

Tagliamonte, S. (2012). *Variationist sociolinguistics: Change, observation, interpretation*. Wiley-Blackwell.

Thompson, G. L. (2015). Understanding the heritage language student: Proficiency and placement. *Journal of Hispanic Higher Education*, *14*(1), 82-96.

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 US southwest: A language in transition.* Iberoamericana/Vervuert.

Torres Cacoullos, R., & Travis, C. (2011). Testing convergence via code-switching: Priming and the structure of variable subject expression. *International Journal of Bilingualism*, *15*(3), 241-267.

Travis, C. (2007). Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change*, *19*, 101-135.

Travis, C., & Torres Cacoullos, R. (2012). What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. In Cognitive Linguistics, 23.4, 711-748.

Valdés, G. (2001). Heritage language students: Profiles and possibilities. In J. K. Peyton, D. A. Ranard, & S. McGinnis (Eds.), *Heritage languages in America: Preserving a national resource* (pp. 37-77). Center for Applied Linguistics/Delta Systems.

Valdés, G., Fairclough, M., Beaudrie, S., & Roca, A. (2016). *Innovative strategies for heritage language teaching: A practical guide for the classroom*. Georgetown University Press.

APPENDIX A
RESULTS FOR L1 MEXICAN SPANISH SPEAKERS

Table 1. Overall rates of overt 1st person SPE among L1 Mexican Spanish speakers in written production

Yo	66/553	12%
Nosotros	11/469	2%

Table 2. Hierarchy of linguistic constraints significantly influencing SPE among L1 speakers' written language

Factor	Weight	%	# Tokens	p Value
Switch Reference				p< 0.0001
First token	.75	25%	3/12	
Switch Reference	.48	10%	51/507	
Same Reference	.26	5%	24/475	
Range 49				
Grammatical Number				p< 0.0001
1st Person singular	.71	12%	66/553	
1st Person plural	.29	2%	11/469	
Range 42				
Reflexivity				p< 0.005
Non reflexive	.69	9%	74/840	
Reflexive	.31	2%	3/182	
Range 38				
Frequency				
Continuous +1 logodds 0.022				

Total N 1,022

Nagelkerke R2 .51

Intercept -4.249

Log.likelihood -223.273