To appear in *Colomina-Almiñana, J. & S. Sessarego (eds.). Patterns in Spanish: Structure, Context and Development. Amsterdam/Philadelphia: John Benjamins*

Title: How preschoolers acquire the null-overt contrast in Mexican Spanish: Evidence from production

Authors:
Hannah Forsythe, University of California Irvine, hforsyth@uci.edu
Daniel Greeson, Michigan State University
Cristina Schmitt, Michigan State University

Key words: acquisition of pro-drop, subject pronoun expression, pronoun resolution, learnability

Abstract: In many so-called true null subject languages, null and overt subject pronouns have contrasting referential preferences: null subjects tend to maintain reference to the preceding subject while overt pronominal subjects do not. We propose that children acquire this contrast by initially restricting their attention to 1st and 2nd person pronouns, whose intended referents are simpler to infer. Production data from Mexican Spanish shows that (i) the null/overt contrast is in principle acquirable from 1st and 2nd person subject pronouns in naturalistic input, (ii) children's production of 1st and 2nd person subjects is conditioned by this contrast early in life (approximately 4½), and (iii) the contrast is generalized to the 3rd person slightly later (by 6).

1. Introduction

One of the first things that anglophone learners of Spanish notice is the fact that subjects can be optionally omitted without causing ungrammaticality, as in (1).

(1) *Primero Ø estábamos comiendo*.
First Ø were eating.
("First we were eating.")

While the alternation between null and overt subjects is shared by many languages, the precise distribution and frequency of use of these forms varies across particular language varieties. The null/overt alternation has long been studied from the perspective of sociolinguistic variation (Bayley & Pease-Alvarez 1997, Flores-Ferrán 2007, Lastra & Butragueño 2015, Michnowicz 2015, Otheguy *et al*. 2007, 2010; Otheguy & Zentella 2011, Shin & Otheguy 2009), parametric variation (Barbosa 2009, 2011 and references therein; Toribio 2000), processing (Carminati 2002, Filiaci *et al*. 2013, Jegerski *et al*. 2011, Keating *et al*. 2011, 2016), and second-language acquisition (Montrul 2004, Pérez-Leroux and Glass 1999). Yet only recently has attention turned toward the question of how the variability between null and overt subjects is acquired (Shin 2012, 2016; Shin & Cairns 2012). This is an important question because, in order for such a widespread, stable alternation to arise in the first place, it must be acquirable by children learning their first language. The question we address in this paper is how children acquire this knowledge about the distribution of null and overt subjects.

Acquiring the null/overt contrast involves two steps. First, the learner must realize that her language licenses both null and overt subjects in tensed clauses, without overextending this alternation to nonfinite clauses or to operator constructions such as relative clauses (2), where the gap receives a different analysis. Language acquisition researchers have long known that this step is accomplished early (Austin *et al*. 1997, Grinstead 2004). In fact, this may be the preferred parameter setting in child language, since even children acquiring non-pro-drop languages pass through a stage in which they produce and accept null subjects alongside overt subjects (Hyams 1986/2012, Orfitelli & Hyams 2012, Valian 1990, Yang 2002, among many others).

(2) *el muchacho que (*él) saludó a Ana*
the boy that (*he) greeted Ana

Second, in those environments where null and overt subjects alternate, the learner must discover the semantic and pragmatic conditions that govern their use. This second step has received much less attention from the L1 research literature and it is where we focus our efforts. We focus specifically on the conditions governing the choice between an overt subject pronoun versus a null pronoun in the dialect of Spanish spoken in Mexico City, Mexico. How do children acquiring this variety of Spanish learn to correctly produce and interpret this contrast?

We begin with a description of the knowledge that must be acquired and an explanation of the challenge that this presents to the learner (Section 2). In Section 3, we present a specific hypothesis about how children solve this problem: namely, that they begin by initially restricting their attention to 1st and 2nd person pronominal subjects. We identify three questions that must be answered in order to support this hypothesis: (i) Is the null/overt distinction in principle acquirable from the 1st and 2nd person alone?; (ii) Do children show knowledge of this distinction

in their own production of 1st and 2nd person pronominal subjects?; and (iii) Do children generalize that knowledge to 3rd person pronominal subjects? In Section 4 we present a brief review of the currently available acquisition results. Section 5 presents a corpus study addressing questions (i)-(iii). Results are consistent with the proposal that children acquire the conditions on subject pronoun realization within the domain of 1st and 2nd person subjects first, before generalizing to the 3rd person, and that this process occurs by age 6. The paper concludes with a summary of our findings, their limitations, and their impact. In particular, this study provides the earliest evidence to date of children's sensitivity to the null/overt contrast.

2. Defining the learning problem

An extensive literature reveals that adult speakers of pro-drop languages tend to associate the null subject with the subject of the preceding clause but have either the opposite bias or no bias at all when interpreting the overt pronoun (Alonso-Ovalle *et al*. 2002, Carminati 2002, de la Fuente 2015, Papadopoulou et al. 2015). First, we describe how this contrast is played out in the production and comprehension of so-called "consistent" null subject languages like Spanish (Barbosa 2009, among others). Next, we illustrate how acquiring this contrast presents a challenge, since it requires the child to engage in the complicated task of pronoun resolution.

In Spanish, listeners presented with grammatically ambiguous examples such as (3) (adapted from de la Fuente 2015) tend to interpret the null subject (3a) as referring to the preceding subject antecedent *Juan,* whereas they have no such interpretive bias for the overt subject (3b). The subject-antecedent reading is commonly referred to as the "same-reference" reading because the same referent is referred to in both subject positions; all other readings tend to be grouped under the term "switch-reference" (Otheguy & Zentella 2011). In addition to judgments of ambiguous sentences like (3), online reading times and offline acceptability judgments of pragmatically disambiguated sentences show that the null subject is easier to process and more acceptable when presented in same-reference contexts, relative to switch-reference contexts, and vice-versa for the overt pronoun (Carminati 2002, Filiaci *et al*. 2013, Jegerski *et al*. 2011, Keating *et al*. 2011, 2016).

(3) a. *Juan llamó a Pedro cuando* ***Ø*** *estaba en casa.*
Juan called Pedro when __ was at home.
b. *Juan llamó a Pedro cuando* ***él*** *estaba en casa.*
Juan called Pedro when **he** was at home.

Sociolinguistic studies reveal the same biases in production as in comprehension. Across a host of language varieties, speakers tend to produce fewer overt pronominal subjects in same-reference contexts relative to switch-reference contexts; in fact, this is consistently the strongest factor conditioning subject pronoun expression next to its person/number features (Otheguy & Zentella 2012, Carvalho, Orozco & Shin 2015 and references therein).

It is important to underscore that this contrast is probabilistic rather than categorical. Even though the null subject in (3a) favors the same-reference reading more strongly than does the overt subject in (3b), both readings are available for both pronouns. Moreover, because pronoun interpretation is sensitive to multiple discourse-pragmatic factors, the reading that adults ultimately prefer can be biased either towards the same-reference reading or the switch-reference

reading, depending on context. For example, were it made known to the listener that Juan was calling Pedro *in order to find out where Pedro was at that moment*, then the switch-reference reading would be favored over the same-reference reading in both (3a) and (3b), i.e., the listener would likely infer that Juan called Pedro when Pedro (not Juan) was at home, regardless of which pronoun the speaker used. Nevertheless, the relative contrast between (3a) and (3b) would still stand, with the overt variant favoring the switch-reference reading *even more* strongly than the null variant.

The task for the L1 learner, then, is to associate null subjects with an increase in probability of a same-reference reading, relative to overt pronominal subjects (rather than acquiring a categorical rule). To succeed at this task, the child must track the relative rates of pronoun realization in same-reference contexts versus switch-reference contexts in her input.

The problem that this task presents to the learner is that the correlation between same- and switch-reference contexts and null and overt pronouns can only be established (i.e. in the child's mind) once the pronoun's referent has already been identified. That is, if a child is to verify that a null subject like (3a) in her input indicates reference to the preceding subject, or that an overt subject like (3b) indicates reference to a non-subject antecedent, she must know in advance what the intended referent of each pronoun is. However, pronoun resolution is a complex task, depending on a multitude of grammatical and discourse factors. Relevant grammatical cues to pronoun reference include the pronoun's person, number, and gender features, as well as syntactic constraints like Principle B[1], and discourse cues include information from the context, such as what the current topic of conversation is, which discourse referents are most prominent, how events and states are pragmatically related to one another, and relevant real-world knowledge about these referents and eventualities, to name just a few.

The complexity of determining pronoun reference can be illustrated by any discourse situation that offers more than one sufficiently salient referent with the right grammatical features. For example, in the context of a story about Juan and Pedro hatching a plan to skip school, being discovered by the principal, and being disciplined by their fathers, a speaker could utter (3a) or (3b) using either the null or overt pronoun; or as illustrated in (4) the speaker could use any number of more semantically restricted noun phrases, such as a proper name or a modified noun phrase. Because pronouns are semantically underspecified, they could in principle refer to *any* of the characters in this story, in contrast to the other more semantically restricted DP subjects, as illustrated in (5).

(4) Context: A story about Juan and Pedro skipping school, getting caught by the (male) principal, and being disciplined by both of their fathers.
Juan llamó a Pedro cuando {*Juan/el niño travieso /ø/él*} *estaba en casa.*
Juan called Pedro when Juan/the naughty boy/*pro*/he was at home.

(5) Potential referents of the subject DP in (4)
a. *Juan*: {Juan}

[1] Principle B refers to the constraint that prohibits pronoun coreference with antecedents in certain structural positions. Roughly, the principle states that pronouns must be free in their domain, where "domain" roughly corresponds to a clause. This explains the contrast between (i), where the pronoun *him* cannot refer to the subject of its clause *Bill*, and (ii), where the pronoun *him* can refer to *Bill* when *Bill* is the subject of a higher clause.
(i.) Bill saw him. *him* ≠ Bill
(ii.) Bill said that John saw him. *him* = Bill

b. *el niño travieso*: {Juan, Pedro}
c. null subject: {Juan, Pedro, principal, Juan's father, Pedro's father}
d. overt subject: {Juan, Pedro, principal, Juan's father, Pedro's father}

What this example shows is not that sentences with null and overt pronominal subjects are impossible to interpret, but that doing so requires an inferential process that should not be taken for granted—the child must consider the set of potential referents and narrow it down to the referent she believes is intended by the speaker. Then—and only then—can she identify the sentence as a same-reference or switch-reference token and create the necessary association between that interpretation and the use of a null or overt subject pronoun.

3. Proposed learning path

In this section we present a specific proposal for how children might establish the probabilistic association between null and overt pronouns, and same- and switch-reference, respectively. In the remainder of the paper we present evidence that this proposal is consistent with both the input and production of children acquiring Spanish.

One possible learning strategy might be to delay acquisition of the null/overt contrast until the child has acquired enough knowledge of pragmatic cues, world knowledge, discourse relations, and other discourse cues to be able to resolve a sufficiently large portion of the subject pronouns in her input. While we agree that this learning path is possible in principle, we argue against it for a conceptual reason. First of all, this explanation forces us to assume that children initially ignore relevant information present in their input. Second, it merely pushes the learning problem back one step. If acquisition of the null/overt distinction depends on prior acquisition of cues to pronoun resolution, how did children acquire those cues in the first place?

We propose that a much simpler learning path becomes available if children initially restrict their attention to a subset of the subjects in their input: 1st and 2nd person pronouns. These pronouns refer to the speaker and addressee, respectively, and therefore present a much more restricted set of potential referents at the outset. Barring special circumstances[2], 1st person pronouns refer directly to the speaker him or herself and are therefore completely unambiguous. Second person pronouns merely require the child to infer who the intended addressee is, which in one-on-one speech is the child herself. Furthermore these pronouns do not compete with a wide range of definite descriptions as 3rd person pronouns do (Pedro, he, the new neighbor, the principal, etc). Because of these restrictions, it should be much simpler to identify the intended referent of 1st and 2nd person pronouns and thence to identify same- and switch-reference contexts like (6) below, compared to similar contexts with 3rd person pronouns like (3) above.

(6) a. *Tú llamaste a Pedro cuando* ***Ø*** *estabas en casa.*
You called Pedro when __ were at home.
b. *Pedro te llamó cuando* ***tú*** *estabas en casa.*

[2] Kratzer (2009) and Heim (2005) identify cases of "fake indexicals" where 1st and 2nd person pronouns do not refer directly to the speaker or hearer, but instead receive a bound variable reading. For example, the sentence *Only I eat what I cook*, implies that there is no other person *x* such that *x* eats what *x* cooks. To produce this meaning, the embedded pronoun must be referentially dependent on the matrix pronoun.

Pedro called you when **you** were at home.

Once the child has formed an association between overt 1st and 2nd person pronominal subjects and an increased probability of switch-reference, she can then generalize this to *all* subject personal pronouns. We assume that this generalization is justified because 1st, 2nd and 3rd person pronouns fall into the same paradigm, differing only in the phi-features that they carry (Charnavel 2015). In sum, once the child learns that 1st and 2nd person null pronouns are statistically more likely to occur with preceding subject referents and overt variants with other referents, and that this applies to 3rd person pronouns in their input as well, the child can use the same variations in production and comprehension of all subject pronouns. Our proposal is summarized in (7).

(7) Proposed learning path
   a. Step 1, Association: Track the realization of 1st and 2nd person pronominal subjects in same- and switch-reference contexts and associate switch-reference readings with an increased probability of overt realization.
   b. Step 2, Generalization: Generalize this association to the production/comprehension of *all* subject pronouns—1st, 2nd and 3rd person.

To show that this proposal meets the challenge of empirical adequacy, we must at minimum answer the following three questions:

Q1: Is the null/overt contrast in principle acquirable from the distribution of 1st and 2nd person pronouns in children's input?
Q2: Do children show sensitivity to this contrast in their own production of 1st and 2nd person pronouns?
Q3: Do children generalize this contrast to the production/comprehension of 3rd person pronominal subjects *after* they have acquired this contrast in the domain of 1st and 2nd person?

We leave it to future work to show that the alternative strategy (namely, delaying acquisition of the null/overt distinction until the child has mastered pronoun resolution) is empirically inadequate, in addition to being conceptually undesirable. Before addressing these questions with data of our own, we will review what is currently known from existing data in languages that display the null/overt contrast.

4. Acquisition background

What we know so far about the developmental path of acquiring the distribution of null and overt subjects comes from studies of spontaneous production, felicity judgments, and pronoun resolution tasks in a variety of languages. We concentrate here on pro-drop languages with rich agreement. Children learning these languages must associate same-reference contexts with a decreased rate of overt pronoun realization, relative to switch-reference contexts, although the strength of this association may vary across language varieties (Filiaci *et al*. 2013, de la Fuente 2015).

Production studies from the acquisition literature find that children acquiring Spanish and Catalán begin to produce overt subjects before age 2 (Bel 2003, Grinstead 2004). However, overt pronominal subjects remain infrequent. Sociolinguistic studies that include child subjects corroborate the low frequency of overt pronominal subjects well into middle and late childhood. In Mexican Spanish, first graders (ages 6-7) overtly realize pronouns at a rate of 6-8% (Shin 2012, 2016), and this rate gradually increases to about 10% in 5th grade (ages 10-11)—far less than the 18-22% rate found among Mexican adults (Lastra & Butragueño 2015, Shin & Erker 2015, Shin & Otheguy 2013). Despite being infrequent, however, overt pronouns are not randomly distributed. Shin (2016) finds that overt realization is positively associated with switch-reference contexts among even the youngest age group (6-7 years).

Felicity judgment tasks show that children are much more accepting than adults of null subjects in switch-reference contexts as well as of overt subjects in same-reference contexts. Sorace *et al*. (2009) asked bilingual children, monolingual children, and adult speakers of Italian to judge which of two speakers produced a better description of an event, with one speaker producing a null subject (8a) and the other an overt subject (8b). Crucially, the event was manipulated such that either *Minnie* (a same reference context) or another person (a switch-reference context) had fallen. Monolingual Italian-acquiring children ages 6-7 and 8–10 were just as likely as adults to prefer the overt pronoun in the switch-reference condition; however, only the 8–10-year-olds were just as likely as adults to prefer the null pronoun in the same-reference condition. Bilinguals were overall less adult-like than their monolingual peers—even those whose other language was Spanish, another pro-drop language.

(8) a. *Minnie ha detto che **ø** è caduta.* [Sorace *et al*. (2009)]
Minnie has said that (**she**) has fallen.
b. *Minnie ha detto che **lei** è caduta.*
Minnie has said that **she** has fallen.

Shin & Cairns (2012) used a similar methodology to probe the preferences of Mexican Spanish-acquiring children from ages 6 to 15. They found a qualitatively similar but quantitatively slower developmental trajectory. In the switch-reference condition, 8–10-year-olds showed a significant preference for an overt subject, although the strength of this preference did not match adults until age 14–15. In the same-reference condition, even the oldest children failed to show a significant preference for the null subject. Finally, studies among L2 learners have shown an even more pronounced version of this pattern of non-target behavior (Belletti *et al*. 2007, Montrul 2004, 2011; White 2011, a.o.). These results suggest a very protracted developmental path; however, these tasks may underestimate what children actually know about the null/overt contrast at earlier stages in development.

The tasks employed in these studies are cognitively complex, requiring the listener to hold two utterances in short-term memory while making a metalinguistic judgment about them. To lighten the cognitive load of the task in order to better get at children's knowledge, Papadopoulou *et al*. (2015) used a less taxing method to probe Greek-acquiring children's sensitivity to the null/overt distinction from ages 6 to 11 and found more adult-like performance. Participants listened to a sentence like (9a) while simultaneously viewing a picture that corresponded to either a same-reference interpretation (*pro* = the old man), or one of two switch-reference interpretations (object interpretation: *pro* = his grandchild; other interpretation: *pro* = another person). At the end of each sentence, the task was simply to judge whether the sentence

matched the picture. Additionally, the audio was self-paced: participants would press the spacebar to hear each phrase of the sentence, and their listening times were measured. A second experiment used the same methodology with overt pronouns, as in (9b).

(9) a. *O papús millúse δinatá ston egonó tu ótan* ***ø*** *δjávaze ena vivlío.*
The old-man spoke-3SG loudly to his grandchild when **(he)** read-3SG a book.
b. *I jajá xerétise tin kipéla ótan* ***aftí*** *pernúse to δromo.*
The old-lady greeted-3SG the girl when **she** crossed-3SG the street.

Like adults, children of all ages accepted the same-reference reading of the null subject nearly all the time, they accepted the object reading less often, and they accepted the "other" reading even less often. In the overt subject experiment, children of all ages were like adults in accepting the object reading most of the time and the "other" reading less often, but unlike adults they over-accepted the same-reference reading of the overt pronoun until ages 10-11. Nevertheless, listening times showed that even in this condition the younger children still processed the object reading faster than the same-reference reading.

Summing up, in languages with a similar null/overt contrast to Spanish, children as young as 6 show sensitivity to the association between subject pronoun expression and pronominal reference in their own production as well as their interpretation of ambiguous pronouns. Unfortunately, not much is known about the trajectory of development before this age, after the onset of overt subjects at around 2. We help fill this gap by contributing production and comprehension data for Spanish-acquiring children ages 3 to 6.

5. Corpus study: subject realization

Our proposal claims that children acquire the conditions governing the null/overt contrast from the distribution of 1st and 2nd person pronominal subjects in their input. The first question to address is whether the input even provides the necessary information—after all, 1st and 2nd person pronouns have a different discourse function and therefore may not show the same patterns. We do this by examining naturalistic child-directed speech by monolingual speakers of Spanish native to Mexico City. The second question is whether children's own production demonstrates awareness of this contrast, and if so, whether it is acquired first in the domain of 1st and 2nd person subjects before being generalized to the 3rd person.

We used data from the Schmitt–Miller corpus (collected in Mexico City in 2008), comprising approximately 649,000 words of spontaneous speech from 11 low-SES and 14 high-SES child-caretaker dyads (child ages 1;6-5;11). Caretakers were recorded during sessions playing with their children, as well as one session each chatting with another adult.

5.1 Subject characteristics

We examined a subset of 103,249 words taken from the speech of five mother-child dyads, whose characteristics are summarized in Figure 1.

Child	Age	Mean Length of Utterance	Mother Word Count	Child Word Count

YGSZ	3;9	3.652	9,608	10,910
YBM	4;8	3.993	11,054	8,373
OMJ	5;2	3.87	11,934	7,314
KUC	4;1	4.522	11,721	9,393
JRC	5;11	3.735	13,114	10,548
	Mean: 4;9	Mean: 4.02	Total: 93,024	Total: 69,552

Fig. 1: Subject characteristics of 5 Mexican Spanish-speaking dyads from the Schmitt-Miller corpus

5.2. Hypotheses and predictions

The learning path hypothesized in Section 3 makes three predictions. First, if the null/overt distinction is learnable from 1st and 2nd person pronouns in the input, we expect that mothers in our sample will produce significantly more overt 1st and 2nd person pronominal subjects in switch-reference contexts than they do in same-reference contexts.

Second, if children become attuned to this statistical pattern at some time before 6 years of age, we expect that some portion of the children in our sample will show the same pattern in their own production of 1st and 2nd person pronominal subjects.

Finally, if children generalize constraints from 1st and 2nd person to the 3rd person (rather than the other way around), then we would predict that a slightly older group of children in our sample will show this pattern in their production of 3rd person pronominal subjects as well.

5.3. Coding methods

From our sample, we extracted all animate, pronominal subjects of tensed verbs that were preceded by at least one other clause in the same turn, defined as the longest uninterrupted string of speech by one person. These subjects were then coded for form (null, overt) and reference (same, switch) following the definition in Otheguy and Zentella (2011, appendix), which we illustrate below in 5.3.2.

5.3.1 Exclusions

We excluded the subjects of tensed verbs in several conditions, as described in (10). Following Otheguy and Zentella (2011), we excluded tensed verbs in conditions (a-b); following Shin (2016) we excluded tensed verbs in conditions (c-d). We excluded all subjects of dative-experiencer predicates (e), since the surface/syntactic subject is not necessarily the same as the underlying/semantic subject in these cases. We excluded generic subjects (f) because these uses are nearly exclusively null. We also excluded any pre-authored material, such as passages from books, song lyrics, or other recited material, since it corresponds to another register and/or speaker.

(10) Exclusions

a. frozen expressions: *quién sabe, ¿sabes?, ¿ves?, ¿ya viste?*, and so on.

b. Subject relatives: *El niño que* ***Ø*** *toca el piano* ("The boy that plays the piano")
c. subjects of imperatives: *Dime* ***Ø*** *cuántos años tienes*
d. the phrase *no sé* ("I don't know")
e. Dative experiencer–predicate constructions: *me gusta eso* ("I like that;" literally, "that pleases me"), *le encanta la playa* ("he loves the beach"; literally, "the beach charms him"), etc.
f. Generic references to "them" or "people," such as *dicen que…* ("they say that…") and *en la escuela me dieron de comer* ("at school (they) gave me something to eat")
g. Passages read from books, song lyrics, nursery rhymes, etc.

5.3.2 Definition of same-reference and switch-reference

We coded a subject as *same-reference* if it referred to the same referent as the subject of the preceding tensed verb within the same speaker turn. In the hypothetical example (11), the null subject of *tienes sueño* ('you're sleepy') refers to the same entity as the subject of *estás bostezando* ('you are yawning') and would therefore be coded as a same-reference token. We coded a subject as *switch-reference* if it referred to a non-subject argument of the preceding tensed verb, to an entirely new referent never mentioned before, or to a referent that had been mentioned before, but in a preceding turn (see also example (13), below, for our operational definition of what counts as an eligible preceding verb). In (11), for example, the null subject of *estás bostezando* ('you are yawning') would be coded as a switch-reference token because it refers to the same referent as the preceding object of *veo* ('I see'). The subject of *vamos* ('we'll leave') would also be coded as switch-reference because it does not match in person and number features with the subject of the preceding verb *tienes sueño* ('you're sleepy'). Finally, we excluded any subjects that were not preceded by another tensed verb in the same turn. For example, the first subject in (11), *yo* ('I') has no preceding subject with which to maintain or switch reference, so it is excluded as a token, though it still serves as a suitable "trigger" for coding the subject of the following verb, *estás*.

(11) ***Yo*** *te veo que* ***Ø*** *estás bostezando. Si* ***Ø*** *tienes sueño,* ***Ø*** *nos vamos.*
I see you that (**you**) are yawning. If (**you**)'re sleepy, (**we**)'ll leave.

In some cases, a subject pronoun was rendered trivially switch-reference by the properties of the preceding subject—usually because the preceding subject was non-referential[3]. In these cases, the preceding subject was skipped and the preceding subject before that was used

[3] We use the term "non-referential subjects" to refer to expletive subjects (i), impersonals (ii), and *wh*-pronouns (iii). In these examples, the subject pronoun *ella* ('she') fails to refer to the same entity as the preceding subject, not because these are genuine cases of switch-reference, but because the preceding subject simply does not refer to an entity in the first place.

(i) *Cuando* ***ø*** *llueve, ¿va a llevar ella su sombrero?*
When (**it**) rains, is she going to take her hat?
(ii) ***ø*** *se dice que ella va a llevar su sombrero.*
(**They**) say that she is going to take her hat.
(iii) *¿****Quién*** *dijo que ella va a llevar su sombrero?*
Who says that she is going to take her hat?

instead. The following criteria were used to determine whether the preceding subject was an eligible "trigger" for coding same- and switch-reference:

(12) Eligibility of the preceding subject for coding same- and switch-reference

a. Subjects of imperatives and inanimate subjects were considered eligible triggers, even though they were not themselves coded as tokens. The exception to this rule was set imperative phrases like *mira*, *oye*, etc.
b. Subjects that do not refer to individuals were considered ineligible (see footnote 3 for some examples). This includes: subjects of frozen expressions, clausal subjects, and expletive subjects (*es que*, *lo que quiero es*, etc.) and weather verb subjects.
c. We excluded *se*-impersonal/passives as well as 3rd person plural subject like *dicen que* ('they say that').
d. We excluded traces of wh-operators.
e. Contra Otheguy & Zentella (2011) we did not consider the grammatical subjects of experiencer-predicate constructions (ex. *me gusta el agua*, literally 'water pleases me') or the subject of presentational *haber* (ex. *habían tres gatos*, 'there were three cats') to be eligible triggers.

5.3.3 Inter-rater reliability

Two different raters coded the form (null, overt) and reference (same-reference, switch-reference) of all subject personal pronouns produced by the children and mothers in our sample. To calculate inter-rater reliability, both raters independently coded a subset consisting of 6 transcripts (comprising 18% of the data) and their ratings for each token were compared. Inter-rater reliability was 96.6% (κ=.84) for pronominal form and 95.7% (κ=.91) for reference.

5.3.4. Examples

Data from our corpus analyses illustrate that both null and overt subjects can occur in same- and switch-reference contexts in children's naturalistic input. In example (13), produced by the mother YBM (4;5) we find two same-reference tokens that illustrate this. The speaker's turn begins with a null subject, and the subjects of both the following verbs (*es, tiene*) maintain reference to the same referent (a dancer on a tightrope). In one case, the mother uses an overt subject (*ella*) and in the second she uses a null subject.

(13) *ø estuvo a punto de caerse,*
(She) was about to fall,
*pero no, porque {**ella**} es una experta bailarina*
but no, because she is an expert dancer
*y {**ø**} tiene todo el equilibrio para poder bailar en una cuerda floja!*
and (she) has all the balance to be able to dance on a tightrope!

The same is true for switch-reference tokens, as illustrated in (14)-(15), from the same mother. The speaker begins her turn with an imperative, whose (null) subject refers to the child, then she switches reference to herself using an overt pronoun (*yo*). Finally, in (15), the mother begins her turn referring to herself (with a null subject) and switches reference to her daughter, this time using a null pronoun.

(14) Context: Mother encourages daughter to sing a lullaby to some lions.
Cántasela
(you) sing it to them,
Para que {***yo***}*me siente un ratito*
so that I can sit down for a sec

(15) Context: Mother pretending to be a doctor prescribing "luneta" candies.
Okey, entonces ø ya no le doy esta receta
Okay, then (I) won't give you this prescription
y ya {**ø**} *no va a comer lunetas nunca más en la vida*
and now (you) will never eat lunetas ever again in your life.

These examples confirm that in child-directed speech, as in adult-directed speech, there is not a categorical requirement to use null subjects in same-reference contexts or to use overt subjects in switch-reference contexts. Rather, if these associations are attested in child-directed speech, they will be probabilistic in nature.

5.4. Results: Overall rate of overt pronoun realization

Mothers and children in our sample produced overt pronouns at similar rates to each other (10.2% for mothers, 9.7% for children). These rates differed somewhat from subject pronoun expression rates found in other studies (see Figure 2 for comparison), which could be due to differences in coding decisions as well as differences in the nature of the interactions that were recorded.

	% overt pronouns
Mothers (this study)	10.2% (169/1,482)
Adult-directed speech, Mexico City (Lastra & Butragueño 2015)	21.7% (443/2,040)
Children ages 3-6 (this study)	9.7% (104/970)
Children ages 6-7, Querétaro & Oaxaca (Shin 2016)	8% (148/1,845)

Fig. 2 Rates of overt pronoun use

The child-directed speech produced by the mothers in our sample contains fewer overt pronouns than adult-directed speech from Mexico City as reported by Lastra & Butragueño (2015) (10.2% versus 21.7%). This may be because child-directed speech tends to reference familiar items, potentially increasing the rate of null pronouns. An additional contributing factor is a difference in the sampling method: because we included only tokens preceded by another clause within the turn, we excluded turn-initial pronouns, potentially under-sampling the overt

variant. When turn-initial pronouns are included in this count, mothers' overt pronoun rates increase slightly to 12.7% (540/4,320) and children's to 12.6% (417/3,314).

On the other hand, the children in our sample produced overt pronouns more often than the slightly older children from other regions in Mexico reported by Shin (2016) (9.7% versus 8%). This difference may be due to the nature of the interactions between speakers. Data from Shin (2016) is mainly from narrative contexts, which tend to feature a single referent to whom a narrator refers multiple times. In other words, narratives are dominated by same-reference segments, thereby encouraging more null subjects. In contrast, the Schmitt-Miller corpus recorded mostly one-on-one interactions between parent and child, which may have led to fewer same-reference segments, and which may also have encouraged children to assimilate to their mothers' input, leading to more overt subjects.

5.5. Results: Pronoun realization in same/switch-reference contexts

The learning path proposed in Section 3 depends upon the assumption that children's input provides the necessary statistical information to acquire the null/overt distinction from 1st and 2nd person subject pronouns alone. Hence, the first prediction to test is whether 1st and 2nd person subject pronouns in child-directed speech are indeed overtly realized more often in switch-reference compared to same-reference contexts. Figure 3 shows the rate and frequency of overt pronouns in mother's speech (left) and children's speech (right) in each of these contexts, across 1st, 2nd and 3rd person domains. To test the first prediction, we employed separate chi-squared tests of proportion to compare the rate of overt pronoun production across same-reference and switch-reference contexts in mother's speech, for each of the three personal pronouns. For mothers, the effect of reference was significant for 1st and 2nd but not 3rd person subjects. That is, mothers produced significantly more overt pronominal subjects in switch-reference contexts relative to same-reference contexts in the 1st ($\chi(1) = 7.54$, $p < 0.007$) and 2nd person ($\chi(1) = 16.903$, $p < 0.001$), but not in the 3rd person ($\chi(1) = 0.15$, $p = 0.70$). Thus, the statistical contrast between null and overt pronominal subjects is not only present in the 1st and 2nd person, it may be stronger there.

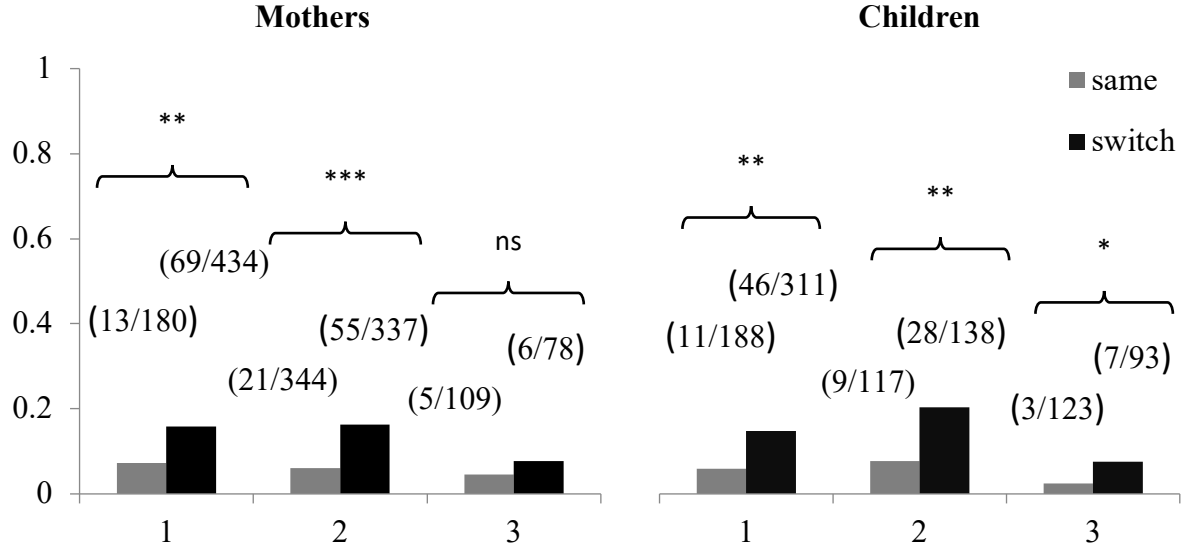

Fig. 3: Rate of overt 1st, 2nd, and 3rd person pronouns appearing in same- and switch-reference contexts in the speech of mothers (left) and children (right).

Given that 1st and 2nd person pronominal subjects seem to provide a robust statistical signal, the second prediction to test is whether children are sensitive to this signal and reflect it in their own production of 1st and 2nd person pronouns. To test this, we ran the same chi-squared tests of proportion on children's production of overt pronominal subjects. Children produced significantly more overt pronominal subjects in switch-reference contexts relative to same-reference contexts across both 1st and 2nd persons (1st person: $\chi(1) = 8.40$, $p < 0.004$; 2nd person: $\chi(1) = 7.71$, $p < 0.006$). Thus, this sample provides evidence that even children under 6 are sensitive to the null/overt contrast.

The third prediction to test is whether children generalize this distinction to the 3rd person. A chi-squared test of proportions finds a significant difference in the realization of 3rd person pronouns across same- and switch-reference contexts ($\chi(1) = 5.48$, $p = 0.019$), indicating that, as a group, these children have generalized the null/overt distinction to the 3rd person as well. Given that this difference was not significant in mothers' speech, this would seem to indicate that children have inferred a statistical pattern that is actually absent in the input. However, this conclusion is not warranted. The failure to reach statistical significance does not constitute evidence that no such difference exists, since a larger sample of the input may reveal differences that our sample failed to detect. Rather, we take the absence of a significant difference in mothers' realization of 3rd person pronominal subjects as an indication that the statistical signal provided by the input may be *weaker* in the 3rd person than it is in the 1st and 2nd persons.

In order to establish that children generalize knowledge acquired in the 1st and 2nd person to the domain of the 3rd person, rather than the other way around, we need evidence that the null/overt distinction is acquired *earlier* in the 1st and 2nd person. Figure 4 provides a more detailed view of the developmental trajectory of the null/overt contrast across individual children, separating 1st and 2nd person from 3rd person. All children in our sample produced overt pronouns more often in switch-reference contexts compared to same-reference contexts, but the difference reached significance at a younger age in the 1st and 2nd person. One-sided chi-squared tests of proportion for each individual child revealed significantly more overt 1st and 2nd person pronouns in switch- versus same-reference contexts for YBM (4;8, $\chi(1) = 6.93$, $p = 0.004$) and OMJ (5;1, $\chi(1) = 3.77$, $p = 0.026$) and a marginal difference for the oldest child, JRC (5;11, $\chi(1) = 1.97$, $p = 0.08$). For 3rd person pronominal subjects, however, this difference was significant only in the speech of the oldest child (JRC, 5;11, $\chi(1) = 2.87$, $p = 0.045$). This suggests that children first acquire the null/overt contrast as it applies to 1st and 2nd person pronominal subjects *before* generalizing to the 3rd person.

If this sample of 5 children is representative of acquisition more generally, then it suggests that typically developing children begin to acquire the null/overt contrast around age 4 ½. However, a larger sample, preferably longitudinal in nature, would be required to make this claim more confidently.

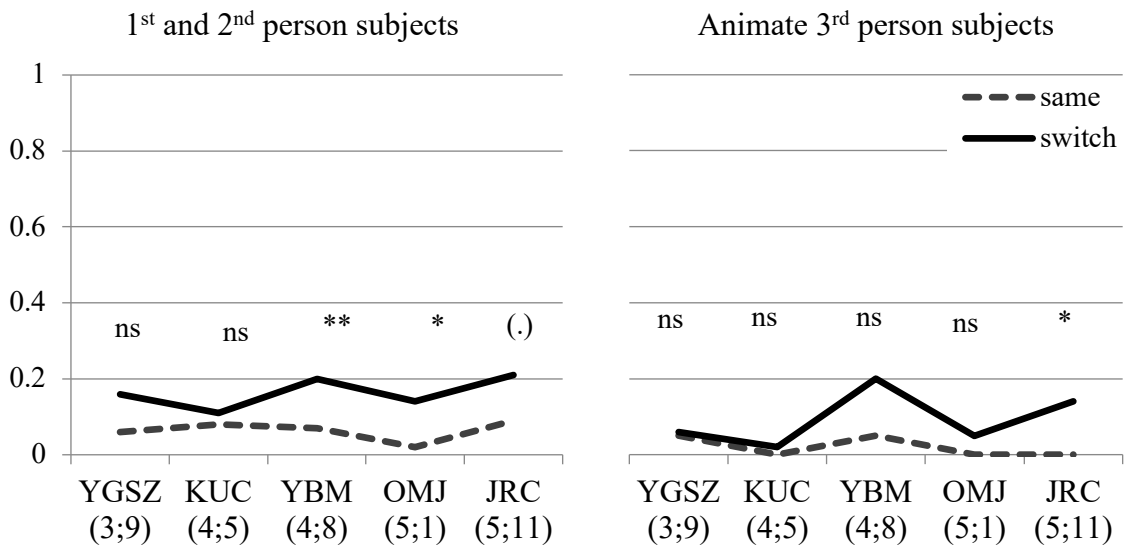

Fig 4: Rate of overt pronominal subjects produced by individual children, in same-reference and switch-reference contexts. Left: 1st and 2nd person pronominal subjects. Right: animate 3rd person pronominal subjects.

5.5 General discussion and study limitations

Our proposal generated three predictions that are supported by the production data in our sample of naturalistic speech. First, analysis of the input reveals that the null/overt distinction is in principle acquirable from 1st and 2nd pronouns alone: mothers produce significantly more overt pronominal subjects in switch-reference contexts relative to same-reference contexts. In fact, this statistical signal may even be stronger in the 1st and 2nd person than it is in the 3rd person, where the same contrast fails to reach significance. Assuming that the input produced by these mothers can be generalized to caretaker speech at similar developmental stages, then this would suggest that children not only *can* learn the null/overt distinction from 1st and 2nd person subjects, it may actually be *easier* to do so.

Second, children's own production reveals that they have picked up on this statistical signal early in life: they too produce significantly more overt 1st and 2nd person pronouns in switch-reference compared to same-reference contexts. Thirdly, children's production of 3rd person subject pronouns shows the same statistical pattern, indicating that they have generalized the null/overt contrast to apply to all persons. Furthermore, the individual-level analysis reveals earlier acquisition of this distinction in the 1st and 2nd persons (between 4;5 and 4;8), relative to the 3rd person (by 5;11), indicating that the generalization is made from 1st and 2nd persons to the 3rd person, rather than running in the other direction.

Our proposal makes one prediction that we have not tested here. If children generalize knowledge of the null/overt contrast from 1st and 2nd person to the 3rd person, then this knowledge should be available to them in the comprehension of 3rd person pronominal subjects as well—not just in production. If children acquire the null/overt contrast around age 4 ½, as our individual-level analysis suggests, then we would expect children of that age to comprehend the contrast between null and overt 3rd person subjects at that same age. Interestingly, comprehension data from Forsythe (2017) shows that children age 4;7-6;4, but not younger, use

the null/overt contrast to help resolve grammatically ambiguous 3rd person subject pronouns (see also Forsythe, Greeson & Schmitt 2019).

There are two main limitations to this study: its small sample size and its limited comparability with the sociolinguistic literature. The overall number of tokens on which this study is based is comparable to that reported in other studies on the spontaneous production of null and overt pronouns (ex. Lastra & Butragueño 2015, 6-7-year-olds in Shin 2016). Nevertheless, the number of individuals involved is smaller, especially considering the wide age range of the children in our sample (3;9-5;11). Currently, we are working to increase our sample size. A second limitation is that our sampling method prevents us from making a direct comparison with standard sociolinguistic studies on this topic. Breaking with the standard of Otheguy & Zentella (2011), we made the decision to limit our analysis to pronoun reference within a single turn, in order to focus our attention on pronouns with locally available antecedents. While we ultimately replicated the effect of same- reference versus switch-reference contexts found by all standard sociolinguistic studies, we did find a slightly different *overall* rate of pronoun realization than other studies on this variety of Spanish, and our results should therefore be viewed with appropriate caution.

6. Conclusion

In this paper, we asked how children acquiring Spanish as their first language can learn the contrast between null and overt subjects given the complexity of the task. Specifically, we sought a way around the problem of having to identify pronominal antecedents in order to acquire the association between null subjects and same-reference, on the one hand, and overt pronominal subjects and switch-reference, on the other. We proposed that children solve this problem by first tracking the realization of 1st and 2nd person pronominal subjects, whose reference is easier to infer, and we provided evidence consistent with that proposal from the Schmitt-Miller corpus of Mexico City parent-child speech.

Finally, in addition to providing a novel proposal for how children acquire the conditions governing the use of null and overt pronominal subjects, this study provides the first evidence we know of that children under age 6 are sensitive to the conditions governing the use of null and overt subjects, filling a crucial gap in the literature on this otherwise well-studied alternation.

References

Alexiadou, A., & Anagnostopoulou, E. (1998). Parametrizing AGR: Word order, V-movement and EPP-checking. *Natural Language & Linguistic Theory*, *16*(3), 491-539.

Alonso-Ovalle, L., Fernández-Solera, S., Frazier, L., & Clifton, C. (2002). Null vs. overt pronouns and the topic-focus articulation in Spanish. *Italian Journal of Linguistics*, *14*, 151-170.

Austin, J., Blume, M., Parkinson, D., del Prado, Z. N., & Lust, B. (1997). The status of pro-drop in the initial state: Results from new analyses of Spanish. *Contemporary perspectives on the acquisition of Spanish*, *1*, 37-54.

Barbosa, P. P. (2009). Two kinds of subject pro. *Studia Linguistica*, *63*(1), 2-58.

Barbosa, P. P. (2011). Pro-drop and Theories of pro in the Minimalist Program Part 1: Consistent Null Subject Languages and the Pronominal-Agr Hypothesis. *Language and Linguistics*

Compass, *5*(8), 551-570.
Bayley, R., & Pease-Alvarez, L. (1997). Null pronoun variation in Mexican-descent children's narrative discourse. *Language variation and change*, *9*(3), 349-371.
Bel, A. (2003). The syntax of subjects in the acquisition of Spanish and Catalan. *Probus,15*(1), 1-26. doi:10.1515/prbs.2003.003
Belletti, A., Bennati, E., & Sorace, A. (2007). Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language & Linguistic Theory*, *25*(4), 657.
Carminati, M. N. (2002). *The processing of Italian subject pronouns* (Doctoral dissertation, ProQuest Information & Learning).
Carvalho, A. M., Orozco, R., & Shin, N. L. (2015). *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Washington, D.C.: Georgetown University Press.
Charnavel, I. (2015). Let You Be Bound to Me (and Me to You). *33rd WCCFL Poster Session Proceedings*. *SFU's Working Papers in Linguistics,* 18-27.
Filiaci, F., A. Sorace and M. Carreiras (2013). Anaphoric biases of null and overt subjects in Italian and Spanish: a cross-linguistic comparison. *Language, Cognition and Neuroscience*, 29:7, 825-843, DOI: 10.1080/01690965.2013.801502
Flores-Ferrán, N. (2007). A bend in the road: Subject personal pronoun expression in Spanish after 30 years of sociolinguistic research. *Language and Linguistics Compass*, *1*(6), 624-652.
Forsythe, H. (2017). *Discourse and Grammatical Cues in the Acquisition of Spanish Pronouns*. (Doctoral dissertation. Michigan State University).
Forsythe, H., Greeson, D. & Schmitt, C. (2019). Learnability in Romance: How indirect input helps children acquire the contrast between null and overt subjects. *Proceedings of the 43rd Boston University Conference on Language Development*.
de la Fuente, I. (2015). *Putting pronoun resolution in context: The role of syntax, semantics, and pragmatics in pronoun interpretation* (Doctoral dissertation, Université Paris Diderot).
Grinstead, J. (2004). Subjects and interface delay in child Spanish and Catalan. *Language*, 40-72.
Heim, I. (2008). Features on Bound Pronouns. Adger, Bejar, and Harbour (eds.), *Phi Theory: Phi Features Across Modules and Interfaces*. Oxford University Press, 35–56.
Hyams, N. (2012). *Language acquisition and the theory of parameters* (Vol. 3). Springer Science & Business Media.
Jegerski, J., VanPatten, B., & Keating, G. D. (2011). Cross-linguistic variation and the acquisition of pronominal reference in L2 Spanish. *Second Language Research*, *27*(4), 481-507.
Keating, G. D., VanPatten, B., & Jegerski, J. (2011). Who was walking on the beach?: Anaphora resolution in Spanish heritage speakers and adult second language learners. *Studies in Second Language Acquisition*, *33*(2), 193-221.
Keating, G. D., Jegerski, J., & Vanpatten, B. (2016). Online processing of subject pronouns in monolingual and heritage bilingual speakers of Mexican Spanish. *Bilingualism: Language and Cognition*, *19*(1), 36-49.
Kratzer, A. (2009). Making a pronoun: Fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry*, *40*(2), 187-237.
Lastra, Y., & Butragueño, P. M. (2015). Subject pronoun expression in oral Mexican Spanish. *Subject pronoun expression in Spanish: A cross-dialectal perspective*, 39-57.

Michnowicz, J. (2015). Subject pronoun expression in contact with Maya in Yucatan Spanish. *Subject pronoun expression in Spanish: A cross-dialectal perspective*, 103-122.
Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and cognition*, *7*(2), 125-142.
Montrul, S. (2011). Multiple interfaces and incomplete acquisition. *Lingua*, *121*(4), 591-604.
Orfitelli, R., & Hyams, N. (2012). Children's grammar of null subjects: Evidence from comprehension. *Linguistic inquiry*, *43*(4), 563-590.
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*, 770-802.
Otheguy, R., Zentella, A. C., & Livert, D. (2010). Generational differences in pronominal usage In Spanish reflecting language and dialect contact in a bilingual setting. *Language contact: New perspectives*, 45-62.
Otheguy, R., & Zentella, A. C. (2011). *Spanish in New York: Language contact, dialectal leveling, and structural continuity*. Oxford University Press.
Papadopoulou, D., Peristeri, E., Plemenou, E., Marinis, T., & Tsimpli, I. (2015). Pronoun ambiguity resolution in Greek: Evidence from monolingual adults and children. *Lingua,155*, 98-120. doi:10.1016/j.lingua.2014.09.006
Pérez-Leroux, A. T., & Glass, W. R. (1999). Null anaphora in Spanish second language acquisition: probabilistic versus generative approaches. *Second Language Research*, *15*(2), 220-249.
Pierce, J.W. (2007). PsychoPy – Psychophysics software in Python. J Neurosci Methods, 162(1-2), 8-13).
Rizzi, L. (1997). The fine structure of the left periphery. In *Elements of grammar* (pp. 281-337). Springer, Dordrecht.
Shin, N. L. (2012). Variable use of Spanish subject pronouns by monolingual children in Mexico. In *Selected proceedings of the 14th Hispanic Linguistics Symposium* (pp. 130-141). Cascadilla Proceedings Project Somerville, MA.
Shin, N. L., & Cairns, H. S. (2012). The Development of NP Selection in School-Age Children: Reference and Spanish Subject Pronouns. *Language Acquisition,19*(1), 3-38. doi:10.1080/10489223.2012.633846
Shin, N. L., & Erker, D. (2015). The emergence of structured variability in morphosyntax: Childhood acquisition of Spanish subject pronouns. *Subject pronoun expression in Spanish: A cross-dialectal perspective*, 169-190.
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*(4), 429-452.
Shin, N. L. (2016). Acquiring constraints on morphosyntactic variation: children's Spanish subject pronoun expression. *Journal of child language*, *43*(4), 914-947.
Sorace, A., Serratrice, L., Filiaci, F., & Baldo, M. (2009). Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua, 119*(3), 460-477. doi:10.1016/j.lingua.2008.09.008
Toribio, A. J. (2000). Setting parametric limits on dialectal variation in Spanish. *Lingua,110*(5), 315-341. doi:10.1016/s0024-3841(99)00044-3
Valian, V. (1990). Null subjects: A problem for parameter-setting models of language acquisition. *Cognition*, *35*(2), 105-122.
White, L. (2011). Second language acquisition at the interfaces. *Lingua*, *121*(4), 577-590.

Yang, Charles D. (2002). *Knowledge and learning in natural language*. Oxford University Press on Demand, 2002
Zagona, K. (2002). *The syntax of Spanish*. Cambridge University Press.
Zubizarreta, M. L. (1998). *Prosody, focus, and word order*. MIT Press.