After the Null Subject Parameter: Acquisition of the Null-Overt Contrast in Spanish

Hannah Forsythe

Language Sciences, University of California, Irvine

2211 Social Science Plaza B, UCI, Irvine, CA 92612, hforsyth@uci.edu

Daniel Greeson

Linguistics and Languages, Michigan State University

Cristina Schmitt

Linguistics and Languages, Michigan State University

After the Null Subject Parameter: Acquisition of the Null-Overt Contrast in Spanish

In many so-called consistent 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 reference is simpler to infer compared to 3rd person pronouns. Spontaneous production 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, and (iii) the contrast is generalized to the 3rd person slightly later in development. A pronoun resolution task shows that children use the null/overt contrast in comprehension at the same age that it becomes apparent in production.

Keywords: null subjects, pronoun resolution, structured variation, acquisition of Spanish, discourse processes

1. Introduction: Revisiting the null subject parameter

Since the seminal work by Hyams (1986/2012), the so-called "null subject parameter" has played a central role in the theory of language acquisition. This research program, situated within the Principles and Parameters approach (Chomsky 1981/1993, 1986), framed the acquisition of null subjects as a process of deciding between two "settings" of a parameter. If the target language licenses null subjects in tensed clauses, then the child's task is to set the parameter to "on," otherwise the child must set the parameter to "off." This program made the surprising discovery that children initially produce null subjects regardless of the language they are exposed to, suggesting that the "on" setting of the null subject parameter is available in the child's grammar from the start.

Subsequent developments in syntax, psycholinguistics, and sociolinguistics have created a clearer picture of the phenomenon of *pro*-drop, helping to refine our understanding of those original findings and, more broadly, what it means to "acquire null subjects." First, cross-linguistic syntactic work has revealed that the appearance of null subjects can be attributed to more than one parameter choice. In Italian-style languages, for instance, the properties of verbal inflection are what license the null subject[1], while in topic-drop languages like Mandarin, null subjects and null objects are both licensed by an operator linked to a topic position (see Huang 1984, Roberts 2010, and Barbosa 2009, to name a few). Second, sociolinguistic and psycholinguistic investigations have revealed that even when the null subject is licensed by the same formal mechanisms, its precise distribution and preferred interpretation can vary probabilistically across even closely related language varieties (Filiaci *et al*. 2013, Keating *et al.* 2011, Carvalho, Orozco & Shin 2015 and references therein).

These findings bring into sharper focus what was already (perhaps implicitly) assumed in the original framing of the acquisition problem: that the acquisition of null subjects is in fact a multi-step process; one in which the child not only identifies the correct parameter setting but also acquires the precise statistical distribution of null and overt subjects in those grammatical environments where the two overlap. Language acquisition researchers have long recognized that the first step is accomplished early. Children exposed to *pro*-drop languages produce both null and overt subjects by age 2 (Grinstead 2004, Bel 2003). And even though children exposed to non-*pro*-drop languages also produce and accept null subjects (Wang *et al.* 1992, Orfiteli & Hyams 2012 and references therein), their rate of null subject production is significantly lower

---

[1] Or, in an alternate analysis, verbal agreement is itself a pronominal subject.

compared to that of their peers (Valian 1990, inter alia) and drops to less than 3% by around two and a half years (Valian 1989).

However, the second step has received much less attention from the L1 research community, and this is where we focus our efforts. One of the fundamental goals of child language research is to understand why language is the way it is, since natural languages can only exhibit the patterns they do if these patterns are acquirable by children. The fact that every language exhibits some form of probabilistic variation therefore suggests that variable patterns are every bit as acquirable as categorical patterns. Although the acquisition of variable patterns is not as well described and understood as that of categorical patterns, children have nevertheless been shown to acquire certain aspects of structured variation early in life (Miller 2007, Smith, Durham & Fortune 2007, 2009). Since the alternation between null and overt subjects in overlapping environments is a particularly widespread and diachronically stable pattern of variation that has been thoroughly studied in adults, we believe that it makes an especially useful case study for understanding the role of acquisition in shaping natural language.

This paper focuses on the division of labor between null and overt subject pronouns in Italian-style *pro*-drop languages. In these varieties, null subjects are probabilistically associated with continued reference to the preceding subject antecedent, while overt subject pronouns are associated with a switch in reference. How do children acquiring these varieties of *pro*-drop learn to produce each variant in its preferred environment, and how do they learn to make use of this contrast in comprehension? We hypothesize that the most efficient strategy for children to pursue is to initially restrict their attention to a subset of the input: 1st and 2nd person pronominal subjects. We support this hypothesis with evidence from the acquisition of subject pronouns in Mexico City Spanish. First, we show that the target knowledge is in principle acquirable solely

from the distribution of 1st and 2nd person subjects in child-directed speech. Second, we provide evidence that children reflect this knowledge in their own production of 1st and 2nd person subjects before subsequently generalizing it to the production of 3rd person pronominal subjects. And finally, we show that the onset of adult-like production coincides with the ability to use the null/overt contrast in comprehension.

2. The null/overt distinction in adult grammars

In so-called canonical null subject languages like mainstream Italian, European Portuguese, and most mainland varieties of Spanish, the null subject is considered to be a pronoun. In some analyses, both null and overt pronominal subjects enter into an agreement relation with the verb[2], while in others the pronominal subject is itself realized as an agreement affix on the verb[3]. Whatever the analysis, the child's task is to discover not just the proper analysis of the target language, but also which environments trigger each variant (null, overt) and how to interpret these pronouns when they appear in the input. In this section, we describe some of the conditions that govern adults' production and comprehension. In section 3, we illustrate how acquiring these conditions presents a difficult challenge to the learner. In section 4 we hypothesize a learning path that could in principle allow learners to efficiently surmount this challenge. The remainder of the paper is dedicated to providing evidence that this is in fact the path taken by child learners of Spanish, as revealed in their production (Section 5) and comprehension (Section 6) of null and overt subject pronouns.

A host of sociolinguistic studies reveals that production of the overt variant is probabilistically conditioned by multiple factors, including the person and number of the subject,

---

[2] The exact properties of the null pronoun vary across different analyses, see Roberts (2010), Holmberg (2005) among others.
[3] Most notably Alexiadou & Anagnostopoulou (1998); see Barbosa (2011a, 2011b) for a review.

the tense, mood and aspect of the verb, and crucially, what the pronoun's antecedent is (Bayley & Pease-Alvarez 1997, Flores-Ferrán 2007, Lastra & Butragueño 2015, Michnowicz 2015, Otheguy *et al.* 2007, 2010; Otheguy & Zentella 2011, Shin & Otheguy 2009). Across varieties of Spanish, speakers tend to produce more null subjects when the intended antecedent is the preceding subject and more overt subjects when the intended antecedent is a non-subject (a preceding direct object, indirect object, oblique, etc.) or when introducing a new referent. For example, speakers are more likely to use (1a) to express the message that *Juan called Pedro when Juan was at home*, and to use (1b) to express that *Juan called Pedro when Pedro was at home*.

(1) a. *Juan llamó a Pedro cuando ∅ estaba en casa.*   [adapted from de la Fuente 2015]
      Juan called Pedro when **pro** was   at home.
    b. *Juan llamó a Pedro cuando **él** estaba en casa.*
      Juan called Pedro when **he** was   at home.

The subject-antecedent reading, where *Juan* is the one at home, is commonly referred to as the "same-reference" reading because the same referent is referred to in subject position both times; all non-subject readings tend to be grouped under the term "switch-reference" (Otheguy & Zentella 2011). The contrast between same-reference and switch-reference contexts is consistently found to be one of the strongest factors conditioning subject pronoun expression (Otheguy & Zentella 2011, Carvalho, Orozco & Shin 2015 and references therein), and it is the factor that we will focus on in this paper.

On the comprehension side, being null or overt is not the only factor that must be considered when trying to determine the pronoun's intended referent. Following widely adopted accounts of personal pronouns (Schlenker 1999, Sauerland 2008, Charnavel 2019), we assume that both null and overt subject pronouns are variables, interpreted by an assignment function, as in (2) from

Büring (2011). The assignment function *g* is an ordered list of referents from the context (ex.

{*Jane, Alice, John, my brothers*}) that takes the index of the pronoun, *i*, and returns the *i*-th

member of that list (ex. *she₁* = *Jane*, *she₂* = *Alice*, etc.). The pronoun's phi-features introduce

presuppositions that restrict the output of this function: if the *i*-the referent fails to satisfy the

presuppositions imposed by the pronoun's features (ex. *my brothers* violates the feminine,

singular features of the pronoun *she*), then the function is undefined, i.e., it returns nothing.

(2) $[\![she_i]\!]^g = g(i)$
    where $g(i)$ returns the *i*-th member of g, a list of referents, and $g(i)$ is a singular female,
    else undefined

On this account, the job of the listener is to construct the assignment function based on what

he or she believes is the referent intended by the speaker. This is a very subtle process that

depends on a multitude of considerations, including not just the pronoun's phi-features, but also

the listener's general world knowledge, the physical environment, the speaker and listener's

shared understanding of the question under discussion (Roberts 1996), and the coherence

relations established between discourse segments (Kehler 2002). For example, in (3a) the hearer

is unlikely to pick out *Bill* as the referent of *he* despite both Bill and John being syntactically

viable options, due to real-world knowledge about the nature of apologies. On the other hand, if

the coherence relation between the two events changes, as in (3b), this preference is reversed.

(3) a. John apologized to Bill because he had been rude.
    b. John apologized to Bill even though he had been rude.

Speakers of canonical *pro*-drop languages can also use pronoun realization (alongside these

other considerations) to guide their interpretation. For example, Spanish speakers presented with

globally ambiguous examples like (1) above tend to interpret the null subject (1a) as referring to

the preceding subject *Juan,* whereas they have no such interpretive bias for the overt subject (1b)

(Alonso-Ovalle *et al.* 2002, de la Fuente 2015). And in both Spanish and Italian, online reading times and offline acceptability judgments of pragmatically disambiguated sentences show that the null subject is easier to process and perceived as 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).

Different proposals have been made to explain the nature of this contrast (Carminati 2002, Blackwell & Quesada 2012, Luján 1985, Frascarelli 2007, and many others; see de la Fuente 2015 for a review) all of which rely on some notion of prominence: the more prominent the intended referent, the more likely the pronoun is to be null. This is part of a universal tendency for more reduced referring expressions to pick out more prominent referents (Ariel 2001, Gundel *et al.* 1993), which in this case means referents that have been referred to in subject position.

It is important to underscore that the null/overt contrast is probabilistic in nature: antecedent position does not uniquely determine the realization of pronominal subjects, nor does pronoun realization categorically determine antecedent position. Both (1a) and (1b) can grammatically be used to refer to the preceding subject, the object, or an extra-sentential referent, and any of these readings can be made more or less likely by the surrounding discourse. 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 object reading would be favored over the subject reading for both the null subject in (1a) and the overt subject in (1b), although the object bias might be slightly weaker for (1a). The task for the L1 learner, in other words, is to associate null and overt pronominal subjects, not with distinct interpretations, but with an increase or decrease in the *probability* of a same-reference interpretation, respectively.

The last point to underscore about this probabilistic contrast is that it varies in strength across different communities. Sociolinguistic studies across varieties of Spanish reveal that the contrast between same-reference and switch-reference contexts can condition overt pronoun production to different degrees, depending on the variety spoken, and the strength of this factor can change when dialects come into contact (Otheguy *et al.* 2007, 2010; Shin & Otheguy 2009). Cross-linguistic variation is also attested in interpretation. For instance, psycholinguistic studies show that Italian more strongly associates the overt variant with a non-subject antecedent compared to Spanish (Filiaci *et al*. 2013, de la Fuente 2015).

How do children acquire this probabilistic association between pronoun realization and pronoun reference? Even if the tendency to associate more reduced referring expressions with more prominent referents is universal, there is still language-particular learning to be done. The child must discover which dimension of prominence is relevant (in this case, the prominence afforded to a referent by being mentioned in the immediately preceding subject position), and she must learn the precise statistics of this association for her particular language variety.

3. Defining the learning problem

The child's first step in the acquisition of subjects is simply to gather the full inventory of referring expressions that may appear in subject position, including names, definite and indefinite noun phrases, demonstratives, and of course null and overt personal pronouns. Setting the null subject parameter correctly is part of this process: it adds both overt *and* null personal pronouns to the inventory of subjects that can appear in tensed clauses. The next step is to search for the conditions governing when to use each one. For null and overt pronouns, this means tracking whether the pronoun in subject position refers to the same entity as the referring expression in the preceding subject position (same-reference) versus a different entity (switch-

reference) and forming an association between the pronoun's reference and its realization. The last step is to fine-tune the statistics of this association to match the target language.

What could help the learner identify the contrast between same-reference and switch-reference as a crucial factor determining subject pronoun realization? This is not an easy problem to solve because identifying same- and switch-reference environments requires the learner to know in advance what the intended referent of the null or overt pronoun is. If a child is to verify that a null subject in her input indicates reference to the preceding subject, or that an overt subject indicates reference to a non-subject antecedent, she must know what the intended referent of each pronoun is in the first place. However, reference resolution is a much more complex task for pronouns than it is for other referring expressions like names and noun phrases, a fact that can be illustrated by any discourse situation that offers more than one potential antecedent for a pronoun.

For example, in the context of a story about Juan and Pedro skipping school, being discovered by the principal, and being disciplined by their fathers, a speaker could utter (4a) or (4b) using either the null or overt pronoun; or as illustrated in (5) the speaker could use any number of more semantically restricted noun phrases, such as a proper name or a 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 expressions, as illustrated in (6).

(4) a. *Juan llamó a Pedro cuando ∅ estaba en casa.*
    Juan called Pedro when **pro** was   at home.
   b. *Juan llamó a Pedro cuando **él** estaba en casa.*
    Juan called Pedro when **he** was   at home.

(5) Context: A story about Juan and Pedro skipping school, getting caught by the (male) principle, 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.

(6) Potential referents of the subject DP in (5)
    a. *Juan* : {Juan}
    b. *el niño travieso*: {Juan, Pedro}
    c. null subject: {Juan, Pedro, principle, Juan's father, Pedro's father}
    d. overt subject: {Juan, Pedro, principle, Juan's father, Pedro's father}

What this example shows is not that sentences like (4a-b) are impossible for learners 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 associate this interpretation with the pronominal form chosen by the speaker.

4. Proposed solution

Charnavel (2019) argues that 1st and 2nd person pronouns have the same formal representation as 3rd person pronouns—all three depend on an assignment function linking their index to the intended referent. The reason for the interpretive differences that researchers have observed between 1st and 2nd person, on the one hand, and 3rd person, on the other, is due to the fact that, for the former, it is simply a lot easier to infer the speaker's intended referent. For first person, the intended referent is nearly always the speaker, and for second person it is almost always the addressee.

Charnavel's analysis suggests two things: first, that it should be much easier for children to identify same-reference versus switch-reference uses of 1st and 2nd person pronouns, and second, that there is nothing preventing children from automatically generalizing knowledge they have acquired about 1st and 2nd person subject pronouns to 3rd person subject pronouns. If this

suggestion is correct, then it offers a way around the challenge faced by the learner when trying

to associate certain pronominal interpretations with certain pronominal forms. If children initially

restrict their attention to tracking the realization and interpretation of 1st and 2nd person subjects,

then this knowledge can be used to identify those same conditions for 3rd person subjects. We

therefore hypothesize that children begin the acquisition process by forming an association

between overt 1st and 2nd person pronominal subjects and switch-reference, on the one hand, and

null 1st and 2nd person pronouns and same-reference, on the other. Once this association is

formed, we hypothesize that they generalize it to the production and comprehension of 3rd person

pronominal subjects. This learning path is summarized below.

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

Of course, it is not enough to argue that following this learning path is a good idea, we must

show that it is consistent with real children's production and comprehension. We make the case

that it does by answering three basic questions.

Q1. Is the null/overt contrast acquirable, in principle, 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 subjects?

Q3. Do children generalize this contrast to the production and comprehension of 3rd person pronominal subjects *after* they have acquired it in the domain of 1st and 2nd person?

In the remainder of the paper we present evidence from the acquisition of Spanish by

children in Mexico City that answers each of these questions with a 'yes.' Section 6 uses a

corpus of naturalistic mother-child speech to address the question of children's input and

children's production. In brief, this shows that the statistical contrast between null and overt

subject reference in children's input is not only significant in the 1st and 2nd persons, but stronger

than it is in the 3rd person. This study also shows that children under age 6 reproduce this

contrast in their own production of pronominal subjects, and that the contrast emerges earlier

within the 1st and 2nd persons as compared to the 3rd person. Section 7 presents a pronoun

resolution task addressing the question of children's comprehension. Results indicate that

children begin using the null/overt contrast to guide their interpretation of grammatically

ambiguous 3rd person pronouns at approximately the same age when their production becomes

adult-like (around 4 ½). Before presenting this data, Section 5 reviews what is currently known

about the developmental path in languages that display the null/overt contrast.

5. Acquisition background

What we know so far about the developmental path comes from studies of spontaneous

production, felicity judgments, and pronoun resolution tasks in a variety of *pro*-drop languages.

We concentrate here on those with rich agreement like Spanish, Catalan, Italian, and Greek.

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 individual varieties (Filiaci *et al*. 2013, de la Fuente 2015).

Production studies from the acquisition literature find that children acquiring Spanish and

Catalan begin to produce overt subjects before age 2 (Grinstead 2004, Bel 2003). However, overt

pronominal subjects remain infrequent. Sociolinguistic studies that include child participants

reveal low rates of overt subject pronoun realization 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 & Otheguy 2013, Shin & Erker 2015). 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 (11a) and the other an overt subject (11b). Crucially, the event was manipulated such that either *Minnie* herself had fallen (a same-reference context) or another person had fallen (a switch-reference context). 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. 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 (12b).

(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 realization and pronominal reference in both spontaneous production and the comprehension of ambiguous pronouns. After this age, through middle childhood and pre-adolescence, children gradually incorporate additional sociolinguistic constraints on subject pronoun realization and slowly develop the ability to make felicity judgments about the appropriate use of the null and overt variants. Unfortunately, not much is known about the trajectory of development before 6, 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.

6. Production study: Subject pronoun realization in spontaneous speech

The first question we address is whether 1st and 2nd person pronouns in the input actually provide the distributional information necessary to acquire the null/overt distinction. In this section, we examine a corpus of mother-child interactions to determine whether null and overt 1st and 2nd person pronominal subjects in child-directed speech are indeed associated with same- and switch-reference environments, respectively. The next 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 use data from the Schmitt–Miller corpus, collected in Mexico City, Mexico in 2008, which records the spontaneous speech of 11 low-SES and 14 high-SES child-caretaker dyads ages 1;6 to 5;11. Caretakers were recorded playing with their children during 2-4 free-play sessions lasting between around 30 minutes, as well as one approximately 30-minute session chatting with another adult. The entire corpus contains approximately 649,000 words.

6.1. Participants

We examined a subset of the corpus, totaling 183,443 words from the speech of nine mother-child dyads, whose characteristics are summarized in Table 1.

	Child	Age	Mean Length of Utterance	Mother Word Count	Child Word Count
Middle Class	JGAV	2;11	2.903	20,873	7,485
	YGSZ	3;9	3.652	9,608	10,910
	EAMR	4;3	4.713	5,758	8,522
	SLV	4;10	4.738	9,590	11,277
Working Class	KDP	3;4	4.754	8,962	7,007
	KUC	4;1	4.522	11,721	9,393
	YBM	4;8	3.993	11,054	8,373
	OMJ	5;2	3.87	11,934	7,314
	JRC	5;11	3.747	14,787	12,170
		Mean: 4;4	Mean: 4.099	Total: 104,287	Total: 82,991

Table 1: Subject characteristics of 9 Mexican Spanish-speaking dyads from the Schmitt-Miller corpus

6.2. Hypotheses and predictions

The learning path hypothesized in Section 4 makes three predictions about the realization of pronominal subjects. First, this path assumes that the null/overt distinction is learnable from the distribution of 1st and 2nd person pronouns in the input, meaning that we expect mothers in our sample to produce significantly more overt 1st and 2nd person pronominal subjects in switch-reference contexts than they do in same-reference contexts. Second, if children begin acquiring this statistical pattern at some time before 6 years of age, as suggested by the production and

comprehension data available to date, then we predict that some portion of the children in our sample will reproduce this same statistical contrast in their own production of 1st and 2nd person pronominal subjects. Finally, since we hypothesize that children generalize knowledge from 1st and 2nd person to the 3rd person, rather than the other way around, we predict that a slightly older subgroup of children in our sample will show this statistical pattern in their production of 3rd person pronominal subjects, as well.

6.3. 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.

6.3.1 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), due to the fact that they are not in a perfect parallel with the subjects of regular transitive verbs 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.

6.3.2 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 speaker's 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, since we excluded any subjects that were not preceded by another tensed verb in the same turn, the first subject in (11), *yo* ('I') would be excluded from coding because it has no preceding subject with which to maintain or switch reference. Nevertheless, it still serves as a suitable "trigger" for coding the subject of the following verb, *estás* ('you are').

(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 tired, (**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[4]. In these

cases, the preceding subject was skipped and the preceding subject before that was used 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 morphologically imperative set phrases like *mira*, *oye*, *ándale*, 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,
           expletive subjects (*es que*, *hay que*, 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.

6.3.3 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.

---

[4] 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?

To calculate inter-rater reliability, both raters independently coded a subset consisting of 6 transcripts (comprising 11% 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.

6.3.4. 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 {ɵ} no va a comer lunetas nunca más en la vida
        and now (you) will never eat lunetas ever again in your life.

These examples demonstrate 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.

6.4. Results: Overall rate of subject realization

 Mothers and children in our sample produced overt pronouns at very similar rates to each other (10.9% for mothers, 13.0% 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)	10.9% (291/2,667)
Adult-directed speech, Mexico City (Lastra & Butragueño 2015)	21.7% (443/2,040)
Children ages 3-6 (this study)	13.0% (259/1,998)
Children ages 6-7, Querétaro & Oaxaca (Shin 2016)	8% (148/1,845)

Table 2 Rates of overt pronoun use by mothers and children in a subset of the Schmitt-Miller corpus and in other corpora of Mexican child and adult speakers of Spanish

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) (10.9% 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 same turn, we excluded turn-initial pronouns, potentially under-sampling the overt variant. When turn-initial pronouns are included in this count, mothers' overt pronoun rate increases slightly to 11.1% (814/7,332).

On the other hand, the children in our sample produced overt pronouns more often in comparison to slightly older children from other regions in Mexico, as reported by Shin (2016) (13.0% versus 8%), and this rate remains steady at 13.0% when turn-initial pronouns are included (782/5,996). 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 in addition may have encouraged children to assimilate to their mothers' input, leading to more overt subjects.

6.5. Results: Subject realization in same- and switch-reference contexts

The learning path proposed in Section 3 depends upon the assumption that children's input provides the statistical information necessary 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.

Table 3 shows the rate and raw 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 subjects. Middle-class dyads (henceforth, MC) and working-class dyads (henceforth, WC) are shown separately from each other. Before performing any analyses we tested for differences between SES groups using two multilevel mixed-effects regression models on mothers' and children's rate of overt pronoun realization, with SES (MC, WC) as a level 1 fixed effect. The model also included reference (same, switch) and person (1st, 2nd, 3rd) as level 1 fixed effects and individual dyad (9 mother-child pairs) as a level 2 random effect. For mothers, switch-reference was strongly associated with an increase in overt realization (β = 1.97, SE = 0.74, p < 0.01), but there was no main effect of SES and no interactions, although the three-way interaction between switch-reference, 2nd person, and the WC group came close to reaching significance (β = 1.76, SE = 0.92, p = 0.054). Additionally, this model failed to significantly outperform an otherwise identical model lacking SES as a main effect (Model 1: DF = 13, AIC = 1777.6; Model 2: DF = 7, AIC = 1773.5, $\chi^2(6)$ = 7.93, p = 0.24). For children, likewise, switch-reference was strongly associated with an increase in overt realization (β = 1.40, SE = 0.41, p < 0.001) and there were no other significant main effects or interactions. This model also failed to significantly outperform the alternative model lacking SES as a main effect (Model 1: DF = 13, AIC = 1435.3; Model 2: DF = 7, AIC = 1425.6, $\chi^2(6)$ = 2.30, p = 0.89). We therefore collapse across SES in all subsequent analyses.
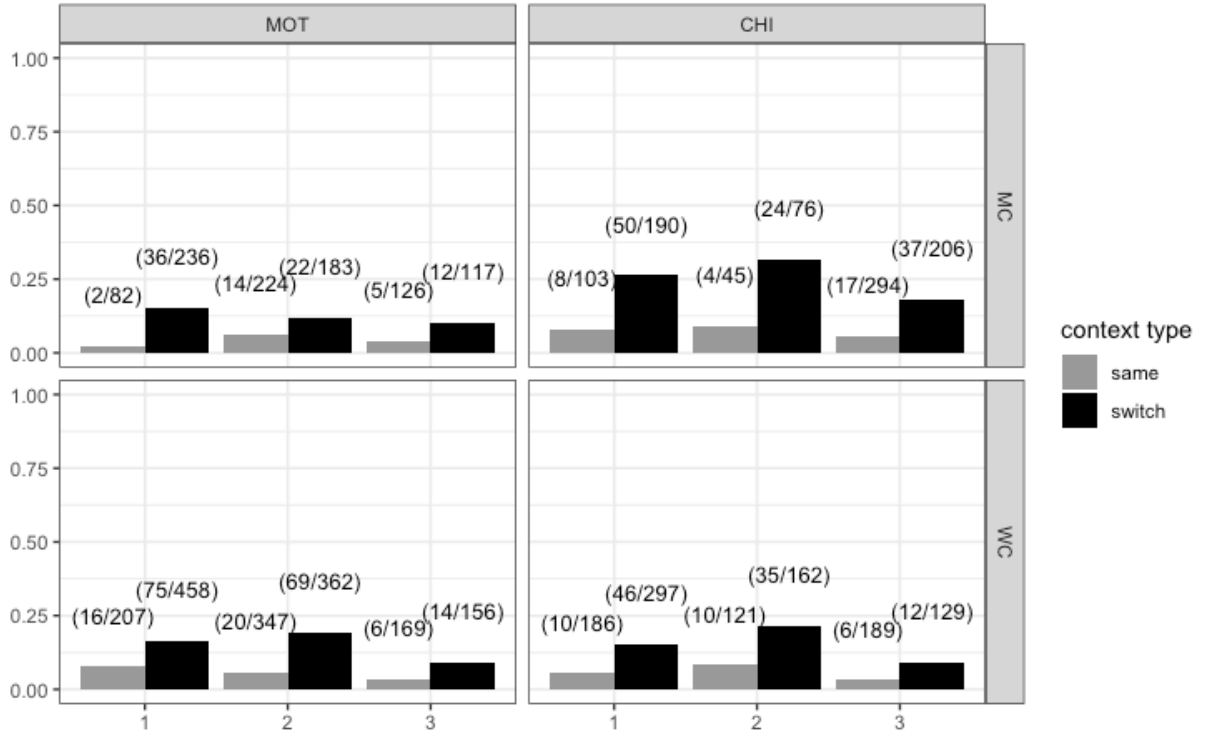
Table 3: Rate of overt 1st, 2nd, and 3rd person pronouns appearing in same- and switch-reference contexts in the speech of mothers (left) and children (right). WC: Working Class (*N*=5 dyads, age range: 3;4-5;11, Tokens: 2,783); MC: Middle Class (*N*=4 dyads, age range: 2;11-4;10, Tokens: 1,882).

To test the first prediction, we employed separate chi-squared tests of proportion comparing the rate of overt pronoun production across same-reference and switch-reference contexts in mother's speech, within each of the three persons. Across all three persons, mothers produced significantly more overt pronominal subjects in switch-reference contexts relative to same-reference contexts (1st person: $\chi(1) = 16.22$, $p < 0.001$; 2nd person: $\chi(1) = 31.28$, $p < 0.001$; 3rd person: $\chi(1) = 6.90$, $p < 0.01$). Thus, the statistical contrast between null and overt pronominal subjects appears to be at least as statistically reliable in the 1st and 2nd persons as it is in the 3rd person.

Given that 1st and 2nd person pronominal subjects in the input seem to provide a robust statistical signal, the next prediction to test is whether children are sensitive to this signal and reflect it in their own production at some point before age 6. For children, this contrast was

significant across the board. That is, in the 1st person ($\chi(1) = 25.25$, $p < 0.001$), in the 2nd person ($\chi(1) = 16.58$, $p < 0.001$), and in the 3rd person ($\chi(1) = 22.77$, $p < 0.001$), children produced significantly more overt pronouns in switch-reference relative to same-reference contexts, indicating that, as a group, they have applied the null/overt distinction to all three persons.

To get a more fine-grained picture of children's developmental trajectory, we compared the rate of overt pronominal subjects produced by each individual child, as shown in Figure 1. Every single child produced overt subject pronouns more often in switch-reference contexts compared to same-reference contexts, but the difference was significant only for the two older MC children (EAMR: 4;3, $\chi(1) = 19.55$, $p < 0.001$; SLV: 4;10, $\chi(1) = 26.86$, $p < 0.001$) and the three oldest WC children (YBM: 4;8, $\chi(1) = 8.48$, $p < 0.01$; OMJ: 5;2, $\chi(1) = 5.83$, $p = 0.016$; JRC: 5;11, $\chi(1) = 11.63$, $p < 0.001$). If this sample of 9 children is representative of acquisition more generally, then this result suggests that typically developing children acquire the null/overt contrast about halfway through the fifth year of life.
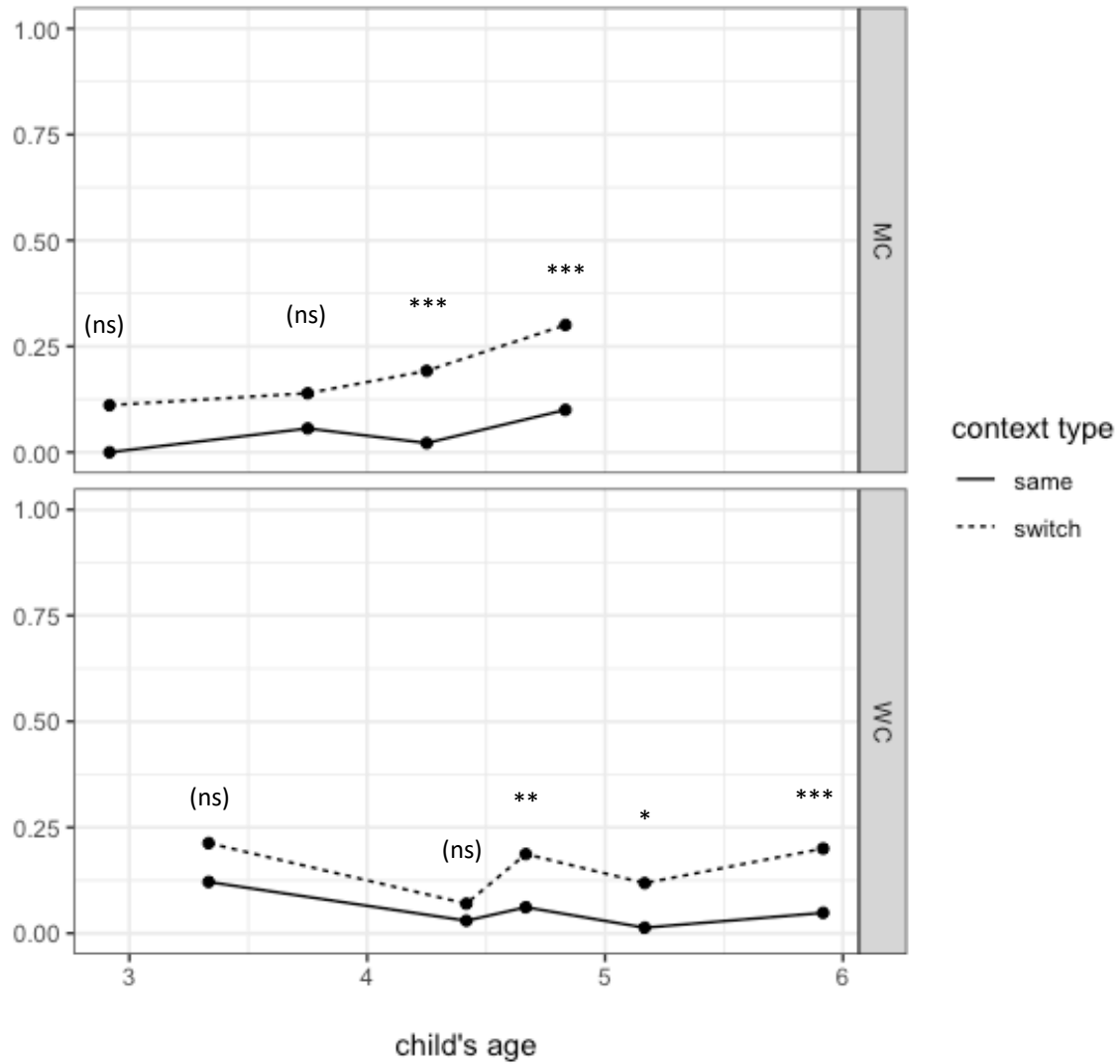
Figure 1: Rate of overt pronominal subjects produced by individual children, in same-reference and switch-reference contexts.

The final prediction to test is whether children generalize knowledge acquired in the 1st and 2nd persons to the 3rd person, rather than the other way around; in other words, that the null/overt distinction is acquired *earlier* in the 1st and 2nd persons compared to the 3rd. To test this prediction, we fit a linear regression model predicting pronoun realization (overt = 1, null = 0) as a function of the pronoun's reference (same vs. switch), the pronoun's person feature (1st/2nd vs. 3rd), and children's age (2.9-5.9 years), in the expectation that there would be a 3-way interaction. The model revealed a significant 3-way interaction ($\beta = 1.30$, SE $= 0.52$, $p = 0.012$),

plus a significant 2-way interaction between pronoun reference and person ($\beta = -6.04$, SE = 2.39, $p = 0.012$; see appendix Table 1 for full details).

To explore these interactions, we divided children into groups by their age in years: under 4 ($N = 3$), 4 to 5 ($N = 4$), and over 5 ($N = 2$). For each age group, we ran separate, one-tailed tests of proportion comparing the rate of overt pronoun realization across same- and switch-reference contexts, within each person (see appendix Table 2 for full details). In the 1st person, this difference was marginally significant for children under 4 ($\chi^2(1) = 2.53$, $p = 0.056$) and significant for the other two age groups (4 to 5: $\chi^2(1) = 17.8$, $p < 0.001$; over 5: $\chi^2(1) = 4.34$, $p = 0.019$). In the 2nd person, this difference was significant for all age groups (all $\chi^2(1) > 2.9$, all $p < 0.05$). In the 3rd person, however, only children over 4 showed the same contrast (4 to 5: $\chi^2(1) = 20.68$, $p < 0.001$; over 5: $\chi^2(1) = 4.65$, $p = 0.016$). 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.

6.6. Discussion and study limitations

This sample of naturalistic input and spontaneous child production from the Schmitt-Miller corpus has provided three crucial pieces of evidence that are consistent with the learning path hypothesized in Section 4. First, analysis of the input reveals that the null/overt distinction is in principle learnable from 1st and 2nd pronouns alone, consistent with the hypothesis that children specifically track this subset of the input in the early stages of acquiring this distinction. Mothers produce significantly more overt 1st and 2nd person pronominal subjects in switch-reference contexts relative to same-reference contexts. Assuming that the input produced by these mothers can be generalized to caretaker speech at similar developmental stages, then this would suggest

that children can learn the null/overt distinction from 1st and 2nd person subjects without needing to include data from 3rd person pronominal subjects.

Second, children's own production is consistent with their having picked up on this statistical signal as it applies to 1st and 2nd person pronouns, consistent with the hypothesis that children not only track but actually internalize this distinction. Like their mothers, they produce significantly more overt 1st and 2nd person pronouns in switch-reference compared to same-reference contexts.

Third and most importantly, children show the same statistical contrast in the 3rd person, but this contrast reaches significance slightly later in development, consistent with the hypothesis that children generalize acquired knowledge from the 1st and 2nd persons to the 3rd person, rather than the other way around.

This study has one main limitation, which 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 speaker turn in order to focus our attention on pronouns with locally available antecedents. This provides a good point of comparison to psycholinguistic studies on the topic, all of which test the comprehension of pronouns with locally available antecedents (some within the same sentence and others across sentence boundaries, but all within a single speaker's turn). However, this limitation makes the comparison to sociolinguistic studies less direct, although it should be noted that we ultimately did replicate the effect of same- versus switch-reference found in the sociolinguistic literature.

7. **Comprehension study: Resolving 3rd person subject pronouns**

The last prediction we test has to do with comprehension. If, as we hypothesize, children generalize the knowledge they have internalized about the realization of 1st and 2nd person subjects to their realization of 3rd person subjects, then they should be able to leverage this same knowledge when interpreting 3rd person subjects, as adults do. Since our production data suggests that children acquire the null/overt contrast around age 4 ½, we would expect children to begin using this contrast in their interpretation of 3rd person pronouns at approximately the same age. For example, we would expect children to associate the null subject in (16a) with the subject antecedent, *Juan*, significantly more often than the overt subject in (16b). In fact, we would expect this contrast to remain stable even across different pragmatic contexts, such as (17), where the casual connective *y por eso* ('and so') triggers a slight object bias.

(16) a. *Juan le pega a Pedro y después ø se va.*
           Juan CL hits A Pedro and then *pro* leaves.
       b. *Juan le pega a Pedro y después él se va.*
           Juan CL hits A Pedro and then he leaves.

(17) a. *Juan le pega a Pedro y por eso ø se va.*
           Juan CL hits A Pedro and so *pro* leaves.
       b. *Juan le pega a Pedro y por eso él se va.*
           Juan CL hits A Pedro and so he leaves.

Previous research with L1 learners of Mexican Spanish has failed to find any evidence that children this age distinguish between null and overt subjects in felicity judgment tasks (see discussion of Shin & Cairns (2012) in Section 5). However, our interest is not in children's awareness of what constitutes felicitous and infelicitous uses of these pronouns, but rather what constitutes a more or less likely *interpretation* of each one. We therefore used a forced-choice picture selection task to probe children's preferred interpretations of the null and overt variants. A second benefit of this task is that it places lower demands on working memory. Rather than asking children to choose between two alternative sentences, which must be held simultaneously

in working memory as they are compared, this task asks children choose between two different visually represented interpretations of the same sentence, which are presented side by side.

7.1. Methods
7.1.1. Participants

A total of 41 adults (35 women) and 76 children (41 girls) ages 2;11 to 6;4 ($M = 4;5$) completed the task. After exclusions (see 7.1.3 below), children were divided into two age groups for analysis: 40 children ages 2;11-4;5 ($M = 3;9$, SD = 5.8 months) and 33 children ages 4;6-6;4 ($M = 5;5$, SD = 5.1 months).

7.1.2. Design and procedure

Subjects participated in a 16-item forced-choice picture selection task in a 2 (null, overt) x 2 (Version 1, Version 2) within-subjects design. In version 1 the two clauses were joined via a discourse connective expressing a temporal relation: *y después* ('and then'); in version 2 the discourse connective expressed a causal relation: *y por eso* ('and so'). This tests whether the contrast between null and overt subjects persists under different pragmatic conditions that may also have an effect on the preferred reading of the null and the overt pronoun.

Items were created by crossing each condition with eight verb-phrase pairs (*alegrar– aplaudir*: 'cheer up–applaud,' *cantar para–bailar*: 'sing for–dance,' *perseguir–cansarse*: 'chase– get tired,' *asustar–gritar*: 'scare–yell,' *pegarle–irse*: 'hit–leave,' *pelearse con–llorar*: 'quarrel with–cry,' *tocar-reírse*: 'poke–laugh,' and *hablar con-sonreír*: 'speak to–smile'). Each subject saw every item in two out of the four conditions.

Prompts were presented in blocks, by condition, with the order of blocks counterbalanced across participants. Between each block participants saw items from a separate study that used some of the same illustrated characters. Items were presented in random order within each block.

Pictures were presented on a computer screen using Psychopy version 1.82.01 (Peirce 2007), and the position of the first-mentioned character (left or right side) was counterbalanced across verb-phrase pairs. Children were read the prompts by a native speaker who recorded their responses on the computer by pressing either the "4" key (left-hand picture) or "9" key (right-hand picture). Adults listened to pre-recorded prompts and pressed the keys for themselves.

Before beginning, the experimenter explained the task and obtained children's verbal assent using the following script.

(18) *Te voy a contar unas historias acerca de mis amigos. Al final de cada historia, vas a*
ver dos fotos, y tú me tienes que decir cuál es la foto que corresponde, ¿sale?
I'm going to tell you some stories about my friends. After each story, you're gonna see
two pictures, and you have to tell me which is the one that matches, sound good?

Next, participants were introduced to the characters used in the pictures, who were all school-aged children with common Mexican names, two male and two female (*María, Sara, Juan, Pedro*). Learning the characters' names was not technically necessary to complete the task, since the first clause of each experimental prompt always identified the characters by name and an accompanying illustration. However, we reasoned that familiarity with character names facilitates processing, and we therefore administered a 4-item name-recognition task to weed out any participants with especially poor name recall. Finally, the experimenter elicited the child's assent a second time and began the experiment.

7.1.3. Exclusions

Subjects providing fewer than two correct answers in the pre-experimental name-recognition task were excluded from analysis (1 adult, 3 children). The remaining adults gave an average of 3.63 correct responses out of 4 and children averaged 3.37 correct out of 4. A small number of

individual responses were eliminated for reasons including inattention, failure to select only one picture, or experimenter error in presenting items (1.16% of total data).

7.2. Results

We estimated participants' sensitivity to pronominal form by measuring the rate at which they chose the picture compatible with a same-reference reading; i.e., the picture in which the character previously mentioned in subject position in the first clause was depicted performing the action described in the second clause. Figure 2 shows the rate of same-reference choices made by adults, older children, and younger children, for null and overt pronouns in each version of the experiment.
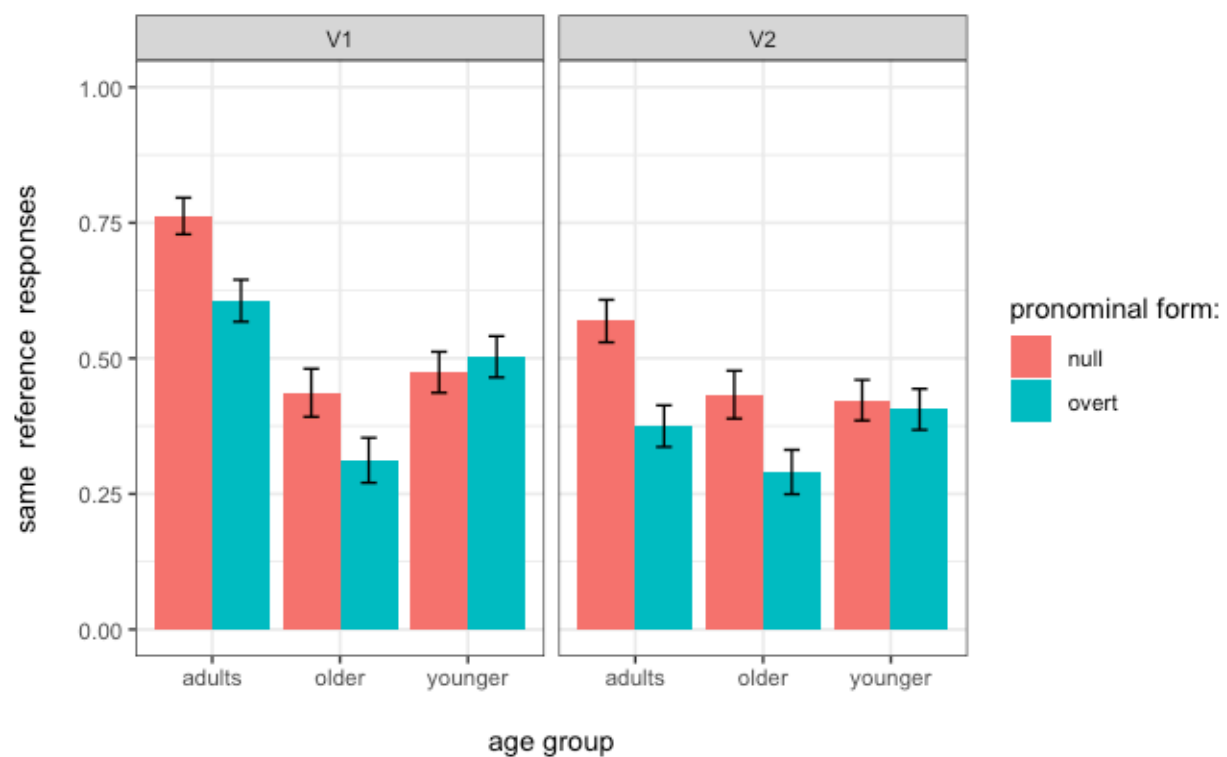
Fig 2: Proportion of subject responses by adults ($N = 40$), older children (4;6-6;4, $N = 33$), and younger children (2;11-4;5, $N = 40$). Left graph shows version 1, with temporal connective *y después* ('and then'); right graphs shows version 2, with causal connective *y por eso* ('and so').

To test the prediction that listeners use the null/overt distinction to help interpret ambiguous subject pronouns, we modeled the likelihood of a same-reference response using multilevel logistic regression models fit to adults, older children, and younger children's data, with pronominal form (null=1, overt=0) and connective (temporal *después* =1, causal *por eso* = 0) as level-1 fixed effects and items and participants as level-2 random intercepts (see appendix Table 3 for full details).

For adults, the null pronoun was associated with a significant increase in the likelihood of a same-reference response (β = 0.95, SE = 0.25, $p <= 0.001$), and so was the temporal connective (β = 1.17, SE = 0.26, $p < 0.001$). There was no interaction, indicating that the contrast between null and overt pronoun preferences remained constant even as the baseline preference changed across different versions of the experimental items.

For older children, the null pronoun was associated with a significant increase in the likelihood of a same-reference response (β = 0.60, SE = 0.28, $p = 0.035$), but there was no effect of connective and no interaction, indicating that their use of the null/overt distinction was stable across different versions of the experimental items. For younger children, the temporal connective was associated with a significant increase in the likelihood of a same-reference response (β = 0.53, SE = 0.24, $p = 0.026$), but there was no effect of pronominal form and no interaction. Neither group showed any evidence of using both pieces of information simultaneously, as adults do.

This result suggests that children initially rely exclusively on the content of lexical connectives to interpret ambiguous pronouns, until age 4 ½ when they switch to relying exclusively on the null/overt distinction. To explicitly test this hypothesis, we designed a regression model for all children's data combined, which included an interaction between age

group and pronominal form and an interaction between age group and connective (see appendix Table 4). This analysis revealed a significant main effect of pronominal form (β = 0.57, SE = 0.20, p < 0.01) and an interaction between pronominal form and age such that being in the younger age group was associated with a reduced effect of pronominal form (β = -0.56, SE = 0.26, p = 0.032). There was no main effect of connective and no interaction between connective and age. In sum, children's use of pronominal form appears to change with age, but there is not solid evidence that their use of connectives changes.

7.3. Discussion and study limitations

This study provides robust evidence that children learn to use pronominal form as a cue pronoun interpretation beginning at age 4 ½ and that children younger than 4 ½ rely on the lexical information provided by temporal and causal connectives. This result is consistent with our hypothesis that children generalize their knowledge of 1st and 2nd person subject realization to not just production but also comprehension of 3rd person pronominal subjects. It also shows that sensitivity to the null/overt contrast begins far earlier than has been previously revealed by felicity judgment tasks (Shin & Cairns 2012, Sorace *et al*. 2009) and even slightly earlier than has been revealed in other pronoun resolution tasks (Papadopoulou *et al*. 2015).

Importantly however, this study does not show that children's comprehension of null and overt pronominal subjects becomes adult-like at age 4 ½. First of all, while children in both age groups were able to incorporate at least one relevant piece of information into their interpretation—lexical connectives for younger children, pronominal form for older children— neither group appears to use both types of information simultaneously. It is unlikely that children

forget the importance of connectives when they reach 4 ½ years of age; rather, we take this as an indication that children may have knowledge of a phenomenon yet still fail to deploy that knowledge under certain circumstances. Older children in our study may know that connectives are relevant to pronoun interpretation yet consider that information less important than newly acquired information like pronominal form, or they may simply be unable to integrate the two types of information together. A second deviation from adult-like behavior is that neither group of children display the same absolute preferences as adults. Both groups appear more strongly biased overall towards the switch-reference reading, as compared to adults. This could be due to processing constraints that render the more recently mentioned non-subject antecedent easier to access, or it could be that children fail to fully incorporate all the relevant pieces of information that adults may be drawing on in this task, such as background knowledge of the world, prior biases and/or heuristic strategies favoring the preceding subject, and so on. Whatever the reason, it is clear that children of this age are still in the midst of sorting out the complex process of pronoun resolution.

The fact that children this age are not done acquiring pronoun resolution should come as no surprise, given our arguments in section 3 about the challenge that pronoun resolution poses to the learner. What is truly impressive is that despite this lack of full competence children this young have already discovered the probabilistic association between pronoun reference and pronoun realization and have internalized it to the point that they can deploy it in both production and comprehension. This state of affairs argues against the idea that children acquire the null/overt contrast by directly tracking the correlation between the realization of 3rd person pronominal subjects and their reference—they do not yet have adult-like command of 3rd person pronoun reference.

8. Conclusion

In this paper we set out to identify the learning path that children follow after they have discovered the target setting of the null subject parameter. Specifically, we sought to understand how children acquiring so-called consistent null subject languages, like mainstream varieties of Spanish and Italian, learn to probabilistically associate the null subject with reference to the preceding subject antecedent, in contrast to the overt pronominal subject. We identified the problem that this task presents to children: namely, that acquiring this contrast requires the child to identify pronoun antecedents in the first place—a non-trivial task. We proposed that children overcome this problem by first tracking the realization of 1st and 2nd person pronominal subjects in their input, whose intended referents are much less complicated to identify than 3rd person pronoun referents.

Our proposal is based on the independently motivated claim by Charnavel (2019) that 1st, 2nd and 3rd person pronouns all share the same underlying formal representation and the idea that the intended referent of 1st and 2nd person pronouns is easier to identify. To show that our proposal is consistent with actual child learning, we first demonstrated that 1st and 2nd person subjects in the input provide the necessary statistical signal (Study 1). Next, we showed that children apply this contrast to their own production of 1st and 2nd person pronominal subjects and that they subsequently generalize this contrast to the realization of 3rd person subject pronouns (Study 1), as well as to the comprehension of ambiguous 3rd person subjects (Study 2), all before the age of 6.

Together, these production and comprehension results suggest a reason for why null and overt subject pronouns have continued their pattern of variable alternation across so many

varieties and so many generations. Simply put, the pattern is acquired early, thanks to children's

ability to zero in on the most informative subsets of their input.