Learning to be inaccurate like an adult: Using computational cognitive modeling to investigate the acquisition of pronoun interpretation in Spanish

Lisa Pearl, Hannah Forsythe

Abstract

When children behave differently from adults in language tasks, it's often unclear if the underlying cause is non-adult-like representation of relevant information, non-adult-like deployment of adult-like representations, or both non-adult-like representations and non-adult-like deployment. We show how computational cognitive modeling can be used to identify which options could lead to specific non-adult-like language behavior, using the case study of Spanish subject pronoun interpretation by typically-developing children. In a picture-selection task, children interpret subject pronouns differently from adults; modeling results suggest that both child and adult pronoun interpretation behavior is best captured by inaccuracy somewhere in the pronoun interpretation process, though how exactly children are inaccurate differs from how exactly adults are inaccurate. So, to become adult-like, children need to learn how to be inaccurate in adult-like ways. We discuss the promise and limitations of the computational cognitive modeling approach demonstrated here for evaluating specific hypotheses about the underlying cognitive computations leading to observed language behavior.

Keywords: Bayesian inference, computational cognitive modeling, inaccurate decision-making, language acquisition, pronoun interpretation, Spanish

1 Introduction

1.1 When children differ from adults

When children produce and interpret language differently from adults, the underlying cause can be unclear: do they have a non-adult-like representation of the target language, or do they simply deploy their representation in a non-adult-like way—or perhaps some of both? For instance, let's consider how someone might interpret a pronoun in context, such as *she* in *Lisa hugged Iona and then she took a nap.* Suppose we believe that adults know that the pronoun in this type of context is interpreted as the subject (e.g., *she* as *Lisa*) 80% of the time. However, suppose we observe a child interpreting the pronoun as a subject 60% of the time. Is this observed behavior because the child represents the relevant probability as 60%, and has accurately deployed that 60% probability? If so, a non-adult-like representation (of probability) is the cause. Or, is this observed behavior because the child has an adult-like probability representation (80%), but has inaccurately deployed that knowledge in the moment (perhaps due to cognitive limitations)? If so, non-adult-like deployment of the representation is the cause. Of course, both non-adult-like representations and non-adult-like deployment of those representations may cause non-adult-like behavior.[1]

A traditional way to try separating out the effects of non-adult-like deployment is to use experimental designs and techniques that facilitate language deployment. If non-adult-like behavior persists, it's more likely due to children having a different representation from adults than to any difficulty deploying their representation; however, if adult-like behavior emerges when deployment is facilitated, this suggests that previous non-adult-like behavior was instead due to deployment difficulties. Deployment can be facilitated in a variety of ways, such as designing tasks with lower processing demands (e.g., Hartshorne *et al.*, 2015; Messenger & Fisher, 2018; Ud Deen *et al.*, 2018) or more natural pragmatics (e.g., Conroy *et al.*, 2009; Spenader *et al.*, 2009), as well as using more sensitive behavioral measures, such as eye gaze (e.g., Brandt-Kobele & Höhle, 2010) instead of pointing and verbal responses.

Still, it can be difficult to know how much facilitation is enough or if the behavioral measure is sensitive enough. Suppose a new task fails to reveal adult-like behavior in children even after improving the task and using a more sensitive performance measure; while that result is less likely due to deployment difficulties, deployment difficulties are nonetheless still possible. Children differ from adults in many ways that can affect their language deployment–probably in ways that have yet to be discovered–so it seems unwise to definitively dismiss deployment difficulties, based on data from behavioral studies. This uncertainty therefore makes it inherently more difficult to interpret children's non-adult-like behavior than to interpret their adult-like behavior.

1.2 Computational cognitive modeling as a tool complementing behavioral techniques

Here, we show how computational cognitive modeling can be fruitfully used to analyze children's non-adult-like behavior. We focus on the case study of Spanish subject pronoun interpretation by typically-developing children. We first present a pronoun interpretation task showing that children acquiring Spanish as their first language interpret subject pronouns differently from adults. In particular, we find that some information about potential pronoun interpretations appears to matter less (or not at all) to children, in contrast with adults.

We next use computational cognitive modeling to explore the potential sources of children's non-adult-like responses, because computational cognitive models allow us to evaluate hypotheses about the mental computations humans do in order to generate observable behavior (Pearl, 2023). Here, we evaluate different hypotheses about how listeners generate their pronoun interpretation responses. These hypotheses involve (i) how listeners represent probability information in their input about pronoun interpretation, and (ii) how listeners deploy that represented information in the moment to interpret a pronoun in an experimental context.

[1]Note that this framing of representation vs. deployment has some overlap with the competence vs. performance distinction introduced by Chomsky (1969). In particular, "competence" is traditionally thought of as knowledge of language, irrespective of cognitive limitations (e.g., memory limitations) or other factors (e.g., attention shifts, distractions). This idea aligns with knowledge "representation" in our framing. In contrast, "performance" is traditionally thought of as the actual use of language with cognitive limitations and other factors in effect – that is, the observed linguistic behavior. In our framing, knowledge "deployment" is subject to cognitive limitations and other factors, and so the observed behavior ("performance") is due to knowledge representations ("competence") being deployed in the moment.

More specifically, we investigate whether the patterns of behavior observed in children are more likely caused by children (i) inaccurately representing probability information available for pronoun interpretation, (ii) deploying accurate probability representations inaccurately, or (iii) both inaccurately representing pronoun probability information and inaccurately deploying those inaccurate probability representations. Importantly, we also use the same modeling techniques to capture adult behavior in the same task, and so identify the target state for children, both for representing probability information and for deploying that information. This allows us to identify more precisely what needs to change in children in order for them to develop adult-like ways of interpreting pronouns.

1.3 "Adult-like" doesn't necessarily mean accurate

Importantly, the observable target of development is adult-like behavior, whatever that happens to be. We might think that "adult-like" always means accurate, such as adults giving correct responses on behavioral tasks. However, sometimes adult behavior isn't accurate, such as when adults need to balance competing constraints. A classic example of balancing constraints in human cognition is the "speed-accuracy tradeoff" (see Heitz 2014 for an overview): a participant's reactions are limited in time, so the participant may sacrifice accuracy of a response in favor of making a speedier response.

So again, the acquisition target is adult-like behavior, even if that behavior isn't accurate. One perspective on adult cognition is that adults have identified "optimal" behaviors, where they achieve the optimal balance of competing constraints (e.g., balancing speed of a decision vs. the accuracy of that decision). More specifically, when it comes to how adults use information to make decisions, there's a rich literature exploring whether–and how–they may be optimal (e.g., see Lieder & Griffiths 2020 for a recent review). For instance, a long-standing hypothesis in cognitive psychology has been the idea of "bounded rationality" (e.g., Simon, 1956; Gigerenzer, 2008; Zhang *et al.*, 2020), where humans intend to be optimal (i.e., "rational" and therefore accurate), but can't actually be optimal in practice because of human cognitive limitations. So, in practice, the optimality of human cognition is "bounded" by the constraints imposed by the implementation of human cognition in the mind and brain. For example, the speed-accuracy balance that adults achieve would be impacted by the limitations of achieving that balance in the medium of the human mind.

More recently, this approach has been extended to the idea that humans may be "resource-rational" (Vul *et al.*, 2014; Lieder & Griffiths, 2020): humans are optimal ("rational") when it comes to how they deploy their limited cognitive resources to accomplish a task. That is, the optimization that the human mind does may be be targeted at optimizing cognitive constraints like energy efficiency or processing time (Friston, 2010; Markman & Otto, 2011; Martin, 2016). So, humans may appear to be suboptimal (i.e., "irrational" and therefore inaccurate) when it comes to using available information if cognitive limitations aren't considered; however, if we consider that optimization includes the efficient use of limited cognitive resources, then human behavior once again may be optimal.

1.4 Resource-rationality in language acquisition

Within language acquisition, optimizing with respect to cognitive resources is actually a cornerstone of at least two approaches we're aware that investigate when a child would decide to adopt a rule or generalization. The first approach uses the Tolerance and Sufficiency Principles (Yang, 2005, 2016, 2018), which both assume that a learner is trying to optimize the average retrieval time for any item that a potential rule or generalization could apply to. That is, this decision process is predicated on humans wanting to make retrieving information as efficient "time-wise" as possible.

The second approach is Minimum Description Length (Li & Vitányi, 1994; Rissanen & Ristad, 1994; Stabler, 1998; Hsu & Chater, 2010; Hsu *et al.*, 2011, 2013; Chater *et al.*, 2015), which assumes the learner is optimizing the amount of space for storing information, with a preference for anything that makes storing information more compact. That is, this decision process is predicated on humans wanting to make information as efficient "space-wise" as possible. In both cases, whether the goal is efficiency in terms of time or efficiency in terms of space, humans are assumed to optimize with respect to limited cognitive resources (e.g., processing time or mental storage space).

So, being inaccurate may in fact be optimal (or at the very least, a plausible adult-like thing to do) once limited cognitive resources are taken into account. More specifically, we might initially think that the optimal approach to pronoun interpretation in context would be to accurately use all information available from all relevant pronoun interpretation cues. However, in the face of cognitive resource limitations, the resource-optimal approach would be *not* to do this; instead, humans may optimize processing efficiency (either time-wise, space-wise, or some other way) to achieve "good enough" pronoun interpretation (similar to "good enough" language comprehension approaches more generally: Ferreira *et al.* 2002; Ferreira & Patson 2007; Traxler 2014).

1.5 Our findings: A preview

Our modeling results suggest that both child and adult pronoun interpretation behavior is best captured by inaccuracy somewhere in the pronoun interpretation process. Child behavior is best captured if children are doing one of two things: (i) they always deploy inaccurate probability representations of their input, or (ii) they selectively deploy accurate probability representations. Adult behavior is best captured if adults are always deploying inaccurate probability representations. Importantly, this means that adult-like behavior – the target of acquisition – isn't accuracy: to become adult-like, children need to learn how to be inaccurate in the right ways for potentially both representation and deployment of probability information relating to pronoun interpretation.

More generally, this case study demonstrates how computational cognitive modeling complements existing behavioral techniques investigating language acquisition: modeling allows us to test specific hypotheses about the underlying cognitive computations that lead to observed behavior, both in children and adults. We conclude with a brief discussion of the promise and limitations of the computational cognitive modeling approach demonstrated here.

2 Behavioral data: Child pronoun interpretation in Spanish

2.1 Pronoun interpretation in Spanish

Pronoun interpretation requires listeners to combine information from multiple sources in order to decide how to interpret that pronoun (i.e., what the pronoun's antecedent is), and this information can conflict. For instance, Spanish listeners can interpret subject pronouns by attending to at least three different kinds of information in the current linguistic context[2]: (i) the pronoun's form (Keating *et al.*, 2011; Otheguy *et al.*, 2010; Carvalho *et al.*, 2015), (ii) the relation (sometimes called "coherence" or "rhetorical relation") between the pronoun's clause and other clauses in the surrounding discourse (Kehler, 2002; Asher & Lascarides, 2003; Kehler *et al.*, 2008; Fukumura & van Gompel, 2010), and (iii) the pronoun's grammatical features (Clahsen *et al.*, 2002; Pérez-Leroux, 2005; Legendre *et al.*, 2014). Each kind of information can be extracted from one or more cues.

For pronoun form information, Spanish subject pronouns have two options: null (e.g., Ø in (1)) and overt (e.g., the form *él* ('he') in (1)). These pronoun forms serve as the cues listeners can attend to, and have different interpretation preferences: the null form tends to be more strongly associated than the overt form with the preceding subject antecedent[3] (Filiaci, 2010; Keating *et al.*, 2011; Otheguy *et al.*, 2010; Carvalho *et al.*, 2015). So, the null pronoun Ø in (1) is more likely to be interpreted as referring to the subject *Juan* than if the overt pronoun *él* were used in the same context (i.e., the probability that Ø refers to Juan > the probability that *él* refers to Juan).

(1) Juan llamó a Pedro cuando **Ø /él** estaba en casa. $p(\text{Ø}\rightarrow \textit{Juan}) > p(\textit{él}\rightarrow \textit{Juan})$
Juan called A Pedro when *pro*/he was in house.
'Juan called Pedro when (he) was at home.'

For clausal relation information, Spanish has a range of syntactic and lexical cues that can be attended to, such as the Spanish connectives *y después* ('and after') and *porque* ('because') in (2). Temporal relations like those indicated by *y después* are biased toward maintaining continued reference to the topic (Asher & Lascarides, 2003). Because the subject is typically the topic, this biases the pronoun interpretation towards the subject; so, the pronoun in (2a) is more likely to refer to the subject antecedent *Juan* than to the non-subject *Pedro* (i.e., the probability that *y después* causes Ø to refer to Juan > the probability that *y después* causes Ø to refer to Pedro).

(2) Juan le dice adiós a Pedro...
Juan DAT says bye to Pedro

a. ...**y después** Ø se va. $p(\textit{y después}\rightarrow \textit{Juan}) > p(\textit{y después}\rightarrow \textit{Pedro})$
...and then *pro* leaves.

b. ...**porque** Ø se va. $p(\textit{porque}\rightarrow \textit{Pedro}) > p(\textit{porque}\rightarrow \textit{Juan})$
...because *pro* leaves.

'Juan is saying bye to Pedro and then/because (he) is leaving.'

[2] We note that the specific information listeners attend to may result from general-purpose causes, such as optimizing communicative efficiency. Our thanks to Titus von der Malsburg for bringing this point to our attention.

[3] One reason for this stronger association may be because the subject is often the topic of discourse, and null subjects prefer to be associated with the discourse topic (see Vogelzang *et al.* 2021 for more discussion of this point).

In contrast, causal sequences like those indicated with the connective *porque* can be biased in either direction, depending on the content of the specific predicates involved (Asher & Lascarides, 2003; Fukumura & van Gompel, 2010). In (2), the specific predicates lead to a bias for interpreting the pronoun as the object antecedent *Pedro* in (2b) (i.e., the probability that *porque* causes Ø to refer to Pedro > the probability that *porque* causes Ø to refer to Juan).

For grammatical feature information, Spanish subject-verb agreement is a cue that can be attended to. Notably, this cue is available even when the pronoun itself happens to be null because verbs in Spanish must carry morphology that overtly signals the subject's person and number. An example of this is in (3), with the third person singular morphology on the verb *sale* ('leaves').

(3) Las niñas saludan a la maestra, y después Ø sal**e** p(*sale*→*la maestra*) ≈ 1
the girls greet A the teacher, and then *pro* leave-3Sg
'The girls wave at the teacher, and then (she) leaves.'

In contrast to the probabilistic cues of pronoun form and lexical connective, number agreement is traditionally thought to be categorical when it comes to its effects on pronoun interpretation. In (3), the verb morphology aligns with the non-subject antecedent: the singular morphology on the verb *sale* indicates that the null pronoun is also singular, which in turn indicates that the pronoun must be interpreted as the singular *la maestra* ('the teacher', the object) rather than the plural *las niñas* ('the girls', the subject). That is, the probability that *sale* causes Ø to refer to *la maestra* is approximately 1. Note that this interpretation persists even though the other cues probabilistically favor the other interpretation. That is, the pronoun's form is null and the connective is the temporal *y después*, both of which favor the subject antecedent (*las niñas*, 'the girls'); yet, the cue of agreement morphology forces the interpretation to be the non-subject antecedent (*la maestra*, 'the teacher').

This last example demonstrates one of the main problems children must learn to solve when interpreting pronouns: how to interpret a pronoun when multiple conflicting cues are available. In particular, children must learn two key things.

First, they must learn the information each cue carries in an adult-like way – we refer to this as the pronoun cue's *representation*. For example, the adult-like representation of a Spanish null subject pronoun could encode a bias towards subject antecedents. That is, the probability that the null form indicates a subject antecedent is greater than the probability that the overt form indicates a subject antecedent: p(Ø→subject) > p(overt pronoun *él*→subject). This probability is one example of the representation children might need to learn for this cue's information.

Second, children must learn how to integrate the information from different cues in an adult-like way – we refer to this as the *deployment* of their representations. For instance, in (3), a null pronoun form and the connective *y después* are used, which both favor the subject interpretation; however, since the accompanying number morphology is singular and the subject antecedent *las niñas* is plural, the pronoun shouldn't be interpreted as referring to the subject. So, children must learn to prefer the antecedent that adults do (e.g., the subject) by integrating the conflicting cue information in an adult-like way.

It's currently unknown both when Spanish-learning children acquire adult-like pronoun cue representations and when they achieve adult-like deployment of these representations. That is, for the probability-based information discussed here, it's unknown when children would achieve adult-like probability representations for the different cue information, and when children would integrate these probability representations in an adult-like way to achieve adult-like pronoun interpretation.

2.2 What we know about when children learn what

All three information sources illustrated above – pronoun form, clausal relations, and grammatical features – have been studied to different degrees on their own. Here, we provide a synthesis of the relevant information reviewed in more detail in Appendix A. In particular, the evidence collectively suggests that by age four or five, children are capable of using a variety of cues to interpret pronouns in an adult-like way, including the pronoun's form, the presence of various lexical and syntactic cues that reveal its semantic relation with surrounding clauses, and the morphological features of verbal agreement. However, children's ability to deploy this knowledge can vary depending on how cognitively-demanding the task is and whether these cues conflict with each other. The variation in children's performance makes it unclear whether children's representations of these pronoun cues are still developing at this age or are instead fully adult-like but simply obscured by problems with deployment.

2.3 Assessing child pronoun interpretation behavior when cues conflict

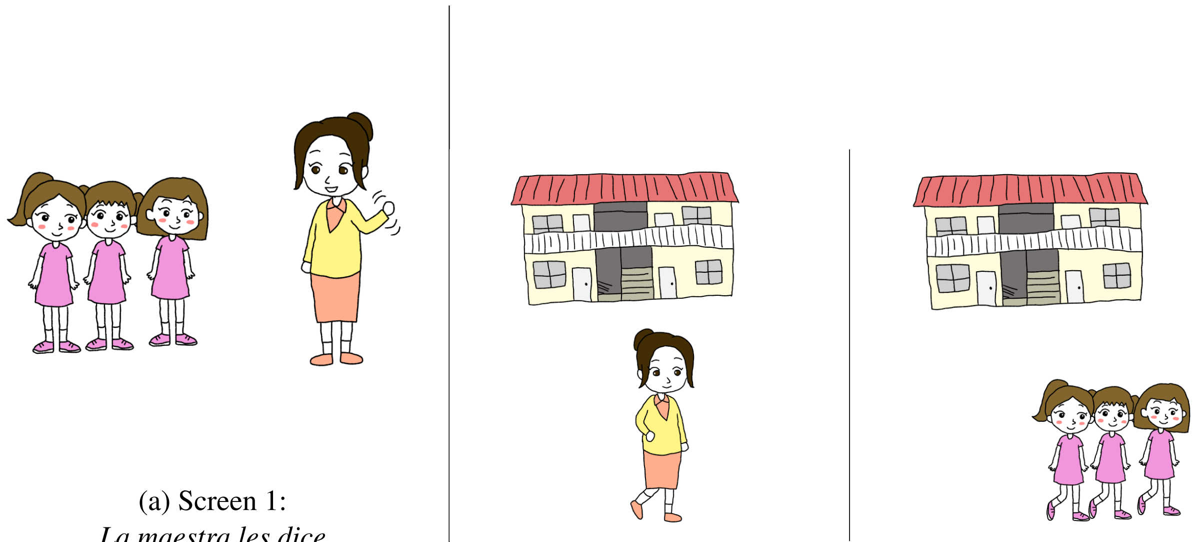
Here, we describe a forced-choice picture-selection task (see Figure 1) used to elicit preferred interpretations of subject pronouns, given each of the three cue types discussed above. This paradigm allows us to observe how children and adults change their preferred interpretations when each individual cue changes, such as switching from a null to an overt pronoun form, switching between different lexical connectives, or switching between subject-aligned vs. non-subject-aligned agreement morphology. The task is similar to earlier picture-selection tasks testing children's comprehension of agreement (e.g. Pérez-Leroux 2005) but incorporates two techniques that lessen the cognitive demand, in order to facilitate children's deployment of their pronoun knowledge. First, both potential antecedents are explicitly mentioned in the clause immediately preceding the pronoun (Screen 1 in Figure 1: the teacher, the girls) in order to make them as prominent in the discourse context as possible.[4] Second, both potential interpretations are depicted side by side (Screen 2 in Figure 1: the teacher leaving, the girls leaving) in order to lessen working memory demands.

All experimental items included one of two pronoun forms (null ø or overt *ella(s)* 'she'/'they (f)'), a discourse connective (temporal *y después* 'and then' or causal *porque* 'because'), and an agreement morphology marker (3rd singular *-ø* or 3rd plural *-n*). To assure that agreement morphology information was only indicated by the verb (rather than also on the pronoun if it was overt), experimental items were designed following the approach of Johnson *et al.* (2005). This approach masks the number marking on specific overt pronoun forms like *ella* and *ellas* (i.e., whether the ending is -ø or -*s*), and is described in more detail below in section 2.3.2.

To test how strongly each cue affects pronoun interpretation, cues were alternately aligned with each other, as in (4a), or set up so that one was pitted against the others, as in (4b)-(4d).[5]

[4]For discussion of recommendations about how to phrase lead-in sentences to facilitate a supportive pragmatic context, see Crain & Thornton (1998).

[5]Table 7 in Appendix B describes the full experimental paradigm.



(a) Screen 1: *La maestra les dice adiòs a las niñas...* "The teacher says goodbye to the girls..."

(b) Screen 2: *...y después ø se va a la casa* "...and then *pronoun* goes home"

Figure 1: Example picture-selection trial. TASK: Listen to the story represented by the images in Screen 1 and choose one picture in Screen 2 that represents the rest of the story. In Screen 2, the subject interpretation is indicated on the lefthand side by the teacher leaving; the non-subject interpretation is indicated on the righthand side by the girls leaving.

(4) Example utterance opening for an experimental item:

"La maestra les dice adiòs a las niñas ..."
The teacher to-them say-3SG goodbye A the girls ...

Possible utterance continuations:

a. All aligned: All cues favor the subject
Form → subject, Connective → subject, Agreement Morphology → subject
"...y después ø se va a la casa."
...and then *pro* SELF go-3SG to the house

b. Form not aligned (it favors the non-subject) while the other two cues favor the subject
Form → **non-subject**, Connective → subject, Agreement Morphology → subject
...y después **ella** se va a la casa.
...and then **she** SELF go-3Sg to the house

c. Connective not aligned (it favors the non-subject) while the other two cues favor the subject
Form → subject, **Connective** → **non-subject**, Agreement Morphology → subject
"...**porque** ø se va a la casa."
...**because** *pro* SELF go-3Sg to the house

d. Agreement Morphology not aligned (it favors the non-subject) while the other two cues favor the subject
Form → subject, Connective → subject, **Agreement Morphology** → **non-subject**

"...y después ø se va**n** a la casa."
...and then *pro* SELF go-**3Pl** to the house

Recall from section 2.1 that two of these cues (pronoun form, lexical connective) are probabilistic, while agreement morphology is assumed to be more categorical (that is, this cue's probability is very near 0 or 1). We therefore expect adults to be much more strongly influenced by agreement morphology than by the other two cues. For example, we would expect a dramatic difference in subject interpretation responses between condition (4a), where all cues signal the subject, and (4d), where agreement morphology signals the non-subject in opposition to the other two cues. However, we would expect a much smaller difference between (4a) and (4b-4c), where only the lexical connective and the pronoun form signal the non-subject, respectively. How and when children differ from adults in these conditions can help us identify which cues children may be heeding.

2.3.1 Participants

Participants were recruited from a daycare in Mexico City, Mexico. Adult participants were all native speakers of Spanish working at the school. Children were typically-developing learners of Spanish as a first language. A total of 47 adults (43 women) and 97 children (57 girls) ages 1;11 to 6;9 completed the task. One adult (n=1) was excluded from analysis due to experimenter error (only half the items were administered). Children were divided into three age groups for analysis: 33 children age three and under (≤3: 1;11-3;10; *M* = 3;3), 35 children age four (4: 4;0-4;11; *M* = 4;5), and 29 children age five and older (≥5: 5;0-6;9; M = 5;8). Individual responses were excluded if the participant failed to choose one of the pictures, or if the experimenter read the prompt incorrectly. The following child responses were excluded for this reason: 29 (of 1225) age ≤3; 22 (of 1308) age 4; 21 (of 1081) age ≥5.

2.3.2 Design and procedure

Experimental stimuli were created by fully crossing the three cues of pronoun form (null, overt), discourse connective (temporal *y después*, causal *porque*), and agreement number morphology (aligns with subject, aligns with non-subject). That is, in addition to the condition in (4a), where all three cues favor the subject interpretation, and the three conditions in (4b)-(4d) where two of three cues favor the subject, there were also three conditions where two of three cues favored the non-subject, and one condition where all three cues favored the non-subject.[6] These 8 conditions were presented in both singular and plural forms, for a total of 16 conditions.

Eight distinct experimental items were created by choosing pairs of verbs that were easily depicted and likely to be known by children under age three (e.g., *sigue–sale*: 'follow–go out'). In addition, as mentioned previously, we followed Johnson *et al.* (2005) to mask agreement morphology information that could come from the singular -ø or plural -*s* marking on the subject of the second clause (i.e., *ella(s)*); we accomplished the masking by using /s/-initial predicates (e.g., *sale*). This way, in overt pronoun conditions, number marking on the pronoun subject *ella(s)* ('she/they (f)') wouldn't provide participants with any additional number cues (e.g., "*ella sal...*" is difficult to distinguish from "*ellas sal...*"). Instead, participants would need to rely on the agreement cue provided by the verbal agreement marker (e.g., *sale(n)*).

[6]This paradigm is described more fully in Appendix B.

Fillers were created by choosing an additional 16 verb pairs, replacing the pronoun with the definite DP *los niños* ('the boys'), and using either 'the girls' or 'the teacher' as a competing antecedent. The same 16 fillers were presented in the same order across all versions of the experiment. See Appendix B for more details of the design and sample items.

Participants were randomly assigned to one of three versions of the experiment with 16 experimental trials and 16 filler trials. Each version tested a different cue type (pronoun form, connective, morphology) by systematically aligning and pitting that cue against the other two. For example, participants assigned to the "morphology" version were exposed to 8 "congruent" trials in which all cues favored the same antecedent (as in 4a) and 8 "incongruent" trials in which agreement morphology favored the opposite antecedent that the pronoun form and connective favored (as in 4d). The same protocol was used for participants assigned to the "pronoun form" and "connective" versions (pronoun form: 4a vs. 4b; connective: 4a vs. 4c). Thus, cue type was between-subject while congruency (i.e., congruent, incongruent) was within-subject.

Within each version of the experiment, fillers and experimental items were presented in a fixed order that formed a coherent narrative arc about two consecutive days at a school with a teacher, a group of girls, and a group of boys. Each experimental item was presented twice, once in its congruent form and once in its incongruent form, which were randomly assigned to appear on either "day 1" or "day 2". Each pair of congruent/incongruent experimental items was randomly assigned to appear in the plural or the singular. The total experiment (3 practice items, 16 experimental items, and 16 fillers) took about 10 minutes to administer.

The task was administered on a 13" MacBook Air using Psychopy version 3.0.0b11 (Peirce *et al.*, 2019). The task began with an introductory screen introducing the characters (a teacher, 3 identically-dressed girls, 3 identically-dressed boys). Next, participants were given an explanation of the task and 3 practice trials. During each practice and experimental trial, participants saw an illustration of the first clause as it was read out loud by the experimenter (child participants) or a recording of her voice was played over headphones (adult participants), as in Screen 1 of Figure 1. Then, participants saw a blank screen and heard the second clause. Then, two illustrations appeared, one corresponding to the subject interpretation and one to the object interpretation (as in Screen 2 of Figure 1); participants chose the picture that matched their own interpretation by pointing (child participants) or by pressing the '4' or '9' key (adult participants). Pictures were randomly placed on the left or right side of the screen. Upon completion, children received a piece of candy and adults received the equivalent of US$10.

2.4 Behavioral results

2.4.1 Practice and filler trials

For the practice trials, we used a one-sided t-test to evaluate whether the correct response rate in each age group was above chance, and found that indeed it was ($\leq$3: M=0.68, t(32)=4.37, p<0.001; 4: M=0.66, t(34)=3.60, p<0.01; $\geq$5: M=0.71, t(28)=4.12, p<0.001; adults: M=0.89, t(46)=17.09, p<0.001). For the filler trials, we again used a one-sided t-test to evaluate whether the correct response rate in each age group was above chance, and found that indeed it was ($\leq$3: M=0.69, t(32)=4.45, p<0.001; 4: M=0.67, t(34)=3.34, p<0.01; $\geq$5: M=0.85, t(28)=7.727, p<0.001; adults: M=0.94, t(46)=19.71, p<0.001).

Because participant performance on both the practice and filler trials was above chance, this suggested that the participants in all age groups understood the task and were capable of completing it as intended. However, because the mean score on the filler trials for children under age five seemed much lower (.67-.69 vs. .85), we report and discuss the child results only for children age five and older (≥5).

2.4.2 Results: Experimental trials

Figure 2 shows the rate of subject antecedent responses (i.e., interpreting the pronoun to refer to the potential antecedent in the subject position) produced for each combination of cues (pronoun form, discourse connective, and agreement morphology) for both children and adults.

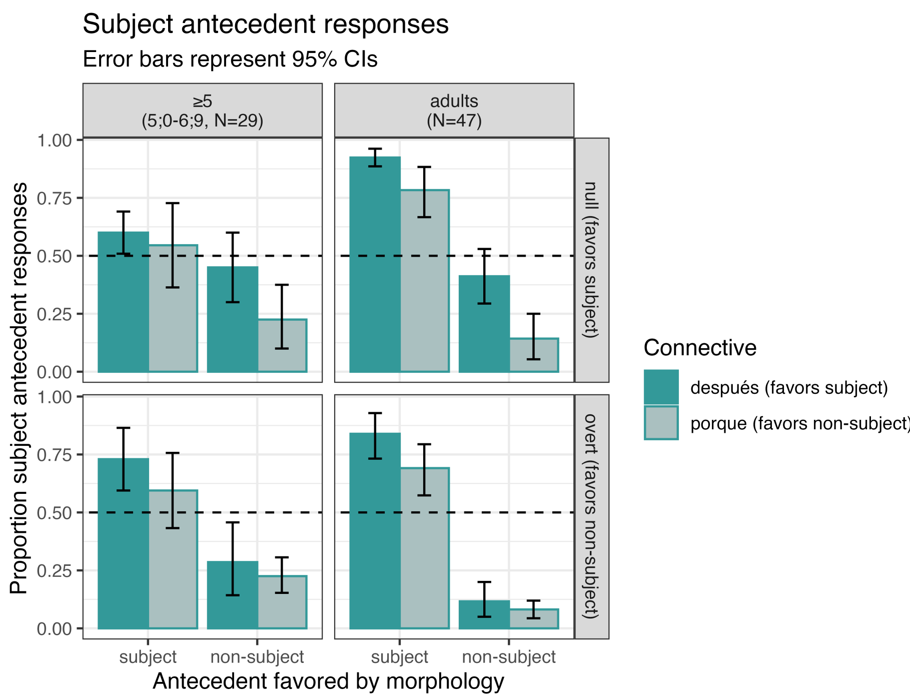


Figure 2: Rate of subject antecedent responses by children (≥5 years old) and adults interpreting Spanish pronouns in utterances where the pronoun form favors the subject (null) or the non-subject (overt), the discourse connective favors the subject (*y después*) or the non-subject (*porque*), and agreement morphology favors the subject or the non-subject. Error bars represent 95% confidence intervals.

In general, we can see that children appear to interpret pronouns differently from adults, even when all three cues favor the same antecedent. For instance, children have fewer subject interpretations in conditions with subject-favoring cues (the null pronoun form, the temporal connective *y después*, and agreement morphology that aligns with the subject). Also, child responses seem much less extreme than adult responses: when adults strongly prefer the subject antecedent, children prefer the subject less strongly; when adults strongly prefer the non-subject antecedent, children prefer the non-subject less strongly.

Interestingly, adult responses indicate that agreement morphology information is perhaps not as categorical at determining adult interpretations as previously thought. That is, there are some contexts where adults seem to ignore the information coming from agreement morphology and interpret the pronoun in a way that conflicts with this information. For instance, when agreement morphology favors the non-subject, but the pronoun form and the discourse connective favor the subject (form=null, connective=*y despuès*), adults interpret the pronoun as the subject 41.2% of the time (SE: 0.06), in spite of the agreement morphology. This adult interpretation behavior underscores how adult behavior isn't always accurate behavior (in this context, "accurate" behavior would always have the pronoun interpreted as the non-subject, to align with the agreement morphology). While this adult interpretation behavior may be surprising at first glance, agreement attraction errors – where agreement morphology information appears to be disregarded – are well-documented in adults (Wagers *et al.*, 2009; Acuña-Fariña *et al.*, 2014; Lago *et al.*, 2015, e.g.). This (mis)interpretation behavior occurs in certain contexts that provide conflicting cues to the agreement morphology information, just as the contexts here do. The question then is why inaccurate interpretation behavior occurs – that is (as with children), it's unknown whether this behavior is caused by inaccurate representations of pronoun information, inaccurate deployment of accurate representations during the experiment, or both inaccurate representations and inaccurate deployment of those inaccurate representations.

To identify if each cue cue significantly affected the probability of interpreting the pronoun as referring to the subject antecedent (whatever the underlying reason), we analyzed the data using mixed-effects logistic regressions with the *lme4* package in R (Bates *et al.*, 2015a).[7] Following a data-driven approach that includes both fixed-effect and random-effect terms if they improve the model fit (Bates *et al.*, 2015b; Matuschek *et al.*, 2017; Sonderegger, 2023), we selected the model structure in (5) that includes two-way interactions with age, by-participant and by-item random intercepts, and by-participant random slopes for agreement morphology (see Appendix C for details of the model selection process).

(5) Model with age interactions and by-participant random slopes for agreement morphology

$$\text{subj-antecedent} \sim \text{form*age + connective*age + morph*age} \\ + (1|\text{item}) + (1 + \text{morph}|\text{participant})$$

The analysis in Table 1 suggests that the agreement morphology cue has the strongest effect (Agremeent Morph: β=0.78, $p < 0.001$), which increases with age (Agreement Morph x Age: β=1.04, $p < 0.001$). In contrast, the other two cues either have a main effect that doesn't increase with age (connectives: Connective β=0.32, $p = 0.025$; Connective x Age: $p = 0.159$) or no significant effect, but one that does change with age (pronoun form: Form p=0.504; Form x Age: β=0.56, $p = .006$). To interpret the significant interactions with age for pronoun form and agreement morphology, we used the `emmeans` package (Lenth, 2021) to conduct simple effects analyses. For each predictor showing a significant interaction with age, we tested (i) the effect of the predictor within each age group separately (simple effects), and (ii) whether these effects differed significantly between age groups (interaction contrasts).

Our analysis in Table 2 suggests that age moderated the cue effects of both pronoun form (β = 1.11, $p = .006$) and agreement morphology (β = 2.07, $p < .001$). Notably, only adults

[7] R code implementation of all statistical analyses and graphical plots is available at *[link to author github repository]*.

Fixed Effects	β	SE	z	p	
Intercept	-0.33	0.16	-2.10	.036	*
Form	-0.10	0.14	-0.67	.504	
Connective	0.32	0.14	2.24	.025	*
Agreement Morph	0.78	0.21	3.79	<.001	***
Age group	0.32	0.18	1.80	.071	
Form × Age	0.56	0.20	2.77	.006	**
Connective × Age	0.28	0.20	1.41	.159	
Agreement Morph × Age	1.04	0.27	3.84	<.001	***

Random Effects	Variance	SD
Participant: Intercept	0.09	0.31
Participant: Morphology	0.77	0.88
Item: Intercept	0.07	0.26

Note. N = 1,181 observations from 76 participants across 8 items.
Random effects correlation between participant intercept and morphology slope = -0.18.

Table 1: Mixed-effects logistic regression results for model `subj-antecedent ∼ form*age + connective*age + morph*age + (1|item) + (1 + morph|participant)`, including β coefficients, standard deviation (SE), z-values, and p-values. Significant effects are reported at alpha level p<0.05 *, p<0.01 **, or p<0.001 ***.

Predictor	**Contrast/Group**	**Est.**	**SE**	z	p	
Pronoun Form						
Simple Effects: Null form vs. Overt form						
	Children: Null - Overt	0.19	0.29	0.67	.504	
	Adults: Null - Overt	-0.92	0.28	-3.25	.001	**
Interaction						
	Children vs. Adults	1.11	0.40	2.77	.006	**
Agreement Morphology						
Simple Effects: Non-Subject-Aligned (Non) vs. Subject-Aligned (Aligned)						
	Children: Non - Aligned	-1.56	0.41	-3.79	<.001	***
	Adults: Non - Aligned	-3.63	0.37	-9.88	<.001	***
Interaction						
	Children vs. Adults	2.07	0.54	3.84	<.001	***

Note. Estimates on log-odds scale. Negative values indicate higher log-odds when predictor favors subject antecedent.
*** $p < .001$, ** $p < .01$

Table 2: Simple effects and interaction contrasts from estimated marginal means analysis, reporting the estimate (Est.), standard deviation (SE), z-values, and p-values.

were sensitive to the null/overt pronoun form distinction (≥5 children: $p = 0.504$; adults: OR = 2.51, $p = .001$). In contrast, both children and adults were sensitive to agreement morphology ($p < 0.001$), but adults were markedly more sensitive (≥5 children: OR = 4.76; adults: OR = 37.7).

Taken together, our analysis suggests that child pronoun interpretations at age five are influenced by agreement morphology and discourse connectives, but not by pronoun form. To become adult-like, children's pronoun interpretations must be influenced more strongly by agreement morphology, and must additionally be influenced by pronoun form.

2.5 Discussion

This pronoun interpretation task reveals that children behave differently from adults when it comes to using a combination of information to interpret pronouns, namely pronoun form, discourse connectives, and agreement morphology. Interestingly, the children tested here don't seem to use pronoun form during comprehension, in contrast to adults who choose a subject antecedent more often for a null form. We note that these results fail to replicate earlier work showing child sensitivity to the cue of pronoun form in both production (Forsythe *et al.*, 2019) and comprehension (Forsythe *et al.*, 2022) by age four and a half. The current task may have been more difficult for children because it provided them with three relevant cues at once instead of two.

In contrast to pronoun form, children seem to use the discourse connective to guide their pronoun interpretation as adults do. These results align broadly with Forsythe *et al.* (2022), who also found sensitivity to the cue of discourse connectives in children; these results also align broadly with Torregrossa *et al.* (2019), who found that discourse factors may matter more than grammatical factors during children's pronoun interpretation.

For agreement morphology, children do use this cue, but not as much as adults do. These results align with previous findings that children's early perception and production of agreement morphology doesn't automatically translate into adult-like use of this cue in comprehension tasks (Johnson *et al.*, 2005; Pérez-Leroux, 2005; Gxilishe *et al.*, 2009; Rastegar *et al.*, 2012; Legendre *et al.*, 2014).

Importantly, when we see deviation from adult behavior (as we do in this pronoun interpretation task), the underlying cause is unclear. Perhaps the differences are due to an inaccurate representation of the information (e.g., children don't represent that a null pronoun form signals the subject antecedent more often); perhaps the differences are due to inaccurate deployment of a representation (e.g., children don't access the adult-like representation of what a null pronoun form signals); perhaps both inaccurate representations and inaccurate deployment are the cause. That is, children may fail to rely on the information available (or not rely on it as much as adults rely on it) for different reasons.

For instance, let's consider children's non-adult-like use of pronoun form. One possibility is that children have an inaccurate representation of pronoun form. For example, perhaps children view the subject-favoring null form as being less indicative of the subject antecedent than it truly is. Suppose also that these children believe that the subject-favoring null form favors the subject less than other non-subject-favoring cues favor the non-subject (like the discourse connective); then, when the null form signals the subject antecedent while the discourse connective signals the non-subject antecedent, these children would favor the non-subject antecedent (unlike adults). This non-adult-like behavior would be due to the inaccurate representation of pronoun form information.

Another possibility for children's non-adult-like behavior with pronoun form is that they deploy its information inaccurately. For instance, perhaps children fail to notice in the moment whether the pronoun element was overt, and so lose access to that form information. Or, perhaps these children correctly perceive the pronoun form, but the information it carries decays too much in short-term memory before they can use it. If the information provided by the pronoun form is lost, then it can't be deployed in the moment and integrated with the information from other cues. So, a child could have an adult-like representation of pronoun form, where the null form favors a subject antecedent, but be unable to deploy that representation appropriately in the moment, due to misperception or memory decay. This deployment difficulty could cause these children to prefer the non-subject, despite having an adult-like representation that the null form favors the subject.

The above possibilities are merely some of the potential underlying causes of non-adult-like behavior in this task. There may well be other cognitive limitations that prevent children from behaving like adults when it comes to interpreting pronouns in this context. While improved versions of this pronoun interpretation task may be able to probe underlying causes (particularly deployment issues), a complementary approach is to use computational cognitive modeling.

3 Using modeling to understand pronoun interpretation

Computational cognitive modeling is a technique that allows us to concretely implement specific cognitive theories in order to evaluate them (Pearl, 2023). Here, we can use computational cognitive modeling to simulate how listeners use their input to generate the behavioral output in the pronoun interpretation task discussed in the previous section (i.e., the rate of choosing the subject antecedent as the pronoun's interpretation in the experimental context). More specifically, we can implement how children or adults represent available probabilistic information for interpreting pronouns on the basis of the input encountered (i.e., the representation of pronoun probability information), as well as how they use those probabilistic cue representations in the moment to identify a pronoun's interpretation in context (i.e., the deployment of those probability representations). In other words, we can model the decision a listener – whether child or adult – makes about how to interpret a pronoun in an experimental context, on the basis of (i) what that listener has learned from her input about pronoun interpretation, and (ii) how the listener uses that learned information at the moment of interpreting a particular pronoun. We can then compare these modeled decisions against the decision data we have from the individual adults and children who participated in the pronoun interpretation task.

In the model, we can also specify how listeners combine information of different kinds to generate a preferred interpretation. Here, we use a Bayesian modeling framework, which has been used to understand a variety of child language acquisition behaviors (e.g., speech segmentation: Pearl *et al.* 2011; Phillips & Pearl 2014a,b, 2015a,b; morphosyntax: Gagliardi & Lidz 2014; Gagliardi *et al.* 2017; syntax: Mitchener & Becker 2010; Perfors *et al.* 2011; Orita *et al.* 2013; Pearl & Mis 2016; Nguyen & Pearl 2019; Pearl & Sprouse 2019), as well as adult language behaviors (e.g., for pronoun interpretation specifically: Haghighi & Klein 2007; Kehler *et al.* 2008; Rohde & Kehler 2014). Bayesian inference implements a particular mechanism of combining information (i.e., the inference mechanism), and models relying on this inference mechanism have been able to capture several child behavioral phenomena. Moreover, there's a considerable body of evidence suggesting that young children are capable of Bayesian inference (3 years: Xu & Tenenbaum 2007; 9 months: Gerken 2006; Dewar & Xu 2010; Gerken 2010; 6 months: Denison *et al.* 2011, among many others). So, Bayesian inference seems a plausible mechanism for a computational cognitive model meant to capture child and adult behavior.

Here, we use a Bayesian computational cognitive model to understand the most likely underlying cause of adult pronoun interpretation behavior in the experimental context discussed above, as this is the target state of acquisition; we can also use this model to understand children's non-adult-like pronoun interpretation behavior in the same task. We follow the Bayesian inference approach of Gagliardi *et al.* (2017), adapting it to the task of pronoun interpretation. The modeled listener will have available the distribution of information found in Spanish input to children of the ages tested. The modeled listener will then use Bayesian inference to combine the information together, in order to interpret subject pronouns in a specific experimental context; that is, the modeled listener will choose whether the pronoun refers to the subject antecedent or the non-subject antecedent for a particular experimental item.

When making this inference, the modeled listener will (i) use either accurate or inaccurate representations of the available probability information, and (ii) either accurately or inaccurately deploy those probability representations. In particular, we compare four versions of this modeled listener: (i) a **baseline** model with accurate probability representations and accurate deployment, (ii) an **inaccurate representation** model with inaccurate probability representations but accurate deployment, (iii) an **inaccurate deployment** model with accurate probability representations but inaccurate deployment, and (iv) a model with **both inaccurate** probability representations and inaccurate deployment.[8]

Below we first describe how Bayesian inference operates to determine a pronoun's interpretation in context, on the basis of the available information. We then describe the corpus sample used to estimate how that information is distributed in Spanish input, which the modeled listener will use to represent relevant information. We then discuss the model implementations for the baseline, inaccurate representation, inaccurate deployment, and both-inaccurate modeled listeners. We then assess how well each model version is able to capture the observed pronoun interpretation behavior for children and adults.

3.1 Bayesian inference for pronoun interpretation

A Bayesian modeled listener calculates the probability of a hypothesis $h \in H$, given some data D (this is the posterior probability $p(h|D)$), as shown in (6). It does this calculation on the basis of both its prior beliefs about that hypothesis (the prior $p(h)$) and how well that hypothesis accounts for the data (the likelihood $p(D|h)$).

(6) $$p(h|D) \propto p(h) \cdot p(D|h)$$

For pronoun interpretation, the hypotheses might be that the pronoun refers to each of the available antecedents α from the preceding clause. For example, if the preceding clause contains a singular subject and a plural object, then there are two hypotheses to consider ($H = \{\alpha_{subj.\text{SG}}$, $\alpha_{\neg subj.\text{PL}}\}$. So, a posterior probability can be calculated for each hypothesis and used to estimate how the Bayesian listener would respond in a particular experimental context. In the example above, if $p(\alpha_{subj.\text{SG}}|D)$=0.6 (and so $p(\alpha_{\neg subj.\text{PL}}|D)$=0.4), this Bayesian listener would prefer the pronoun to have the singular subject antecedent rather than the plural object antecedent. In addition, we might expect a group of these Bayesian listeners to select the subject antecedent 60%

[8]We note that there are of course many other possible models, including different implementations of inaccurate representations and inaccurate deployment. We leave these other possibilities as interesting avenues of future work.

of the time and the non-subject antecedent 40% of the time. In this way, we can map the posterior probabilities of a Bayesian listener to the group-level results from the pronoun interpretation experiment.

We can then use this framework to calculate the posterior probabilities for a particular experimental context. In this context, the "data" D correspond to the pronoun information in that context (i.e., pronoun form FORM$\in\{$ø, overt$\}$, discourse connective CON$\in\{$*y después*, *porque*$\}$, and agreement morphology MOR$\in\{$SG, PL$\}$). So, we can calculate the posterior $p(\alpha_{subj.\text{SG}}|\ \text{FORM}, \text{CON}, \text{MOR})$, as in (7). If we assume each cue is independent[9], then the likelihood calculation can be further divided as in (8).

(7) $$p(\alpha_{subj.\text{SG}}|\text{FORM}, \text{CON}, \text{MOR}) \propto p(\alpha_{subj.\text{SG}}) \cdot p(\text{FORM}, \text{CON}, \text{MOR}|\alpha_{subj.\text{SG}})$$

(8) $$\begin{aligned} p(\alpha_{subj.\text{SG}}|\text{FORM}, \text{CON}, \text{MOR}) \propto p(\alpha_{subj.\text{SG}}) & \cdot p(\text{FORM}|\alpha_{subj.\text{SG}}) \\ & \cdot p(\text{CON}|\alpha_{subj.\text{SG}}) \\ & \cdot p(\text{MOR}|\alpha_{subj.\text{SG}}) \end{aligned}$$

In a specific experimental context, the pronoun probability information may or may not favor the singular subject antecedent – this is reflected in the likelihoods $p(\text{FORM}|\alpha_{subj.\text{SG}})$, $p(\text{CON}|\alpha_{subj.\text{SG}})$, and $p(\text{MOR}|\alpha_{subj.\text{SG}})$. For instance, suppose we have an experimental context like (4), with a singular subject antecedent (*la maestra*) and a plural object antecedent (*las niñas*). Suppose then that the speaker uses the cues in (4a): a null pronoun (FORM=ø), the discourse connective *y después* (CON=*y después*), and singular agreement morphology (MOR=SG). The probability information in the likelihoods then reflects how much each of these values is associated with each antecedent. For example, our corpus analysis of child-directed speech (presented in more detail in section 3.2 and summarized in Table 3) suggests the following values: $p($FORM=ø$|\alpha_{subj.\text{SG}})$=0.938, $p($CON=*y después*$|\alpha_{subj.\text{SG}})$=0.324, and $p($MOR=SG$|\alpha_{subj.\text{SG}})$=0.998.

To calculate the probability that the pronoun refers to the subject versus the non-subject antecedent in this specific experimental context, we use (8) to calculate the posteriors: $p(\alpha_{subj.\text{SG}}|\text{FORM}$, CON, MOR$)$ for the singular subject antecedent and $p(\alpha_{obj.\text{PL}}|\text{FORM}$, CON, MOR$)$ for the plural non-subject antecedent. Let's again consider the experimental context of (4a). To calculate the posterior probability $p(\alpha_{subj.\text{SG}}|$FORM=ø, CON=*y después*, MOR=SG$)$, we also need to know the prior probability $p(\alpha_{subj.\text{SG}})$ of a pronoun referring to a singular subject antecedent like *la maestra*. Our corpus analysis of child-directed speech, suggests this is 0.362 (i.e., this kind of antecedent tends to occur about a third of the time). If we use (8) (repeated below as (9)), we can see the resulting posterior probability, which is fairly low. However, it's important to compare this probability to the posterior for the other antecedent, which in (4a) was the plural object *las niñas*. Our corpus analysis of child-directed speech suggests the prior is $p(\alpha_{\neg subj.\text{PL}})$=0.129, and the likelihoods are $p($FORM=ø$|\alpha_{\neg subj.\text{PL}})$=0.959, $p($CON=*y después*$|\alpha_{\neg subj.\text{PL}})$=0.394, and $p($MOR=SG$|\alpha_{\neg subj.\text{PL}})$=0.005. We then arrive at the posterior in (10), which is even lower.

[9]By assuming feature independence, this Bayesian approach reduces to a Naive Bayes classifier (Jurafsky, 2020). We discuss the implications of this idealizing assumption in section 4.2.

$$p(\alpha_{subj.\text{SG}}|\text{FORM},\text{CON},\text{MOR}) \propto p(\alpha_{subj.\text{SG}}) \cdot p(\text{FORM}|\alpha_{subj.\text{SG}}) \cdot p(\text{CON}|\alpha_{subj.\text{SG}}) \cdot p(\text{MOR}|\alpha_{subj.\text{SG}})$$
$$\propto 0.362 \cdot 0.938 \cdot 0.324 \cdot 0.998 \qquad (9)$$
$$\propto 0.110$$

(10)
$$p(\alpha_{\neg subj.\text{PL}}|\text{FORM},\text{CON},\text{MOR}) \propto p(\alpha_{\neg subj.\text{PL}}) \cdot p(\text{FORM}|\alpha_{\neg subj.\text{PL}}) \cdot p(\text{CON}|\alpha_{\neg subj.\text{PL}}) \cdot p(\text{MOR}|\alpha_{\neg subj.\text{PL}})$$
$$\propto 0.129 \cdot 0.959 \cdot 0.394 \cdot 0.005$$
$$\propto 0.000244$$

Comparing the posteriors for the two antecedent options (i.e., normalizing the probabilities so they sum to 1), we can see that this cue combination would cause this Bayesian listener to significantly favor the subject antecedent in this experimental context ($\frac{0.110}{0.110+0.000244} \approx 0.998$). This Bayesian implementation serves as our baseline modeled listener. It assumes accurate representation of the probability information about the pronoun form, connective, and agreement morphology, as based on the likelihoods[10], as well as accurate representation of antecedent probability information, based on the prior for that antecedent type; this implementation also assumes accurate deployment of those probability representations, as these priors and likelihoods all contribute to the posterior calculation.

As we can see from this example, this modeled Bayesian listener requires estimates of relevant priors and likelihoods for a given experimental context. That is, the input to the modeled listener takes the form of the relevant priors and likelihoods used to calculate the posteriors. We turn next to how we estimate the priors and likelihoods from naturalistic child-directed speech corpora. In this way, our modeled Bayesian listener will have as input the same input that young Spanish-learning children encounter.[11]

3.2 Modeled listener input

To estimate the input probabilities used for the priors and likelihoods in our modeled Bayesian listeners, we relied on samples of child-directed speech from the Schmitt–Miller corpus (Miller & Schmitt, 2012). These data were taken from spontaneous interactions between caregivers and their children (ages 1;6-5;11) born and raised in Mexico City, Mexico. Caregivers were recorded during two to four free-play sessions with their children, lasting around 30 minutes each, as well as one approximately 30-minute session chatting with another adult. These probabilities are shown in Table 3, separated out by the type of antecedent (i.e., whether the antecedent was the subject or not, and whether the antecedent had singular agreement morphology or plural agreement morphology). See Appendix E for details about the corpus, including raw counts for each case in Table 3.

We can see from these probabilities that child-directed speech favors singular antecedents, wherever they occur (i.e., priors for singular antecedents = 0.362 and 0.438 vs. plural antecedents = 0.071 and 0.129). The likelihoods demonstrate that the null form of the pronoun is generally favored, irrespective of the antecedent type (the likelihoods for the ø form = 0.817-0.984). Agreement morphology nearly categorically favors the antecedent type with matching morphology: the likelihood for singular morphology given a singular antecedent is 0.998, and for plural morphology given a plural antecedent is 0.995 (irrespective of antecedent position). In contrast, the likelihood of a particular discourse connective seems to sometimes depend on the antecedent type: plural subject antecedents favor *y después* (0.750), while the other antecedent types favor *porque* (0.606-0.868).

[10] As noted before, it additionally assumes independence of cues, when representing this probability information.

[11] We note that this input may not be as realistic for adults, who have access to other kinds of speech besides the child-directed speech that they themselves produce. We discuss this limitation in section 4.2.

antecedent type		prior $p(\alpha)$	likelihoods $p(\text{FORM}\|\alpha)$ ø	overt	$p(\text{CON}\|\alpha)$ *y después*	*porque*	$p(\text{MOR}\|\alpha)$ SG	PL
SUBJ	SG	0.362	0.938	0.062	0.324	0.676	0.998	0.002
	PL	0.071	0.984	0.016	0.750	0.250	0.005	0.995
¬SUBJ	SG	0.438	0.817	0.183	0.132	0.868	0.998	0.002
	PL	0.129	0.959	0.041	0.394	0.606	0.005	0.995

Table 3: Priors and likelihoods for different types of antecedents α (subject (SUBJ) or non-subject (¬SUBJ), singular (SG) or plural (PL)) estimated from the prononun information distributions in naturalistic child-directed speech. Each row corresponds to one antecedent type (e.g., the first row corresponds to α=SUBJ,SG).

3.3 Bayesian listener implementations

We first describe the baseline modeled Bayesian listener more fully, and then the implementations of modeled listeners with either inaccurate probability representations only, inaccurate deployment only, or both inaccurate probability representations and inaccurate deployment. In particular, we follow the approach of Gagliardi *et al.* (2017) by modeling inaccuracy as noise: there's either noise in the modeled listener's representation of the probability information available (about cues or antecedents), or noise in the modeled listener's ability to reliably use that probability information in novel situations, such as an experimental task. So, the baseline Bayesian listener is adapted to incorporate noise in the probability representations (inaccurate representations), noise in the integration of information (inaccurate deployment), or both.[12]

3.3.1 Baseline Bayesian listener incorporating antecedent type

As described above, our baseline Bayesian listener has accurate representations of both antecedent probability information and pronoun probability information (pronoun form, discourse connective, and agreement morphology) on the basis of its input. The modeled listener is also able to integrate this probability information accurately, using Bayesian inference.

(11)
$$\begin{aligned} p(\alpha_{num}, \alpha_{subj?}|\text{FORM, CON, MOR}) \propto\ & p(\alpha_{num}, \alpha_{subj?}) \\ & \cdot p(\text{FORM}|\alpha_{num}, \alpha_{subj?}) \\ & \cdot p(\text{CON}|\alpha_{num}, \alpha_{subj?}) \\ & \cdot p(\text{MOR}|\alpha_{num}, \alpha_{subj?}) \end{aligned}$$

Because priors and likelihoods vary depending on the number and position of the antecedent under consideration, (11) explicitly indicates that the precise value of the prior and each likelihood is different for different antecedent types. That is, the number of the antecedent is either singular or plural ($\alpha_{num}\in\{\text{SG, PL}\}$), and the position of the antecedent is either the subject position or some non-subject position, like the object ($\alpha_{subj?}\in\{\text{SUBJ}, \neg\text{SUBJ}\}$). For instance, a singular subject antecedent ($\alpha_{subj.\text{SG}}$) would have α_{num}=SG and $\alpha_{subj?}$=SUBJ. A plural object antecedent ($\alpha_{\neg subj.\text{PL}}$) would have α_{num}=PL and $\alpha_{subj?}$=¬SUBJ.

Likewise, both the priors and the likelihoods depend on the values of these two variables; this accords with the probabilities from our corpus analysis in Table 3, which also varied by these properties of the

[12] Python code implementation of all modeled learners described here is available at *[link to author github repository].*

antecedent. The posterior probability of a specific antecedent in a given experimental context can now be calculated by using the appropriate priors and likelihoods from Table 3. So, for example, to calculate the posterior for a potential singular antecedent in the subject position, the probabilities from the first row of Table 2 can be used (SUBJ, SG).

3.3.2 Bayesian listener with inaccurate representations

Our listener implementation with inaccurate representations involves the listener calculating posterior probabilities the same way, but relying on inaccurate probability representations to do so. These probability representations correspond to the modeled listener's priors and likelihoods: priors reflect the baseline probability of antecedents with different properties (e.g., singular antecedents in subject position); likelihoods reflect the probability of a particular pronoun cue value (e.g., a null form or singular morphology), given an antecedent with certain properties (e.g., a singular subject antecedent).

Inaccurate representations by incorporating noise. We use the same approach for implementing inaccurate representations of the probabilities corresponding to both the priors and likelihoods, which involves incorporating noise into these probability representations. The modeled listener uses the softmax function ($e^{\sigma ln(probability)}$=$probability^{\sigma}$), where σ serves as a "contrast" parameter. When σ=1, probabilities are kept the same (i.e., equal to the values in Table 3); when $\sigma<1$, probabilities are made more uniform (i.e., relative contrasts are decreased, and so smoothed away); when $\sigma>1$, probabilities are made more extreme (i.e., relative contrasts are increased, and so sharpened). Thus, our modeled learner can decrease the contrast between relative probabilities (e.g., using σ=0.5 to smooth 0.324 vs. 0.676 into 0.409 vs. 0.591); it can also increase the contrast between relative probabilities (e.g., using σ=2 to sharpen 0.324 vs. .676 into .187 vs. 0.813).

We note that the softmax function is a standard component of models of human decision-making tasks, including language tasks (e.g., Frank & Goodman (2012); Goodman & Stuhlmüller (2013); Scontras & Goodman (2017)). Here, we investigate contrast parameter values $0.0\leq\sigma\leq4.0$.[13]

Inaccurate priors. The prior has a single contrast parameter value (σ_α), which holds for all antecedents (i.e., σ_α is the same for all α_{num} and $\alpha_{subj?}$ values). This represents the same noise level for the prior on potential antecedents, irrespective of the number of the potential antecedent or its position.[14]

Inaccurate likelihoods. Each pronoun cue has its own contrast parameter value (i.e., σ_{form}, σ_{con}, σ_{mor}), representing (potentially) different noise levels associated with each cue representation.

Implementing inaccurate representations. Equation (12) shows the way the modeled listener implements potentially inaccurate priors and likelihoods, using the different σ contrast parameters.

[13]Note that 0.0 corresponds to completely smoothing the relative contrasts (i.e., smoothing 0.324 vs. 0.676 to 0.500 vs. 0.500), and so is a natural lower bound. The upper bound of 4.0 is arbitrary, but does represent an extreme sharpening that verges on categorical use (i.e., sharpening 0.324 vs. 0.676 to 0.050 vs. 0.950). We did explore values up to 100.0, with no qualitative change in the results – for those instances where 4.0 was selected as the best-fitting value in the current implementation, the highest value explored was selected as best-fitting in the exploratory implementations.

[14]We note that this assumption of the same contrast value for all antecedent types could be relaxed in future work investigating different modeled listener implementations. We discuss this possibility more in section 4.2.

$$p_\sigma(\alpha_{num}, \alpha_{subj?}|\text{FORM}, \text{CON}, \text{MOR}) \propto p(\alpha_{num}, \alpha_{subj?})^{\sigma_\alpha} \cdot p(\text{FORM}|\alpha_{num}, \alpha_{subj?})^{\sigma_{form}} \cdot p(\text{CON}|\alpha_{num}, \alpha_{subj?})^{\sigma_{con}} \cdot p(\text{MOR}|\alpha_{num}, \alpha_{subj?})^{\sigma_{mor}} \tag{12}$$

We can then determine the contrast parameter values that allow the modeled Bayesian listener to best match the child and adult behavior from our experiment. By doing so, we can get a sense of if and how the true probability representations would need to be distorted in order for this modeled listener to reproduce the behavior we observe. We note that if the best-fitting value is σ=1 for any information type, then this implies that the accurate representation for that probability information (prior or likelihood) was the best fit, rather than a distorted representation. So, this implementation of the inaccurate representations listener allows for the possibility of accurate probability representations. More generally, we can also determine if a better fit occurs when relative probability differences are decreased (σ<1), increased (σ>1), or left undistorted (σ=1). This would then correspond to whether listeners smooth, sharpen, or transparently use the relative probability differences from their input.

3.3.3 Bayesian listener with inaccurate deployment

Inaccurate deployment by selective dropout. We use the same approach for implementing inaccurate deployment of the represented probabilities for both the priors and likelihoods, which involves selectively dropping out the information from these probability representations. More specifically, our implementation of a listener with inaccurate deployment involves the listener having accurate probability representations, but not always using them. That is, when the listener calculates the posterior probabilities, the listener has available accurate representations of the probabilities for the priors and likelihoods, but simply "misses" the information from one or more of these probability representations in the moment of a particular experimental item.

Inaccurate deployment of priors. When the listener calculates the posterior probabilities, the listener has available an accurate probability representation of possible antecedents with the corresponding number and position information (i.e., the prior). However, as mentioned above, the listener simply "misses" the probability information from the prior in the moment of a particular experimental item.

For example, a listener with inaccurate deployment like this might have an accurate representation of how probable it is to have a singular antecedent in subject position (i.e., an accurate prior $p(\alpha_{num}$=SG, $\alpha_{subj?}$=SUBJ) = 0.362). Yet, when calculating the posterior, this probability representation is unavailable; in this case, the listener defaults to an uninformative prior, which is a uniform distribution over the possible antecedents (i.e., if there are two possible antecedents, then $p(\alpha_{num}, \alpha_{subj?})$=0.5). So, the posterior calculation proceeds without that informative prior (which we indicate with p_{UNIF}). An example is shown in (13) when the listener uses all three cues of pronoun form, discourse connective, and agreement morphology, but doesn't use the prior; $p(\text{UNIF})$ represents a uniform (i.e., uninformative) distribution over possible antecedents.

$$p_{\text{UNIF}}(\alpha_{num}, \alpha_{subj?}|\text{FORM}, \text{CON}, \text{MOR}) \propto p(\text{UNIF}) \cdot p(\text{FORM}|\alpha_{num}, \alpha_{subj?}) \cdot p(\text{CON}|\alpha_{num}, \alpha_{subj?}) \cdot p(\text{MOR}|\alpha_{num}, \alpha_{subj?}) \tag{13}$$

In this way, the accurate prior information drops out and is effectively ignored by the listener for that particular experimental item–even though the modeled listener has an accurate representation of this prior information (and may even use it for other experimental items). In this way, inaccurate deployment of the prior is implemented via a selective dropout of the information from the prior. This means the listener's posterior probability relies solely on the probability information coming from the pronoun form, the discourse connective, and the agreement morphology, represented by the likelihoods.

Inaccurate deployment of likelihoods. When the listener calculates the posterior probabilities, the listener has available accurate probability representations (i.e., likelihoods) for the different cues. However, as before, the listener simply "misses" the probability information from one or more cues in the moment of a particular experimental item.

For example, a listener with inaccurate deployment like this might have an accurate probability representation of the agreement morphology cue (i.e., an accurate likelihood $p(\text{MOR}|\alpha_{num}, \alpha_{subj?})$), but be unable to use it when calculating the posterior. That is, the posterior calculation proceeds without that cue, as in (14), where only the probability information from the pronoun form and the discourse connective are used, along with the probability information from the prior.

(14)
$$\begin{aligned} p(\alpha_{num}, \alpha_{subj?}|\text{FORM}, \text{CON}) \propto\ & p(\alpha_{num}, \alpha_{subj?}) \\ & \cdot p(\text{FORM}|\alpha_{num}, \alpha_{subj?}) \\ & \cdot p(\text{CON}|\alpha_{num}, \alpha_{subj?}) \end{aligned}$$

In this way, the accurate probability information from the morphology cue drops out and is effectively ignored by the listener for that particular experimental item–again, even though the modeled listener has an accurate representation of this cue's probability information (and may even use it for other experimental items). In this way, inaccurate deployment is again implemented via a selective dropout of probability information, this time from one or more cues.

Implementing selective deployment. We implement the frequency of this dropout process with a "use" parameter β that encodes how often information for the prior or a particular cue's likelihood is used. That is, with probability β_α, the informative prior for that antecedent will be used in the posterior calculation; with probability (1-β_α), it will be ignored and an uninformative prior used instead. Similarly, with probability β_{cue}, a particular cue will be used in the posterior calculation; with probability (1-β_{cue}), it will be ignored. Each cue therefore has its own β (β_{form}, β_{con}, β_{mor}).[15] Each β ranges between 0 and 1 ($0.0 \leq \beta \leq 1.0$).

Because the prior could be used (β_α) or ignored (1-β_α), this yields two possibilities for the prior for any particular experimental item. Similarly, because each cue could be used or ignored individually, there are eight possibilities for the cues for any particular experimental item: the pronoun form's probability is used (β_{form}) or ignored (1-β_{form}); the discourse connective's probability is used (β_{con}) or ignored (1-β_{con}); the agreement morphology's probability is used (β_{mor}) or ignored (1-β_{mor}). So, the posterior for a particular experimental item ($p_\beta(\alpha|\text{FORM,CON,MOR},\alpha_{num}, \alpha_{subj?})$) is a mix of these 2 x 8 possibilities.[16] We note that if all three cues are ignored, the modeled listener uses the prior over potential antecedents alone–whether informative ($p(\alpha_{num}, \alpha_{subj?})$) or uninformative ($p(\text{UNIF})$)–to calculate the posterior.

We can then determine the use parameter β values that allow the modeled Bayesian listener to best match the child and adult behavioral data on pronoun interpretation in our experiment. By doing so, we can

[15] As with the contrast parameter σ, future work can investigate modeled listener implementations with different β_{cue} values for different antecedent types.

[16] See Appendix F for the full specification of how this modeled listener calculates the posterior probability.

get a sense of how noisy this integration process would need to be (and in what particular ways the process is noisy) in order to best account for participant behavior, under this inaccurate deployment model. We note that if the best-fitting value is β=1 for the prior or any cue's likelihood, then this signals that accurate integration for that information (prior or likelihood) was the best fit, rather than noisy integration. That is, as with the inaccurate representation listener, the inaccurate deployment listener allows for the possibility of accuracy (in this case, for always deploying available information). More generally, we can also determine the best-fitting relative "use rate" of the available information for both children and adults.

3.3.4 Bayesian listener with both inaccurate representations and inaccurate deployment

Our implementation of a listener with both inaccurate probability representations and inaccurate deployment of those representations combines the implementations of the listeners with only inaccurate probability representations and the listeners with only inaccurate deployment. More specifically, the softmax function with contrast parameter σ is used for the representations of the prior and likelihood probabilities, as shown in p_σ in (12), using σ_α, σ_{form}, σ_{con}, and σ_{mor}; the selective dropout function with use parameter β is used for the deployment of the prior and likelihood probabilities, as shown in p_β in (17) in Appendix F, using β_α, β_{form}, β_{con}, and β_{mor}. More concretely, the both-inaccurate listener calculates the posterior $p_{\sigma,\beta}$, which uses the inaccurate representation posterior calculation p_σ within the mixture model defined for the inaccurate deployment posterior calculation p_β. So, $p_{\sigma,\beta}$ looks identical to p_β, except that $p_\sigma(\alpha|\text{cues}, \alpha_{num},\alpha_{subj?})$ is used in place of $p(\alpha|\text{cues}, \alpha_{num},\alpha_{subj?})$ in (17) in Appendix F; similarly, $p_{\text{UNIF},\sigma}(\alpha_{num},\alpha_{subj?}|\text{cues})$ is used in place of $p_{\text{UNIF}}(\alpha_{num},\alpha_{subj?}|\text{cues})$, with (15) showing this implementation when all three cues are used.

(15)
$$\begin{aligned} p_{\text{UNIF},\sigma}(\alpha_{num}, \alpha_{subj?}|\text{FORM, CON, MOR}) \propto\; & p(\text{UNIF}) \\ & * p(\text{FORM}|\alpha_{num}, \alpha_{subj?})^{\sigma_{form}} \\ & * p(\text{CON}|\alpha_{num}, \alpha_{subj?})^{\sigma_{con}} \\ & * p(\text{MOR}|\alpha_{num}, \alpha_{subj?})^{\sigma_{mor}} \end{aligned}$$

3.3.5 Comparing Bayesian listeners

To sum up, we defined two ways that modeled Bayesian listeners can be inaccurate when it comes to the posterior calculation they use to decide how a pronoun is interpreted in a particular context. First, listeners can have inaccurate representations of the probabilities for the priors and likelihoods that go into the posterior calculation. Second, listeners can inaccurately deploy their represented probabilities when doing that posterior calculation, specifically by only using some (or none) of the probabilities. These two options lead to four modeled Bayesian listener types: (i) the baseline listener with accurate representations of probability information about potential antecedents and cues, who deploys those representations accurately, (ii) the listener with inaccurate representations of probability information about potential antecedents and cues, who deploys those probability representations accurately, (iii) the listener with inaccurate deployment of accurate probability representations, and (iv) the listener with both inaccurate probability representations and inaccurate deployment.

We can now compare these modeled listeners on their ability to best capture the child and adult data from the pronoun interpretation task (in particular the responses summarized in Figure 2 that correspond to the subject and non-subject responses provided by participants in each condition). More specifically, a modeled listener can generate the probability of a data point being produced (e.g., a subject or non-subject response in a particular experimental context), given that listener's implemented hypotheses about how judgments are generated in context; here, the context is a certain combination of pronoun form, discourse connective, and agreement morphology. The probabilities of all participant responses can then be aggregated together, representing the probability that the modeled listener would generate the entire set of responses. This will allow us to evaluate which hypothesis, as implemented by the different modeled listeners, provides the best explanation for children's observed behavior and how this compares to the best explanation for adults' observed behavior.

We can also look within the best-fitting modeled listener to find the parameter values that yield the best fit. This will give us an estimate of exactly how inaccurate participants' probability representations are (i.e., how much σ differs from 1) and/or how inaccurate their deployment is (i.e., how much β differs from 1), and whether this changes as children become adults.

3.4 Results

3.4.1 Which modeled listener?

To fit each modeled listener type, we varied the value of the contrast parameters ($0.0 \leq \sigma \leq 4.0$) and/or the use parameters ($0.0 \leq \beta \leq 1.0$) in increments of 0.01, and chose the combination with the best log probability of the participant responses (i.e., log likelihood score[17]) in the pronoun interpretation experiment.

To determine which of the fitted modeled listener types best matches participant pronoun interpretation behavior, we used the Bayesian information criterion (**BIC**) (Schwarz, 1978), shown in (16); the BIC balances a model's ability to fit the data against the number of parameters m that it uses. Since models with more parameters have a natural advantage, the BIC only rewards models with more parameters if they provide a substantially better fit to the data than models with fewer parameters. In our case, this means that to have a better BIC score, the both-inaccurate listener (with 4 σ parameters and 4 β parameters, for a total of 8 parameters) has to have a substantially better fit than the inaccurate representation listener (with 4 σ parameters) and the inaccurate deployment listener (with 4 β parameters); similarly, all three inaccurate listeners have to have a substantially better fit than the baseline listener (with 0 parameters) to have better BIC scores.

(16) $$BIC = m \cdot log(|data|) - 2 \cdot log(modelfit)$$

BIC evaluates a model's fit to the data using the log likelihood of the data, and the best-fitting parameter values for that model.[18] The data in this case are the number of antecedent responses provided by participants in each condition (e.g., adults provided 736 responses) ; so, we calculate P(all participant responses across all conditions | best-fitting parameter values) for each modeled listener type, and take the log – this is the log likelihood score. A better BIC score, which incorporates the log likelihood score, is closer to 0. BIC scores for all modeled listener types are shown in Table 4 for children ≥5 years old and adults.

We see that the baseline modeled listener, with both accurate probability representations and accurate deployment, fares the worst (children ≥5 BIC: 1931.33 vs. <607.12; adults BIC: 1323.82 vs. <668.58); this result suggests that both children and adults are being inaccurate somewhere. That is, they're using inaccurate probability representations, inaccurately deploying accurate probability representations, or both using inaccurate probability representations and deploying those representations inaccurately. Importantly for the target state of acquisition, adult behavior isn't best captured by a totally accurate listener. That is, becoming adult-like in pronoun interpretation doesn't mean that children should represent and deploy

[17] A better log likelihood score is closer to 0. We can see this demonstrated by comparing e^{-3} (≈0.05) vs. e^{-6} (≈0.002). The $\ln(e^{-3})$=-3, while the $\ln(e^{-6})$=-6. 0.05>0.002, and -3 is closer than -6 to 0.

[18] Appendix G has the full set of best-fitting parameter values for each modeled listener type. Below, we discuss only the parameter values for the modeled listeners with the best BIC scores.

age	baseline	inaccurate representation $\sigma \neq 1$	inaccurate deployment $\beta \neq 1$	both inaccurate $\sigma \neq 1$, $\beta \neq 1$
≥ 5	1931.33	**590.04**	**590.70**	607.12
adults	1323.82	**642.48**	646.96	668.58

Table 4: BIC scores for modeled learners with accurate representation and deployment (baseline), potentially inaccurate representations only ($\sigma \neq 1$), potentially inaccurate deployment only ($\beta \neq 1$), or both ($\sigma \neq 1$, $\beta \neq 1$). The best-performing modeled listener's BIC score(s) (closest to 0) for each age group is/are **bolded**. Scores within 2.0 of each other are considered equivalent, following Kass & Raferty (1995).

available pronoun probability information with total accuracy. Rather, children have to become more adult-like in how they inaccurately do these things.

We can see in Table 4 that children's behavior is best captured by the modeled listener types that use either inaccurate probability representations (BIC: 590.04) or inaccurate deployment (BIC: 590.70), but not both (BIC: 607.12). So, these results suggest that children either (i) have inaccurate probability representations, but deploy them accurately, or (ii) have accurate probability representations, but deploy them inaccurately. Table 4 also suggests that adult behavior is best captured by the modeled listener type using inaccurate probability representations, but deploying those representations accurately (BIC: 642.48 vs. 646.96 and 668.58).[19]

So, it could be that children are basically doing the same thing as adults: accurately deploying inaccurate probability representations. In this case, children wouldn't need to qualitatively shift their approach to interpreting pronouns in order to become adult-like. However, our results are also compatible with children doing something fundamentally different from adults, by inaccurately deploying accurate probability representations. In this case, children would need to qualitatively shift their pronoun interpretation approach as part of development; more specifically, children would need to learn how to (perhaps strategically) distort their probability representations of available pronoun information, but accurately deploy those distorted representations in the moment.

3.4.2 What's changing?

To better understand the potential change between children and adults, we can examine the best-fitting model parameters that yielded the BIC scores in Table 4. In particular, we can look at the best-fitting σ contrast values for the inaccurate representations listener to understand what children and adults may be doing, as well as the best-fitting β use values for the inaccurate deployment listener to understand what else children may be doing. These parameter values are shown in Table 5.[20]

Inaccurate representations in adults and possibly in children. Adult behavior is best captured by a modeled listener with inaccurate probability representations, and this is one of the two modeled listener

[19] We do note that the robustness of these results isn't clear. Given the length of the model runtime, it was impractical to conduct a kind of sensitivity analysis based on hundreds of dataset perturbations. With more efficient model code, future work may be able to investigate how sensitive this pattern is to the specific distributions of the participant dataset we have. For now, we discuss the implications if the qualitative pattern (i.e., which modeled listeners have the best BIC score) does indeed hold.

[20] The best-fitting parameter values for all modeled listener types are in Table A1 in Appendix G.

	σ_{for}	σ_{con}	σ_{mor}	σ_{α}	β_{for}	β_{con}	β_{mor}	β_{α}
	children: inaccurate representations							
$\geq$5	0.02	0.28	0.11	0.00	1	1	1	1
	children: inaccurate deployment							
$\geq$5	1	1	1	1	0.00	0.43	0.30	0.00
	adults: inaccurate representations							
adults	0.25	0.33	0.28	0.00	1	1	1	1

Table 5: Best-fitting parameter values for the modeled listeners that best capture the pronoun interpretation behavior of children and adults, based on BIC scores. σ values are used for inaccurate probability representations, while β values are used for inaccurate deployment of representations. Accurate deployment in the inaccurate representations learner uses β values of 1; accurate representations in the inaccurate deployment learner uses σ values of 1.

types that children's behavior is also best captured by. In Table 5, we see that the way in which these probability representations are inaccurate is qualitatively the same between adults and children. In particular, the best-fitting σ values for both adults and children are all less than 1 (adults: 0.00-0.33, children: 0.00-0.28). These values mean that both adults and children tend to smooth away any relative probability differences that the pronoun information provides – and in some cases, to do so quite dramatically. For instance, a σ value of 0.33 (the highest among the adults' σ values) would take a probability distribution of 0.75 vs. 0.25, and transform it into 0.59 vs. 0.41. Lower σ values have even stronger smoothing effects: a σ value of 0.28 (the highest among the children's σ values) transforms 0.75 vs. 0.25 into 0.58 vs. 0.42; σ=0.11 transforms 0.75 vs. 0.25 into 0.53 vs. 0.47; σ=0.02 transforms 0.75 vs. 0.25 into 0.505 vs. 0.495 (i.e., nearly a uniform distribution). So, these results suggest that even the least-distorted representation information for both adults and children has smoothed away a lot of the probability contrasts from the input.

Interestingly, for some information, any probability contrasts present in the input are completely smoothed away, with σ=0.00 (σ_{α} for adults and all children). This can be interpreted as participants completely ignoring this probability information, as a σ of 0.00 yields a uniform distribution. For both adults and children, the information about the prior is completely smoothed away, yielding a distribution over potential antecedents that's uniform. So, children wouldn't need to change at all to become adult-like in this respect: that is, it's adult-like to ignore how often a pronoun refers to a potential antecedent in general, irrespective of the cues available in the current context.

However, for the cue of pronoun form (represented by σ_{for}), children would need to smooth away less of the probability contrast available from the input (adults: 0.25; children: 0.02). Similarly, children would need to smooth away less of the probability contrasts available for the information about discourse connectives and agreement morphology (adult σ_{con}: 0.33; child σ_{con}: 0.28; adult σ_{mor}: 0.28, child σ_{mor}: 0.11). More generally, under this view, acquisition involves tuning how much probability information is smoothed away, lessening the amount smoothed away to adult-like levels.

Inaccurate deployment possibly in children. Another possibility for children's observable pronoun interpretation behavior is that they have accurate representations of pronoun probability information, but are inaccurately deploying them. In particular, children would completely ignore information about the prior (i.e., children use only probability information from the current pronoun context) and completely ignore probability information about the pronoun form, as β_{α}=β_{for}=0.00. The other two pronoun information types of discourse connectives and agreement morphology are ignored less (β_{for}=0.43, β_{mor}=0.30). In contrast, adults would be deploying their representations accurately (all βs=1), rather than selectively deploying them.

So, to become adult-like in deployment, children would need to make a qualitative shift in how they interpret pronouns by accurately, rather than inaccurately, deploying those probability representations. We do note that adult-like smoothing of probability representations involves completely smoothing away information about the prior distribution over antecedents (σ_{alpha}=0.00); this total smoothing is behaviorally equivalent to never using that probability information in the moment (β_{α}=0.00), because it means children completely ignore the probability information from the input. So, for probability information about prior antecedents, children wouldn't necessarily need to qualitatively change what they end up doing, which is to completely ignore that probability information.

Becoming adult-like: Summary. Taken together, our modeling results suggest one of two options for children: they're either always using highly-smoothed representations of the pronoun probability information available (inaccurate representations) or they're selectively using accurate representations of the pronoun probability information available (inaccurate deployment). Notably, the best-fitting inaccurate representation listener for children completely smoothed away all information for the prior distribution over antecedents (σ_{α}=0.00) and nearly always did so for the pronoun form (σ_{for}=0.02); smoothing away all information in a listener relying on inaccurate probability representations yields an equivalent result to the best-fitting inaccurate deployment listener that never deploys these information types (β_{α}=β_{for}=0.00). That is, children's behavior is best captured by ignoring information about the prior antecedent distribution and the pronoun form (either by smoothing away all probability contrasts available in the input, or never deploying that available information in the moment).

In contrast, adults are best fit by an inaccurate representation listener that ignores only the information about the prior antecedent distribution (σ_{α}=0.00), and thus completely smooths away any probability contrasts available from the input. So, to become adult-like, children need to pay attention to pronoun form–whether by smoothing an inaccurate probability representation less (to achieve adult-like σ_{for}=0.25), or making a qualitative shift to always deploy (i.e., β_{for}=1) a probability representation that's inaccurate from being smoothed (i.e., σ_{for}=0.25). Children also need to pay attention to information about the discourse connective and agreement morphology more than they do, again by either smoothing these probability representations less or by making a qualitative shift to always using a smoothed probability representation.

3.4.3 Comparison to behavioral findings

Recall that we observed some specific behaviors in children and adults (discussed in section 2.4). In particular, the logistic regression analyses suggested that both discourse connectives and agreement morphology mattered for children age five and older, and all three cues (pronoun form, discourse connectives, and agreement morphology) mattered for adults. Our modeling results align with these findings, additionally uncovering more specifically how these information types may matter for each group.

In particular, if we look at the modeling results for children, the best-fitting modeled listeners were either (i) relying on inaccurate probability representations that were smoothed, or (ii) inaccurately deploying accurate probability representations. Either option results in children ignoring specific types of probability information, in line with the logistic regression analysis from before. For children age five and older, both the discourse connective probabilities and the agreement morphology probabilities were either less smoothed (σ=0.11-0.28 vs. 0.02) or deployed more often (β=0.30-0.43 vs. 0.00). So, the information types whose representations were either less-smoothed (and so more adult-like) or deployed more often (and so more adult-like) are the same ones our behavioral analysis identified as mattering for explaining children's behavior.

If we look at the modeling results for adults, the best-fitting modeled listener relied on inaccurate representations that smoothed away some –but not all– of the probability contrasts available in the input for the three pronoun cues. So, the modeled learner aligns with the behavioral results by showing that for adults, all cue information mattered.

For both children and adults, what the modeling results additionally uncover is exactly how the probability information that mattered might actually matter. For children, either the representations of probability information perceived as relevant were relatively sharper (i.e., less smooth) or were deployed relatively more often than information that didn't seem to matter to them. For adults, the representations of the probability information perceived as relevant were always equally relevant (though smoothed), and always deployed to interpret pronouns in context.

4 General discussion

Here, we've looked at a case study of pronoun interpretation in Spanish, where cues are available from the pronoun form, discourse connectives, and agreement morphology about how to interpret the pronoun, and these cues can conflict with each other. Our behavioral data suggested that children and adults do indeed have differences in interpretation behavior in these contexts, highlighting that children need to change something in order to become adult-like. Analyses of the experimental data suggested that children and adults differed on whether and how much each cue to pronoun interpretation mattered. By using a computational cognitive model of the pronoun interpretation process, we were able to specify more precise differences in how these cues may matter to children and adults.

More specifically, the modeling allowed us to identify potential differences in child and adult representations of probability information relevant for interpreting pronouns as well as potential differences in child and adult deployment of that probability information. In this way, the modeling complemented the behavioral data analysis, providing a more detailed explanation for why we potentially observed the differences we did in child and adult pronoun interpretation behavior. This detail then allowed us to offer a more concrete, specific acquisition theory about what needs to change in children for them to become adult-like. In particular, children either need to change (i) only the way they represent probability information about pronoun cues (they learn to smooth the probabilities less), or (ii) both the way they represent probability information about pronoun cues (they learn to smooth the probabilities some) and the way they deploy that probability information (they always deploy it).

More specifically, our results suggested one of two options for children's observed behavior. First, children could have probability representations that are both overly-smoothed and unequally-smoothed for pronoun cue information, though they would deploy these representations all the time. This would mean that, to become adult-like, children would need to shift the way they inaccurately represent some pronoun probability information: they should smooth these probability representations less and do so equally. Notably, this shift wouldn't be a qualitative shift, as the basic way that children are interpreting pronouns is similar to that of adults: smoothed representations of pronoun cue probabilities that are always deployed. What changes is simply how much these probability representations are smoothed. More generally, what changes is not the fact that the probability representations are inaccurate, but the way that the probability representations are inaccurate.

The other option is for children to have accurate probability representations that are inaccurately deployed (i.e., selectively used), rather than being used all the time. This would mean that, to become adult-like, children would need to make two qualitative shifts. First, they would need to make their probability representations inaccurate by smoothing away relative differences the same way adults do; second, they would need to make their deployment accurate, by always using these smoothed probability representations. As with the previous option, this again underscores that being adult-like doesn't mean being accurate; it means being inaccurate in an adult-like way.

So, in a broad sense, children are similar to adults when it comes to how they rely on probability information for interpreting a pronoun in context: both children and adults are inaccurate. Our findings for adult pronoun interpretation may seem surprising at first glance – specifically that adults aren't accurate in their representation of relevant probability information for interpreting pronouns. However, recall from the introduction how a resource-optimal approach to human cognition may lead adults to be strategically inaccurate. Below, we discuss the plausibility of the specific findings about how adults appear to be inaccurate in this case study of pronoun interpretation. We then discuss limitations of the current approach, open questions, and future directions.

4.1 How plausible is it for adults to be inaccurate this way?

As discussed in the introduction, being inaccurate may be optimal when limited cognitive resources are available. So, being strategically inaccurate may be a plausible adult-like strategy for deciding how to interpret a pronoun in context. More specifically here, being resource-optimal could well lead adults to inaccurate representations of relevant probability information for pronouns.

For the representation of pronoun cue probabilities, our results suggested that adults had fairly smoothed representations, with σ=0.25-0.33 for the probability information about pronoun form, discourse connectives, and agreement morphology.[21] Within decision theory, it's long been assumed that decision makers would first "interpret" available information (Kahneman & Tversky, 1979), rather than using it accurately. Here, that would correspond to distorting the available probability contrasts in some way.

Some specific proposals for how humans seem to distort available probability information have suggested that probability distortion results from a limitation on the "dynamic range of the neural representation of probability" (Zhang & Maloney, 2012; Zhang *et al.*, 2020). That is, there's a biological limitation in the human brain related to the representation of probability information. Thus, with this limited resource as the basis for information representation, the human mind must optimize. This particular limitation leads to probabilities near the endpoints (i.e., 0.0 and 1.0) being smoothed to values closer to 0.5, narrowing the range of represented probabilities. For relative probabilities (sometimes called "judged relative frequencies"), adults tend to narrow the range to something between 0.16 and 0.80 (Zhang & Maloney, 2012; Zhang *et al.*, 2020).

Our modeling results here seem to align with this idea. In particular, the specific smoothing that our modeling results suggested takes the more extreme probabilities observed in the input for each type of pronoun information and smooths them much closer to (and often into) the range of 0.16-0.80 (see Table 6). For example, agreement morphology probabilities ($p(\textsc{mor}|\alpha)$) of 0.995-0.998 are smoothed to 0.815-0.851 with a σ_{mor}=0.25, while values of 0.002-0.005 are smoothed to 0.149-0.185.

Our findings thus seem consistent with the idea that adults in a pronoun interpretation task would distort probability distributions that are more extreme, smoothing them into less extreme relative probabilities. That is, the resource-rational solution adults have come up with, given the limited ability of their neural representations to encode the range of probabilities, is to smooth their representations of the actual probabilities available in the input.

For the representation of the prior distribution over possible antecedents, our modeling results suggest that adults completely smooth away relative contrasts from the input, effectively ignoring this information (see Table 6, where σ_α=0.00 yields an uninformative distribution). Within the realm of adult speech perception, this kind of selective ignoring has also been observed: Richardson *et al.* (2015) find that adult behavior in certain contexts is best captured by a modeled listener who attends only to the most salient phonetic feature of a phonetic category, rather than all available informative features. As Richardson *et al.* (2015) note, this strategy can allow the listener to generalize more efficiently by attributing observed variation to as few features as possible. In this way, the listener views the preferred feature(s) as "informative enough"

[21] Adults also had complete smoothing of the prior distribution over possible antecedents, σ_α=0.00. We discuss this more below.

original input probabilities								
		prior	likelihoods					
antecedent type		$p(\alpha)$	$p(\text{FORM}\vert\alpha)$ ø	$p(\text{FORM}\vert\alpha)$ overt	$p(\text{CON}\vert\alpha)$ *después*	$p(\text{CON}\vert\alpha)$ *porque*	$p(\text{MOR}\vert\alpha)$ SG	$p(\text{MOR}\vert\alpha)$ PL
SUBJ	SG	0.362	0.938	0.062	0.324	0.676	0.998	0.002
	PL	0.071	0.984	0.016	0.750	0.250	0.005	0.995
¬SUBJ	SG	0.438	0.817	0.183	0.132	0.868	0.998	0.002
	PL	0.129	0.959	0.041	0.394	0.606	0.005	0.995
smoothed input probabilities								
adult σ		0.00	0.25		0.33		0.28	
		prior	likelihoods					
antecedent type		$p(\alpha)$	$p(\text{FORM}\vert\alpha)$ ø	$p(\text{FORM}\vert\alpha)$ overt	$p(\text{CON}\vert\alpha)$ *después*	$p(\text{CON}\vert\alpha)$ *porque*	$p(\text{MOR}\vert\alpha)$ SG	$p(\text{MOR}\vert\alpha)$ PL
SUBJ	SG	0.250	0.664	0.336	0.440	0.560	0.851	0.149
	PL	0.250	0.737	0.263	0.590	0.410	0.185	0.815
¬SUBJ	SG	0.250	0.592	0.408	0.349	0.651	0.851	0.149
	PL	0.250	0.687	0.323	0.465	0.535	0.185	0.815

Table 6: Original and smoothed input probabilities for different types of antecedents (subject (SUBJ) or non-subject (¬SUBJ), singular (SG) or plural (PL)). The original input probabilities are estimated from the pronoun information distributions in naturalistic child-directed speech. The smoothed input probabilities are distorted based on the contrast parameter σ estimated for the best-fitting modeled listener for adult pronoun interpretation.

for efficient communication–and presumably achieves some cognitive resource savings by not needing to attend to, and integrate information from, other less-informative features.

4.2 Limitations, open questions, and future directions

Our results here are predicated on both the data we had available and the specific implementation choices we made. So, it can be useful to note specific components that could benefit from future investigation in order to further validate–or refute, and thus refine–both the results found here and the acquisition trajectory we posited based on those results.

First, we used the child-directed speech from the Schmitt-Miller corpus (Miller & Schmitt, 2012) as input to both the modeled child listeners and the modeled adult listeners. While adults do indeed hear child-directed speech, they also hear (presumably far more) adult-directed speech. We don't know if the probabilities we derived (i.e., the priors and likelihoods from Table 3) would change if they were instead derived from adult-directed speech; adult-directed speech is known to differ from child-directed speech in many ways (Ferguson, 1964; Snow, 1977; Grieser & Kuhl, 1988; Fernald *et al.*, 1989), though sometimes child-directed and adult-directed speech can be similar, especially for more complex representations (e.g., syntactic dependency distributions: Pearl & Sprouse 2013). Future work can examine adult-directed speech samples to derive more accurate estimates of adult pronoun interpretation priors and likelihoods, and then see if the modeling results based on those input probabilities qualitatively change.

More generally, our entire approach can be expanded to different dialects of Spanish and different languages for which we have reasonable estimates of child-directed and adult-directed speech, as well as the ability to collect behavioral judgment data in a pronoun interpretation task. We can then see if the specific hypotheses evaluated here about inaccuracy in the pronoun interpretation process hold for both children and adults.

Second, we assumed that the modeled listener believed the cues to pronoun interpretation we investigated–pronoun form, discourse connectives, and agreement morphology–were independent. So, for instance, the listener assumed the form a pronoun took wasn't related to the discourse connective that was used or the agreement morphology available in context. We noted previously that this was an idealization, which had the effect of simplifying the calculation of the Bayesian baseline model posterior probabilities. More specifically, in our implementation, this meant that the likelihood of a specific set of cue values (e.g., null form, *porque*, singular morphology) was simply the product of their individual likelihoods (e.g., the likelihood for the null form, multiplied by the likelihood for *porque*, multiplied by the likelihood of singular morphology). The modeled listener didn't need to track all the combinations of feature values, and their potential interactions.

If instead the modeled listener did have to track all the possible combinations, the information in the input would be much scarcer. For instance, to calculate the likelihood of the null form, *porque*, and singular morphology, the listener would need to estimate the following likelihood combinations from the input, in addition to the individual likelihoods noted before: (i) the likelihood of the null form and *porque*, (ii) the likelihood of the null form and singular morphology, (iii) the likelihood of *porque* and singular morphology, and (iv) the likelihood of the null form, *porque*, and singular morphology. One or more of these combinations may appear rarely, making accurate estimation from the input difficult. Even with estimates for all the likelihoods in hand, the modeled listener would then need to aggregate the information from all these estimates in a principled way. This aggregation process would need to be specified, as it's part of the implemented theory of how the modeled listener integrates information from multiple cues.

We note here that assuming independence may be plausible for humans to do when multiple features are available; in the realm of visual perception, Vul & Rich (2010) find that human behavior is best accounted for by a modeled observer who considers the probability distribution of each potential feature independently, rather than probability distributions for collections of features. This strategy in turn may have its origin in being resource-rational: assuming features are independent may yield "good enough" perception in the visual domain or "good enough" comprehension in the language domain, while being frugal with cognitive resources. Still, assuming pronoun cues are independent may not be what adult humans actually do, and future work can look at the consequences of relaxing this idealizing assumption.

Another modeling assumption that could be relaxed in future investigations involves the treatment of agreement morphology. Here, the modeled listeners had a single parameter for both singular and plural agreement morphology. However, the additional analyses in Appendix D.2 suggest that children are impacted differently by the agreement morphology number. So, it may be worthwhile to have one parameter for plural agreement morphology and another for singular agreement morphology. Such a distinction may allow those future modeled listeners to better fit the observed data, and would correspond to the hypothesis that there could be separate processes affecting plural vs. singular agreement morphology in the listener's mind. Importantly, such modeled listeners would also be more complex than the modeled listeners here because they would have more parameters. So, to offset that added complexity, those future modeled listeners would have to fit the observed data substantially better. It remains to be seen if those modeled listeners would therefore better explain children's interpretation behavior (and/or adults' interpretation behavior).

In addition, as we noted when first introducing our implementations of modeled listeners with inaccurate representations, inaccurate deployment, or both, there are many other reasonable ways to implement these ideas concretely in modeled listeners. We aimed to make reasonable, cognitively-motivated choices in our implementations, but there are surely many other options. Future work can investigate other cognitively-motivated implementation options for inaccurate representations and inaccurate deployment, and see if the results we found here about the differences between child and adult pronoun interpretation hold up.

For instance, the ACT-R computational cognitive modeling framework (Anderson, 2009) can be used to concretely implement inaccurate deployment due to a variety of cognitive factors, such as working memory and processing speed limitations. Vogelzang *et al.* (2021) have used this approach to investigate the explanatory power of inaccurate deployment alone for pronoun interpretation behavior by children and adults in Italian. Future work can investigate whether this same ACT-R implementation, which encodes specific causes of inaccurate deployment while maintaining accurate representations of relevant knowledge, can capture the child and adult pronoun interpretation behavior found here for Spanish. If so, this would support an explanation of child and adult behavior that rests on inaccurate deployment of accurate representations for both children and adults, due to the cognitive factors identified in the ACT-R implementation. More generally, future work can also investigate other cognitively-motivated models, including other Bayesian and ACT-R implementations, that may be able to explain the observed differences between child and adult pronoun interpretation behavior in these contexts.

5 Conclusion

We used a combination of behavioral experiments and computational cognitive modeling to better understand the differences between child and adult pronoun interpretation in context; the findings from this combined methodological approach allowed us to propose a more concrete acquisition theory about pronoun interpretation. A key finding is that being adult-like doesn't mean being accurate when it comes to the information available for pronoun interpretation: our results suggest that adults are inaccurate in how they represent probability information relevant for interpreting pronouns in context. It's likely that the specific ways that adults are inaccurate are useful for language comprehension, given that there are limited cognitive resources to deploy (even for adults); so, children must learn not to become accurate, but rather how to become strategically inaccurate.

We also hope to have shown how computational cognitive modeling, when empirically-grounded, can complement behavioral studies. Here, we used computational cognitive modeling to uncover the potential underlying causes of the behavior observed in children and adults. The modeling offered concrete explanations both for what potentially causes non-adult-like behavior in children, and also what potentially causes adult-like behavior. Of course, just as with behavioral studies, computational studies involve making simplifying assumptions; how much we believe the results of those studies rests on the plausibility of the assumptions that went into the studies. More generally, we believe that behavioral and computational methods, when used together, can help us make significant progress when it comes to defining and refining our theories about what acquisition actually is in any particular domain. We hope to have done that here for pronoun interpretation.

References

Acuña-Fariña, J Carlos, Meseguer, Enrique, & Carreiras, Manuel. 2014. Gender and number agreement in comprehension in Spanish. *Lingua*, **143**, 108–128.

Alonso-Ovalle, Luis, Fernandez-Solera, Susana, Frazier, Lyn, & Clifton, Charles. 2002. Null vs. overt pronouns and the topic-focus articulation in Spanish. *Rivista di Linguisitica*, **14**(2), 1–19.

Anderson, John R. 2009. *How can the human mind occur in the physical universe?* Oxford University Press.

Arnold, Jennifer, Eisenband, Janet, Brown-Schmidt, Sarah, & Trueswell, John C. 2000. The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, **76**, B13–B26.

Asher, Nicholas, & Lascarides, Alex. 2003. *Logics of Conversation*. Cambridge: Cambridge University Press.

Bates, Douglas, Maechler, Martin, Bolker, Ben, Walker, Steven, Christensen, Rune Haubo Bojesen, Singmann, Henrik, Dai, Bin, Grothendieck, Gabor, Green, Peter, & Bolker, Maintainer Ben. 2015a. Package 'lme4'. *convergence*, **12**(1), 2.

Bates, Douglas, Kliegl, Reinhold, Vasishth, Shravan, & Baayen, Harald. 2015b. Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Brandt-Kobele, Oda-Christina, & Höhle, Barbara. 2010. What asymmetries within comprehension reveal about asymmetries between comprehension and production: The case of verb inflection in language acquisition. *Lingua*, **120**(8), 1910–1925.

Brown, Roger. 1973. *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Carminati, Maria Nella. 2002. *The Processing of Italian Subject Pronouns*. Ph.D. thesis, University of Massachusetts, Amherst.

Carvalho, Ana M., Orozco, Rafael, & Shin, Naomi Lapidus (eds). 2015. *Subject pronoun expression in Spanish: a cross-dialectal perspective*. Georgetown University Press.

Chambers, Craig, & Smyth, Ron. 1998. Structural Parallelism and Discourse Coherence: A test of centering theory. *Journal of Memory and Language*, **39**, 593–608.

Chater, Nick, Clark, Alexander, Goldsmith, John, & Perfors, Amy. 2015. *Empiricism and language learnability*. Oxford University Press.

Chomsky, Noam. 1969. *Aspects of the Theory of Syntax*. Vol. 11. MIT press.

Clahsen, Harald, Aveledo, Friabet, & Roca, Iggy. 2002. The development of regular and irregular verb inflection in Spanish child language. *Journal of Child Language*, **29**, 591–622.

Conroy, Anastasia, Takahashi, Eri, Lidz, Jeffrey, & Phillips, Colin. 2009. Equal Treatment for All Antecedents: How Children Succeed with Principle B. *Linguistic Inquiry*, **40**(3), 446–486.

Crain, Stephen, & Thornton, Rosalind. 1998. *Investigations in Universal Grammar*.

Crawley, Rosalind A., Stevenson, Rosemary, & Kleinman, David. 1990. The Use of Heuristic Strategies in the Interpretation of Pronouns. *Journal of Psycholinguistic Research*, **19**(4), 245–264.

Davies, Benjamin, Xu Rattanasone, Nan, & Demuth, Katherine. 2017. Two-year-olds' sensitivity to inflectional plural morphology: Allomorphic effects. *Language Learning and Development*, **13**(1), 38–53.

Denison, Stephanie, Reed, Christie, & Xu, Fei. 2011. The emergence of probabilistic reasoning in very young infants. *In: Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 33.

Dewar, Kathryn M, & Xu, Fei. 2010. Induction, overhypothesis, and the origin of abstract knowledge: Evidence from 9-month-old infants. *Psychological Science*, **21**(12), 1871–1877.

Ferguson, Charles A. 1964. Baby talk in six languages. *American Anthropologist*, **66**(6_PART2), 103–114.

Fernald, Anne, Taeschner, Traute, Dunn, Judy, Papousek, Mechthild, de Boysson-Bardies, Bénédicte, & Fukui, Ikuko. 1989. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, **16**(3), 477–501.

Ferreira, Fernanda, & Patson, Nikole D. 2007. The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, **1**(1-2), 71–83.

Ferreira, Fernanda, Bailey, Karl GD, & Ferraro, Vittoria. 2002. Good-enough representations in language comprehension. *Current Directions in Psychological Science*, **11**(1), 11–15.

Filiaci, Francesca. 2010. Null and overt Subject Biases in Spanish and Italian: A cross-linguistic comparison. *Pages 171–182 of:* Borgonovo, C, Español-Echevarria, M, & Prévost, P (eds), *Selected Proceedings of the 12th Hispanic Linguistics Symposium*. Cascadilla Press.

Forsythe, Hannah, & Schmitt, Cristina. 2021. Considering the whole paradigm: Preschoolers' comprehension of agreement is not uniformly late. *Language Acquisition*, **28**(3), 272–293.

Forsythe, Hannah, Greeson, Daniel, & Schmitt, Cristina. 2019. How preschoolers acquire the null-overt contrast in Mexican Spanish: Evidence from production. *In:* Colomina-Almiñana, Juan, & Sessarego, Sandro (eds), *Patterns in Spanish: Structure, Context and Development*, vol. 1. John Benjamins.

Forsythe, Hannah, Greeson, Daniel, & Schmitt, Cristina. 2022. After the Null Subject Parameter: Acquisition of the Null-Overt Contrast in Spanish. *Language Learning and Development*, **18**(2), 171–200.

Frank, Michael C, & Goodman, Noah D. 2012. Predicting pragmatic reasoning in language games. *Science*, **336**(6084), 998–998.

Friston, Karl. 2010. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, **11**(2), 127–138.

Fukumura, Kumiko, & van Gompel, Roger. 2010. Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language*, **62**, 52–66.

Gagliardi, Annie, & Lidz, Jeffrey. 2014. Statistical insensitivity in the acquisition of Tsez noun classes. *Language*, **90**(1), 58–89.

Gagliardi, Annie, Feldman, Naomi H, & Lidz, Jeffrey. 2017. Modeling statistical insensitivity: Sources of suboptimal behavior. *Cognitive Science*, **41**(1), 188–217.

Gerken, LouAnn. 2006. Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition*, **98**(3), B67–B74.

Gerken, LouAnn. 2010. Infants use rational decision criteria for choosing among models of their input. *Cognition*, **115**(2), 362–366.

Gigerenzer, Gerd. 2008. *Rationality for mortals: How people cope with uncertainty*. Oxford University Press.

González-Gómez, Nayeli, Hsin, Lisa, Barriere, Isabelle, Nazzi, Thierry, & Legendre, Geraldine. 2017. Agarra, agarran: Evidence of early comprehension of subject–verb agreement in Spanish. *Journal of Experimental Child Psychology*, **160**, 33–49.

Goodman, Noah D, & Stuhlmüller, Andreas. 2013. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*, **5**(1), 173–184.

Gor, Vera. 2020. *Experimental investigations of Principle C at the syntax-pragmatics interface*. Ph.D. thesis, Rutgers The State University of New Jersey, School of Graduate Studies.

Gor, Vera, & Syrett, Kristen. 2018. Beyond Principle C:(Not)-at-issueness and plausibility influence acceptability of coconstrual. *Pages 109–123 of: Proceedings from the Annual Meeting of the Chicago Linguistic Society*, vol. 54. Chicago Linguistic Society.

Grieser, DiAnne L, & Kuhl, Patricia K. 1988. Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology*, **24**(1), 14.

Gxilishe, Sandile, Smouse, Mantoa Rose, Xhalisa, Thabisa, & deVilliers, Jill. 2009. Children's Insensitivity to Information from the Target of Agreement: The Case of Xhosa. *Pages 46–53 of:* Crawford, Jean, Otaki, Koichi, & Takahashi, Masahiko (eds), *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA 2008)*. Somerville, MA: Cascadilla Proceedings Project.

Haghighi, Aria, & Klein, Dan. 2007. Unsupervised coreference resolution in a nonparametric Bayesian model. *Pages 848–855 of: Proceedings of the 45th annual meeting of the Association of Computational Linguistics*.

Hartshorne, Joshua K., Nappa, Rebecca, & Snedeker, Jesse. 2015. Development of the First-Mention Bias. *Journal of Child Language*, **42**(2).

Heitz, Richard P. 2014. The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, **8**, 150.

Hobbs, Jerry R. 1979. Coherence and coreference. *Cognitive Science*, **3**, 67–90.

Hsu, Anne S, & Chater, Nick. 2010. The logical problem of language acquisition: A probabilistic perspective. *Cognitive Science*, **34**(6), 972–1016.

Hsu, Anne S, Chater, Nick, & Vitányi, Paul MB. 2011. The probabilistic analysis of language acquisition: Theoretical, computational, and experimental analysis. *Cognition*, **120**(3), 380–390.

Hsu, Anne S, Chater, Nick, & Vitányi, Paul. 2013. Language Learning From Positive Evidence, Reconsidered: A Simplicity-Based Approach. *Topics in Cognitive Science*, **5**(1), 35–55.

Järvikivi, Juhani, van Gompel, Roger, Hyönä, Jukka, & Bertram, Raymond. 2005. Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science*, **16**(4), 260–264.

Jegerski, Jill, VanPatten, Bill, & Keating, Gregory D. 2011. Cross-linguistic variation and the acquisition of pronominal reference in L2 Spanish. *Second Language Research*, **27**(4), 481–507.

Johnson, Valerie, de Villiers, Jill, & Seymore, Harry. 2005. Agreement without understanding? The case of third person singular /s/. *First Language*, **25**(3), 317–330.

Jurafsky, Dan. 2020. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Prentice hall.

Kahneman, Daniel, & Tversky, Amos. 1979. Prospect theory: An analysis of decision under risk. *Econometrica*, **47**(2), 363–391.

Kass, Robert E, & Raftery, Adrian E. 1995. Bayes Factors. *Journal of the American Statistical Association*, **90**(430), 773–795.

Keating, Gregory D, VanPatten, Bill, & Jegerski, Jill. 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, Gregory D, Jegerski, Jill, & Vanpatten, Bill. 2016. Online processing of subject pronouns in monolingual and heritage bilingual speakers of Mexican Spanish. *Bilingualism: Language and Cognition*, **19**(1), 36–49.

Kehler, Andrew. 2002. *Coherence, Reference, and the Theory of Grammar*. CSLI Publications.

Kehler, Andrew, Kertz, Laura, Rohde, Hannah, & Elman, Jeffrey L. 2008. Coherence and coreference revisited. *Journal of Semantics*, **25**(1), 1–44.

Kouider, Sid, Halberda, Justin, Wood, Justin, & Carey, Susan. 2006. Acquisition of English number marking: The singular-plural distinction. *Language Learning and Development*, **2**(1), 1–25.

Lago, Sol, Shalom, Diego E, Sigman, Mariano, Lau, Ellen F, & Phillips, Colin. 2015. Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, **82**, 133–149.

Legendre, Geraldine, Barriere, Isabelle, Goyet, Louise, & Nazzi, Thierry. 2010. Comprehension of Infrequent Subject–Verb Agreement Forms: Evidence From French-Learning Children. *Child Development*, **81**(6), 1859–1875.

Legendre, Geraldine, Barriere, Isabelle, Goyet, Louise, & Nazzi, Thierry. 2011. On the acquisition of implicated presuppositions: Evidence from French personal pronouns. *Pages 150–162 of:* Pirvulescu, Michaela (ed), *Selected Proceedings of the 4th Conference on Generative Approaches to Language Acquisition North America*. Somerville, MA: Cascadilla Proceedings Project.

Legendre, Geraldine, Culbertson, Jennifer, Zaroukian, Erin, Hsin, Lisa, Barrière, Isabelle, & Nazzi, Thierry. 2014. Is children's comprehension of subject-verb agreement universally late? Comparative evidence from French, English, and Spanish. *Lingua*.

Lenth, Russell. 2021. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.7.1-1.

Li, Ming, & Vitányi, Paul MB. 1994. Inductive reasoning. *In:* Ristad, E. (ed), *Language computations*. Philadelphia: American Mathematical Society.

Lieder, Falk, & Griffiths, Thomas L. 2020. Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, **43**, 1–60.

Maratsos, Michael. 1974. Preschool Children's Use of Definite and Indefinite Articles. *Child Development*, **45**(2), 446–455.

Markman, Arthur B, & Otto, A Ross. 2011. Cognitive systems optimize energy rather than information. *Behavioral and Brain Sciences*, **34**(207), 10–1017.

Martin, Andrea E. 2016. Language processing as cue integration: Grounding the psychology of language in perception and neurophysiology. *Frontiers in Psychology*, **7**, 120–136.

Matuschek, Hannes, Kliegl, Reinhold, Vasishth, Shravan, Baayen, Harald, & Bates, Douglas. 2017. Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, **94**, 305–315.

Messenger, Katherine, & Fisher, Cynthia. 2018. Mistakes weren't made: Three-year-olds' comprehension of novel-verb passives provides evidence for early abstract syntax. *Cognition*, **178**, 118–132.

Miller, Karen, & Schmitt, Cristina. 2012. Variable Input and the Acquisition of Plural Morphology. *Language Acquisition*, **19**, 223–266.

Mitchener, William Garrett, & Becker, Misha. 2010. Computational models of learning the raising-control distinction. *Research on Language and Computation*, **8**(2-3), 169–207.

Nguyen, Emma, & Pearl, Lisa. 2019. Using Developmental Modeling to Specify Learning and Representation of the Passive in English Children. *Pages 469–482 of:* Brown, Megan M., & Dailey, Brady (eds), *Proceedings of the 43rd annual Boston University Conference on Language Development*.

Orita, Naho, McKeown, Rebecca, Feldman, Naomi H, Lidz, Jeffrey, & Boyd-Graber, Jordan. 2013. Discovering pronoun categories using discourse information. *In: Proceedings of the 35th annual conference of the cognitive science society*.

Otheguy, Ricardo, Zentella, Ana Celia, & Livert, David. 2010. Generational differences in pronominal usage in Spanish reflecting language and dialect contact in a bilingual setting. *Pages 45–62 of: Language contact: New perspectives*. Amsterdam: John Benjamins.

Papadopoulou, Despina, Peristeri, Eleni, Plemenou, Evagelia, Marinis, Theodoros, & Tsimpli, Ianthi. 2015. Pronoun ambiguity resolution in Greek: Evidence from monolingual adults and children. *Lingua*, **155**, 98–120.

Pearl, Lisa. 2023. Modeling syntactic acquisition. *Pages 209–270 of:* Sprouse, Jon (ed), *Oxford Handbook of Experimental Syntax*. Oxford University Press.

Pearl, Lisa, & Mis, Benjamin. 2016. The role of indirect positive evidence in syntactic acquisition: A look at anaphoric one. *Language*, **92**(1), 1–30.

Pearl, Lisa, & Sprouse, Jon. 2013. Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition*, **20**, 19–64.

Pearl, Lisa, & Sprouse, Jon. 2019. Comparing solutions to the linking problem using an integrated quantitative framework of language acquisition. *Language*, **95**(4), 583–611.

Pearl, Lisa, Goldwater, Sharon, & Steyvers, Mark. 2011. Online Learning Mechanisms for Bayesian Models of Word Segmentation. *Research on Language and Computation*, **8**(2), 107–132.

Peirce, Jonathan, Gray, Jeremy R, Simpson, Sol, MacAskill, Michael, Höchenberger, Richard, Sogo, Hiroyuki, Kastman, Erik, & Lindeløv, Jonas Kristoffer. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, **51**(1), 195–203.

Pérez-Leroux, Ana T, & Glass, William R. 1999. Null anaphora in Spanish second language acquisition: probabilistic versus generative approaches. *Second Language Research*, **15**(2), 220–249.

Perfors, Amy, Tenenbaum, Joshua, & Regier, Terry. 2011. The learnability of abstract syntactic principles. *Cognition*, **118**, 306–338.

Phillips, Lawrence, & Pearl, Lisa. 2014a. Bayesian inference as a cross-linguistic word segmentation strategy: Always learning useful things. *In: Proceedings of the Computational and Cognitive Models of Language Acquisition and Language Processing Workshop*. Gothenberg, Sweden: EACL.

Phillips, Lawrence, & Pearl, Lisa. 2014b. Bayesian inference as a viable cross-linguistic word segmentation strategy: It's all about what's useful. *Pages 2775–2780 of: Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Quebec City, CA: Cognitive Science Society.

Phillips, Lawrence, & Pearl, Lisa. 2015a. Utility-based evaluation metrics for models of language acquisition: A look at speech segmentation. *Proceedings of CMCL*, 68–78.

Phillips, Lawrence, & Pearl, Lisa. 2015b. The utility of cognitive plausibility in language acquisition modeling: Evidence from word segmentation. *Cognitive Science*, **doi:10.1111/cogs.12217**.

Pyykkönen, Pirita, Matthews, Danielle, & Järvikivi, Juhani. 2010. Three-year-olds are sensitive to semantic prominence during online language comprehension: A visual world study of pronoun resolution. *Language and Cognitive Processes*, **25**(1), 115–129.

Pérez-Leroux, Ana Teresa. 2005. Number Problems in Children. *Pages 1–12 of:* Gurski, C. (ed), *Proceedings of the 2005 Canadian Linguistics Association Annual Conference*.

Rastegar, Zahra, Shirazi, Haghighi, & Sadighi, Firooz. 2012. An Amazing Conundrum in Children's Comprehension and Production of Verb Inflection. *World Applied Sciences Journal*, **18**(8), 1095–1101.

Richardson, Rachael R, Feldman, Naomi H, & Idsardi, William. 2015. What defines a category? Evidence that listeners' perception is governed by generalizations. *In: Proceedings of the Annual Conference of the Cognitive Science Society*.

Rissanen, Jorma, & Ristad, Eric Sven. 1994. Language acquisition in the MDL framework. *Pages 149–166 of:* Ristad, Eric Sven (ed), *Language computations*.

Rohde, Hannah, & Kehler, Andrew. 2014. Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience*, **29**(8), 912–927.

Schwarz, Gideon. 1978. Estimating the dimension of a model. *Annals of Statistics*, **6**(2), 461–464.

Scontras, Gregory, & Goodman, Noah D. 2017. Resolving uncertainty in plural predication. *Cognition*, **168**, 294–311.

Shin, Naomi Lapidus. 2016. Acquiring constraints on morphosyntactic variation: Children's Spanish subject pronoun expression. *Journal of Child Language*, **43**, 914–947.

Shin, Naomi Lapidus, & Cairns, Hellen Smith. 2012. The Development of NP Selection in School-Age Children: Reference and Spanish Subject Pronouns. *Language Acquisition*, **19**, 3–38.

Simon, Herbert A. 1956. Rational choice and the structure of the environment. *Psychological Review*, **63**(2), 129–138.

Smyth, Ron. 1994. Grammatical Determinants of Ambiguous Pronoun Resolution. *Journal of Psycholinguistic Research*, **23**(3), 197–229.

Snow, Catherine E. 1977. The development of conversation between mothers and babies. *Journal of Child Language*, **4**(1), 1–22.

Soderstrom, Melanie. 2002. *The acquisition of inflection morphology in early perceptual knowledge of syntax*. Ph.D. thesis, Johns Hopkins University, Baltimore, MD.

Soderstrom, Melanie, Wexler, Ken, & Jusczyk, P.W. 2002. English-learning toddlers' sensitivity to agreement morphology in receptive grammar. *Pages 643–52 of:* Skarabela, B., Fish, S., & Do, A. H.-J. (eds), *Proceedings of the 26th Boston University Conference on Language Development (BUCLD26)*, vol. 2. Sommerville, MA: Cascadilla Proceedings Project.

Soderstrom, Melanie, White, Katherine S., Conwell, Erin, & Morgan, James L. 2007. Receptive Grammatical Knowledge of Familiar Content Words and Inflection in 16-Month-Olds. *Infancy*, **12**(1), 1–29.

Sonderegger, Morgan. 2023. *Regression modeling for linguistic data*. MIT Press.

Song, Hyun-Joo, & Fisher, Cynthia. 2005. Who's "She?": Discourse prominence influences preschooler's comprehension of pronouns. *Journal of Memory and Language*, **52**(1), 29–57.

Song, Hyun-Joo, & Fisher, Cynthia. 2007. Discourse prominence effects on 2.5-year-old children's interpretation of pronouns. *Lingua*, **117**(11), 1959–1987.

Sorace, Antonella, Serratrice, Ludovica, Filiaci, Francesca, & Baldo, Machela. 2009. Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua*, **119**, 460–477.

Spenader, Jennifer, Smits, Erik-Jan, & Hendriks, Petra. 2009. Coherent discourse solves the pronoun interpretation problem. *Journal of Child Language*, **36**, 23–52.

Stabler, Edward P. 1998. Acquiring languages with movement. *Syntax*, **1**(1), 72–97.

Theakston, Anna L., Lieven, E. V., & Tomasello, Michael. 2003. The role of the input in the acquisition of third person singular verbs in English. *Journal of Speech, Language, and Hearing Research*, **46**, 863–877.

Torregrossa, Jacopo, Bongartz, Christiane, & Tsimpli, Ianthi Maria. 2019. Bilingual reference production: A cognitive-computational account. *Linguistic Approaches to Bilingualism*, **9**(4-5), 569–599.

Traxler, Matthew J. 2014. Trends in syntactic parsing: Anticipation, Bayesian estimation, and good-enough parsing. *Trends in Cognitive Sciences*, **18**(11), 605–611.

Ud Deen, K, Bondoc, Ivan, Camp, Amber, Estioca, Sharon, Hwang, Haerim, Shin, Gyu-Ho, Takahashi, Maho, Zenker, Fred, & Zhong, J. 2018. Repetition brings success: Revealing knowledge of the passive voice. *Pages 200–213 of: Proceedings of the 42nd Annual Boston University Conference on Language Development*.

Verhagen, Josje, & Blom, Elma. 2014. Asymmetries in the acquisition of subject-verb agreement in Dutch? Evidence from comprehension and production. *First Language*, **34**(4), 315–335.

Vogelzang, Margreet, Guasti, Maria Teresa, van Rijn, Hedderik, & Hendriks, Petra. 2021. How children process reduced forms: a computational cognitive modeling approach to pronoun processing in discourse. *Cognitive Science*, **45**(4), e12951.

Vul, Edward, & Rich, Anina N. 2010. Independent sampling of features enables conscious perception of bound objects. *Psychological Science*, **21**(8), 1168–1175.

Vul, Edward, Goodman, Noah, Griffiths, Thomas L, & Tenenbaum, Joshua B. 2014. One and done? Optimal decisions from very few samples. *Cognitive Science*, **38**(4), 599–637.

Wagers, Matthew W, Lau, Ellen F, & Phillips, Colin. 2009. Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, **61**(2), 206–237.

Wood, Justin N, Kouider, Sid, & Carey, Susan. 2009. Acquisition of singular-plural morphology. *Developmental Psychology*, **45**(1), 202.

Wykes, Til. 1981. Inference and Children's Comprehension of Pronouns. *Journal of Experimental and Child Psychology*, **32**, 264–278.

Xu, Fei, & Tenenbaum, Joshua. 2007. Word Learning as Bayesian Inference. *Psychological Review*, **114**(2), 245–272.

Yang, Charles. 2005. On productivity. *Yearbook of Language Variation*, **5**, 333–370.

Yang, Charles. 2016. *The price of linguistic productivity: How children learn to break the rules of language*. MIT Press.

Yang, Charles. 2018. A User's Guide to the Tolerance Principle. *lingbuzz*, lingbuzz/004146.

Zhang, Hang, & Maloney, Laurence T. 2012. Ubiquitous log odds: a common representation of probability and frequency distortion in perception, action, and cognition. *Frontiers in Neuroscience*, **6**, 1.

Zhang, Hang, Ren, Xiangjuan, & Maloney, Laurence T. 2020. The bounded rationality of probability distortion. *Proceedings of the National Academy of Sciences*, **117**(36), 22024–22034.

A Children's pronoun interpretation: Prior work

We briefly summarize the existing literature on children's production and comprehension of the specific linguistic cues that Spanish-learning children can use to extract these different types of information: null and overt subject pronouns, clausal relations, and the grammatical feature of subject-verb agreement.

Null vs. overt pronouns. In so-called "canonical null subject" languages like Spanish (as well as Italian and Greek), adults are more likely to interpret subject pronouns as referring to the preceding subject antecedent when the form is null, in contrast to when the form is overt (Carminati, 2002; Alonso-Ovalle *et al.*, 2002; Filiaci, 2010; Keating *et al.*, 2016). Adult learners of these languages are generally very slow to acquire this contrast (e.g. Pérez-Leroux & Glass 1999; Keating *et al.* 2011; Jegerski *et al.* 2011), but for typically-developing monolingual children, performance varies by task.

In spontaneous production, monolingual children as young as four to five years old respect this contrast (Shin, 2016; Forsythe *et al.*, 2019) – that is, they spontaneously use null pronouns more often when referring to subject antecedents than to other antecedents. In pronoun interpretation tasks like the one in (6), children begin displaying different antecedent preferences for null and overt pronouns by age four and a half to six in Spanish (Forsythe *et al.*, 2022) and by age six to seven in Greek (Papadopoulou *et al.*, 2015).

(6) Juan le pega a Pedro y después **Ø /él** se va.
Juan DAT hits A Pedro and then *pro*/he leaves.
TASK: Choose one picture: (i) Juan leaving vs. (ii) Pedro leaving

That is, when asked to interpret sentences like (6), children select the subject antecedent *Juan* more often when the null pronoun is used, compared to when the overt pronoun *él* is used. Taken together, these results would suggest that older children (four and a half to six in Spanish, six to seven in Greek) have the correct representation of the null form (i.e., it favors the subject antecedent) and can successfully deploy that representation in in the moment.

However, in felicity judgment tasks like the one in (7), adult-like preferences are much slower to develop (Shin & Cairns, 2012; Sorace *et al.*, 2009).

(7) CONTEXT: María and José sing; María sings a ranchera; José sings the Pimpón song.

a. Investigator: María y José cantan canciones. María canta una ranchera.
Maria and José sing songs. Maria sings a ranchera.

b. Puppet A: Luego **Ø** canta la de Pimpón.
Then *pro* sings one about Pimpón.

c. Puppet B: Luego **él** canta la de Pimpón.
Then he sings one about Pimpón.

TASK: Choose the more appropriate description (Puppet A or B).

In (7), the context establishes that *José* is the antecedent of the pronoun used by each of the puppets – this is because José is the one who sings the Pimpón song after María sings a ranchera. Importantly, José isn't the subject of the preceding (second) sentence in (7a) – María is. Therefore, it's not adult-like for Puppet A to use a null pronoun in (7b), since the null pronoun is biased towards María; so, the child should choose the description provided by Puppet B in (7c).

Shin & Cairns (2012) report that up until age eight, children acquiring Mexican Spanish failed to reliably choose the overt form in contexts like (7). Conversely, in contexts where María sings all three times and the pronoun therefore refers to the preceding subject *María*, the most appropriate choice would be to use a null pronoun, instead of the overt pronoun *ella* ('she'). However, even up to ages fourteen and fifteen, children failed to reliably choose the puppet who used the null pronoun in these all-María contexts.

The dramatic difference in children's performance across studies (and therefore, the inferred age of acquisition when it comes to the appropriate pronoun cue representation) underscores the impact of an experimental task's cognitive demands. Felicity judgment tasks like (7) seem much more cognitively-demanding than pronoun interpretation tasks like (6); this is because the participant has to reason about what speakers are likely to say in a given context (e.g., that it's better – but not categorically required – to choose the overt pronoun when referring to the non-subject antecedent *José*). This kind of pragmatic reasoning seems fairly sophisticated. Moreover, felicity judgment tasks place heavy demands on children's working memory load by requiring them to activate two utterances for as long as needed to decide which one is more appropriate for the scene.

In contrast, the pronoun interpretation task in (6) asks children to compare two different interpretations, both of which are presented visually. So, the child is reasoning about what the speaker was referring to (a more natural comprehension task), and the child doesn't have to maintain the two interpretations in memory while deciding which one to choose. These and other task differences could well be why we see such large age differences in adult-like pronoun interpretation behavior (i.e., eight and 14-15 years old in felicity judgment tasks vs. four to five years old in pronoun interpretation tasks).

Given these two very different tasks, it seems that children's representation of Spanish subject pronouns is at least fairly mature (though perhaps not fully adult-like) by four to five years old, because they understand that null pronouns are more biased towards the subject antecedent than overt pronouns. However, children seem unable to deploy these representations in more cognitively-demanding tasks until much later (i.e., eight or 14-15, depending on the representation).

Clausal relations. Pronoun interpretation can also be impacted by the relation between the clause containing the pronoun and other clauses in the discourse (Kehler, 2002; Asher & Lascarides, 2003; Kehler *et al.*, 2008). The relation can be signaled by a variety of syntactic and lexical cues. When these cues change, the underlying relation between clauses also changes, potentially triggering a change in the preferred pronoun interpretation.

For example, narrative relations arise when one event follows another, which can be signaled by the lexical connective *then*, as in *Lisa sang to Lindy, and then she took a nap*. Asher & Lascarides (2003) argue that this type of relation constrains pronoun interpretation by biasing clauses to maintain reference to the same topic. One way to mark a sentential topic is to mention it first and/or in the subject position; so, either of these cues should bias the interpretation of the pronoun *she* toward the antecedent *Lisa*. More generally, the commonly observed "first-mention bias" (Crawley *et al.*, 1990; Arnold *et al.*, 2000; Järvikivi *et al.*, 2005), in which listeners favor the antecedent mentioned first and/or in subject position, falls out naturally from this information about event sequences.

Syntactic structure can also serve as a clausal cue, with subject pronouns preferring subject antecedents and object pronouns preferring object antecedents (Chambers & Smyth, 1998); this tendency is heightened when clause-internal constituents have the same number and placement (Smyth, 1994), as in (8). Here, both clauses include the structure *subject-verb-object* (*Samuel threatened Justin...*, *he/Erin blindfolded Erin/him*). So, when the pronoun in the second clause is in the subject position, it's more often interpreted as the first clause's subject *Samuel*; when the pronoun in the second clause is in the object position, it's more often interpreted as the first clause's object *Justin* (Kehler *et al.*, 2008). Parallel syntactic structures are argued to induce maximally parallel interpretations (Kehler, 2002; Asher & Lascarides, 2003), which is why the pronoun interpretation shifts this way.

(8) Parallel structure (Kehler *et al.* (2008), Expt.1)

a. Samuel threatened Justin with a knife, and he blindfolded Erin.

b. Samuel threatened Justin with a knife, and Erin blindfolded him.

However, syntactic cues like parallel structure can be overcome by the pragmatic cue of real world knowledge (typically signaled by specific lexical items), which causes listeners to pick the most situationally-appropriate interpretation (Hobbs, 1979; Gor & Syrett, 2018; Gor, 2020). An example is in (9) (Kehler *et al.*, 2008), where the clauses again have parallel syntactic structure (i.e., *subject-verb-object*), but real world knowledge (cued by the meaning of the specific lexical items involved) overrides those preferences. In (9a), the knowledge that the people who *alert security* are more likely to be the ones *threatened* (like the object *Justin*) causes the subject pronoun *he* to be interpreted as the object *Justin*. Similarly, in (9b), the knowledge that people who *threaten* other people are more likely to be the ones *stopped* (like the subject *Samuel*) causes object pronoun *him* to be interpreted as the subject *Samuel*.

(9) Pragmatic context (Kehler *et al.* (2008), Expt.1)

a. Samuel threatened Justin with a knife, and he alerted security.

b. Samuel threatened Justin with a knife, and Erin stopped him.

While there are few studies related to whether children are sensitive to discourse coherence in pronoun interpretation, we have some evidence that preschool children can leverage these lexical and syntactic cues to interpret pronouns. For clausal relation cues, children ages three to five have a "first mention" or subject-antecedent bias in a variety of contexts (Song & Fisher, 2005, 2007; Pyykkönen *et al.*, 2010; Hartshorne *et al.*, 2015). For the syntactic cue of parallel structure, children as young as three tend to interpret pronouns in parallel syntactic contexts in the appropriate parallel syntactic position (Maratsos, 1974). For example, three-year-olds tend to act out sentences like *Susie jumped over the old woman, and then Harry jumped over her*, in ways that show that the three-year-olds link the object pronoun *her* to the syntactic object *old woman*. For the pragmatic cue of real world knowledge, five-year-olds can interpret pronouns in situationally-appropriate ways in act-out tasks (Wykes, 1981). For example, consider this utterance sequence: *Jane needed Susan's pencil. She gave it to her.* Five-year-olds tend to act out this sequence by interpreting *she* as *Susan* (who possesses the pencil and could therefore give it to someone else) and interpreting *her* as *Jane* (who needs a pencil, and therefore would be a plausible recipient of pencil-giving).

Taken together, these studies suggest that children can apply adult-like pronoun interpretation strategies under a variety of discourse conditions, though of course much work remains to be done in this area. There is less evidence about what they do when clausal relation cues conflict with each other, as with the syntactic parallel structure cue and the pragmatic real world knowledge cue in (9) above. (Recall in this case that adults override the syntactic cue in favor of the pragmatic cue.) However, Forsythe *et al.* (2022) offers some information about how children resolve conflicts between clausal relation cues and other cues to pronoun interpretation. In particular, Forsythe *et al.* (2022) manipulated both pronoun form (null vs. overt) and lexical cues to clausal relations; the clausal relation was signaled either by the lexical items *y después* ('and then'), a temporal connective which favors the subject, or the lexical item *y por eso* ('and for that'), a result connective which doesn't. Children under four and a half paid attention to the contrast between lexical connectives, choosing more subject antecedents in the *y después* condition compared to the *por eso* condition; children over four and a half paid attention to the pronoun form, choosing more subject antecedents in the null pronoun condition compared to the overt pronoun condition. Neither group behaved like adults, who incorporated both contrasts into their pronoun interpretations. This result highlights how children's deployment of representations may change over time. That is, children in the older group presumably have some representation of lexical connectives, given that younger children demonstrate adult-like use of that cue in their pronoun interpretations; however, the older children appear unable to deploy that representation when it conflicts with the cue of pronoun form. Children seem to resolve the conflict by ignoring one piece of information, rather than relying on both pieces of information together like adults do.

Subject-verb agreement. Interestingly, there seems to be an asymmetry between children's early adult-like perception and production of agreement and their apparent inability to use it in comprehension tasks. For example, English-learning children younger than two readily perceive a range of grammatical violations involving the verbal agreement marker /-s/, distinguishing between grammatical and ungrammatical sentence pairs like *A team bake**s**/*bake bread* (Soderstrom *et al.*, 2002) and *A boy does bake/*does bake**s** bread* (Soderstrom, 2002). English-learning children also spontaneously produce verbal /-s/ in over 90% of obligatory contexts by age 2;2-3;10 (Brown, 1973) and reliably produce it in elicited production tasks by age 3;5 (Theakston *et al.*, 2003). However, English-learning children as old as five can fail to use the presence or absence of verbal /-s/ to interpret when the speaker is referring to a singular versus plural subject. For instance, the contrast between *The X swim**s*** and *The X swim* would indicate whether X is a singular or plural subject (e.g., one duck swimming or multiple ducks swimming); yet, in at least two different studies, five-year-olds didn't seem able to make this inference in a picture-selection task (Johnson *et al.*, 2005; Legendre *et al.*, 2014).

Similarly in Spanish, children correctly produce agreement morphology by age two (Clahsen *et al.*, 2002). Yet, in picture-selection tasks using a null subject pronoun, they fail to use third person plural agreement on the verb (e.g., *nada**n*** '(they) swim') to reliably select a plural picture until age three and a half. Likewise, children fail to reliably use third person singular agreement on the verb (e.g., *nada* '(it) swims') to select a singular picture until age five or later (Pérez-Leroux, 2005; Legendre *et al.*, 2014). Similar production-comprehension asymmetries are found in Xhosa (Gxilishe *et al.*, 2009) and Arabic (Rastegar *et al.*, 2012).

Yet there is also evidence that younger children (24 months and 36 months: Kouider *et al.* 2006; Wood *et al.* 2009) can use verbal agreement morphology for comprehension when it's accompanied by other agreement markers, such as those on nouns and quantifiers. For instance, these younger children could choose the correct picture in a preferential looking task for *There is a car* vs. *There are some cars*, with the verbal morphology *is*/*are* is paired with the appropriate noun morphology -*Ø*/-*s* and appropriate quantifier *a*/*some*. Moreover, 24-month-olds (Davies *et al.*, 2017) and 36-month-olds (Kouider *et al.*, 2006) also seem able to use agreement morphology on the noun without other accompanying cues, particularly when the plural morphology variant is more perceptually salient (e.g., voiceless /s/, which has a longer duration) (Davies *et al.*, 2017).

One way to interpret these behavioral asymmetries is that children's representation of (some) agreement morphology (e.g., verbal agreement morphology) matures earlier than their ability to deploy those representations in real time comprehension. For example, English-learning five-year-olds may very well recognize that sentences like **The ducks swim**s*** are ungrammatical because the morphological /-s/ marker on the verb signals singular while the subject is plural; so, in a passive looking task (e.g., the Headturn Preference Procedure used in work by Soderstrom and colleagues (Soderstrom, 2002; Soderstrom *et al.*, 2002, 2007)) children would show a preference for the grammatical version. However, when trying to use this knowledge in real time to infer which of two pictures the speaker is referring to (e.g., one duck versus multiple ducks), children can't deploy it accurately and/or quickly enough to select the target picture. Similarly, Spanish-learning five-year-olds may recognize that the singular agreement on *nada* is appropriate for a singular subject, but like the English five-year-olds, they might not deploy it accurately enough to select a picture with a single duck swimming.

In support of the idea that children inaccurately deploy the information from agreement cues (rather than having an inaccurate representation of what this cue signals), studies often find improved performance when experimental task demands are lessened. For example, a passive looking task reveals better performance than a more-demanding looking-and-pointing task among German three-year-olds (Brandt-Kobele & Höhle, 2010). As another example, a picture-selection task with familiar words reveals better performance than one with nonce words among Spanish three- to five-year-olds (González-Gómez *et al.*, 2017).

Conversely, when tasks get harder, children don't perform as well. For instance, recall that children's naturalistic production of agreement morphology is often more accurate than their use of agreement morphology to correctly comprehend an utterance's meaning. Yet, when Verhagen & Blom (2014) used an elicited production task, which is more cognitively-demanding than spontaneous production, they didn't find better production performance (compared with children's comprehension). As another example of a harder task decreasing children's performance with agreement morphology, Forsythe & Schmitt (2021) found that children's performance was worse when the interpretation cues (which included agreement morphology) conflicted; in contrast, when the cues aligned (and so the task was easier), children's performance was better. These examples highlight the role of task difficulty, which can affect children's deployment of relevant information.

Still, while the prior studies suggest that inaccurate deployment is one cause of children's non-adult-like pronoun interpretation, children's representations of some agreement morphology may also be inaccurate. This is because pronoun-interpretation performance also varies based on the semantic and phonological characteristics of different agreement markers (Pérez-Leroux, 2005; Legendre *et al.*, 2010, 2011, 2014; Forsythe & Schmitt, 2021).

B Design of the picture-selection task

Eight distinct experimental items were created by choosing pairs of verbs that were easily depicted and likely to be known by children under age three: *sigue–sube*: 'follow–get up,' *busca–se esconde*: 'seek–hide,' *sigue a X–sigue a Y*: 'follow X–follow Y,' *sigue–sale* 'follow–go out,' *echa porra–salta la cuerda*: 'cheer on–jump rope,' *tapa–se acuesta*: 'cover–sleep,' *canta–saca pastel*: 'sing–take out a cake,' *dice adiós–se va*: 'say goodbye–leave'. In addition, as mentioned in the main text, we masked agreement morphology information that could come from the singular -ø or plural -*s* marking on the subject of the second clause (i.e., *ella(s))* by using /s/-initial predicates (e.g., *sale*). This way, in overt pronoun conditions, number marking on the pronoun subject *ella(s)* ('she/they') wouldn't provide participants with any additional number cues (e.g., "*ella sal...*" is difficult to distinguish from "*ellas sal...*"). Instead, participants would need to rely on the agreement cue provided by the verbal agreement marker (e.g., *sale(n)*). Potential experimental stimuli were constructed as in Table 7.

Below are the filler and experimental items used in the pronoun interpretation experiment. A stimuli set for a participant was constructed via a combination of fillers (16) and experimental items (16), with the fillers interleaved between the experimental items. The 16 experimental items were selected by randomly choosing 4 of the items where all cues favored the subject, the corresponding items where 1 cue disfavored the subject (either pronoun form, discourse connective, or agreement morphology), 4 of the items where all cues favored the object, and the corresponding items where 1 cue disfavored the object (either pronoun form, discourse connective, or agreement morphology).

Fillers

1. Todos saludan. Las niñas saludan a los niños, y los niños saludan a la maestra.
   Everyone greet-3Pl. The girls greet-3Pl A the boys, and the boys greet-3Pl A the teacher.

   TARGET: Boys greeting teacher
   DISTRACTOR: Girls greeting teacher

2. Cuando llegan, está cerrado. Encuentran la llave, y los niños abren la puerta.
   When arrive-3P, is-3Sg closed. find-3Pl the key, and the boys open-3Pl the gate.

Utterance opening (singular subject): "La maestra les dice adiós a las niñas ..."				
Utterance Continuation	Cue instance favors subject or object			Cue behavior
	Form	Connective	Agr Morph	
"...y después ø se va..."	subject	subject	subject	All cues favor the subject
"...y después ella se va..."	object	subject	subject	One cue disfavors the subject
"...porque ø se va..."	subject	object	subject	
"...y después ø se van..."	subject	subject	object	
"...porque ø se van..."	subject	object	object	One cue disfavors the object
"...y después ellas se van..."	object	subject	object	
"...porque ella se va...".	object	object	subject	
"...porque ellas se van..."	object	object	object	All cues favor the object
Utterance opening (plural subject): "Las niñas le dicen adiós a la maestra ..."				
Utterance Continuation	Cue instance favors subject or object			Cue behavior
	Form	Connective	Agr Morph	
"...y después ø se van..."	subject	subject	subject	All cues favor the subject
"...y después ellas se van..."	object	subject	subject	One cue disfavors the subject
"...porque ø se van..."	subject	object	subject	
"...y después ø se va..."	subject	subject	object	
"...porque ø se va..."	subject	object	object	One cue disfavors the object
"...y después ella se va..."	object	subject	object	
"...porque ellas se van..."	object	object	subject	
"...porque ella se va..."	object	object	object	All cues favor the object

Table 7: Experimental item stimuli construction for an example item with the three cues of pronoun form (Form), lexical connective (Connective), and agreement morphology (Agr Morph) on the verb. For instances with a singular subject (top 8 rows), the utterance begins the same way (*La maestra les dice adiòs a las niñas* 'The teacher to-them say-3SG goodbye A the girls'), and then continues in one of eight ways, depending on whether each cue instance favors the subject (*La maestra* 'The teacher') or object (*las niñas* 'the girls') in the previous clause as the pronoun's antecedent. Cues instances that favor the subject are null form ø, connective *y despuès* ('and then'), and singular agreement morphology -ø. Cue instances that favor the object are overt form *ella(s)* ('she'/'they'), connective *porque* ('because'), and plural agreement morphology -*n*. The same structure applies for utterances that begin with a plural subject (bottom 8 rows). Note that agreement morphology on the overt pronoun form (*ella* -ø vs. *ellas* -*s*) is masked because the following word begins with an *s*-.

TARGET: Boys opening park gate
DISTRACTOR: Teacher opening park gate

3. Ahora van a comer. Tienen mucha hambre, y los niños se acaban la comida.
   Now go-3Pl A to-eat. Have-3Pl much hunger, and the boys eat-up-3Pl the food.

   TARGET: Boys finishing their plates
   DISTRACTOR: The girls finishing their plates

4. Vuelven a la escuela. Se sientan en el salón, y los niños enseñan la clase.
   Return-3Pl to the school. Sit-3Pl in the classroom, and the boys teach-3Pl the class.

   TARGET: Boys teaching at a chalkboard
   DISTRACTOR: Teacher teaching at a chalkboard

5. Ay no, el salón quedó sucio. Hay lodo en el piso, y los niños se quedan a limpiar.
   Oh no, the classroom got-3Sg dirty. There-is mud on the floor, and the boys stay-3Pl A to-clean.

   TARGET: Boys cleaning
   DISTRACTOR: Girls cleaning

6. Es la clase de arte. Sacan papel y colores, los niños pintan una flor.
   Be-3Sg the class of art. Take-out-3Pl paper and colors, the boys paint-3Pl a flower.

   TARGET: Boys painting a flower
   DISTRACTOR: Teacher painting a flower

7. Ya acabó la siesta. Se ponen los zapatos, y los niños se peinan.
   Already finished-3Sg the nap. Put-on-3Pl the shoes, and the boys brush-3Pl their hair.

   TARGET: Boys brushing their hair
   DISTRACTOR: Girls brushing their hair

8. Ahora hay que limpiar. Limpian la mesa, y los niños se lavan las manos.
   Now must-be to-clean. Clean-3Pl the table, and the boys wash-3Pl the hands.

   TARGET: Boys washing their hands
   DISTRACTOR: Teacher washing her hands

9. Todos saludan. Las niñas saludan a los niños, y los niños saludan a la maestra.
   All greet-3Pl. The girls greet-3Pl A the boys, and the boys greet-3Pl A the teacher.

   TARGET: Boys waving to teacher
   DISTRACTOR: Teacher waving to boys

10. Camino al parque, toman la comida en un restorán. Comen sopa de fideos, y los
    The-walk to-the park, take-3Pl the food in a restaurant. Eat-3Pl soup of noodles, and the
    niños piden postre.
    boys ask-for-3Pl dessert.

    TARGET: Boys asking for dessert
    DISTRACTOR: Girls asking for dessert

11. Cuando vuelven del parque están embarrados. Se lavan las manos, y los niños
    When return-3Pl from-the park be-3Pl muddy. Wash-3Pl the hands, and the boys
    se lavan la cara.
    wash-3Pl the face

TARGET: Boys washing their faces
DISTRACTOR: Teacher washing her face

12. Ya va a empezar la clase. Entran al salón, y los niños se sientan.
Already go-3Sg A to-start the class. Enter-3Pl to-the living room, and the boys sit-down-3Pl

TARGET: Boys sitting down
DISTRACTOR: Girls sitting down

13. Ay no, se olvidó algo. Dejaron las luces encendidas, y los niños apagan las luces.
Oh no, forgot-3Sg something. Left-3Pl the lights lit, and the boys turn-off-3Pl the lights.

TARGET: Boys turning the lights off
DISTRACTOR: Girls turning the lights off

14. Es la clase de música. Cantan una canción, y los niños bailan.
Be-3Sg the class of music. Sing-3Pl a song, and the boys dance-3Pl

TARGET: Boys dancing
DISTRACTOR: Girls dancing

15. Ya acabó la siesta. Se despiertan, y los niños abren las cortinas.
Already finished-3Sg the party. Wake-up-3Pl, and the boys open-3Pl the curtains

TARGET: Boys opening the curtains
DISTRACTOR: Girls opening the curtains

16. Después de la fiesta hay que ordenar. Guardan los platos, y los niños limpian la mesa.
After of the party must to-put-in-order. Put-away-3Pl the dishes, and the boys clean-3Pl the table

TARGET: Boys cleaning the table
DISTRACTOR: Girls cleaning the table

Experimental item sets

1. **Opening**: Hoy van a pasear al parque.
Today they-go A to-walk at-the park.

Subject-favoring: Las niñas corren tras la maestra ...
The girls run after the teacher ...

- All cues favor subject: ... y después ø suben ...
... and then *pro* get-3Pl ...
- One cue disfavors the subject

- Pronoun form: ... y después ellas suben ...
  ... and then they get-3Pl ...
- Discourse connective: ... porque ø suben ...
  ... because *pro* get-3Pl ...
- Agreement morphology: ... y después ø sube ...
  ... and then *pro* get-3Sg ...

Object-favoring: La maestra corre tras las niñas ...
The teacher run-3Sg after the girls ...

- All cues favor object: ... porque ellas suben ...
  ... because they get-3Pl ...
- One cue disfavors the object
  - Pronoun form: ... porque ø suben ...
    ... because *pro* get-3Pl ...
  - Discourse connective: ... y después ellas suben ...
    ... and then they get-3Pl ...
  - Agreement morphology: ... porque ella sube ...
    ... because she get-3Sg ...

Closing: ... al camión.
... on-the truck.

Picture 1: Girls getting on truck
Picture 2: Teacher getting on truck

2. **Opening**: En el parque, juegan al escondite.
   In the park play-3Pl hide and seek.

Subject-favoring: Las niñas buscan a la maestra ...
The girls look-for-3Pl A the teacher ...

- All cues favor subject: ... y después ø se esconden.
  ... and then *pro* hide-3Pl.
- One cue disfavors the subject
  - Pronoun form: ... y después ellas se esconden.
    ... and then they hide-3Pl.
  - Discourse connective: ... porque ø se esconden.
    ... because *pro* hide-3Pl.
  - Agreement morphology: ... y después ø se esconde.
    ... and then *pro* hide-3Sg.

Object-favoring: La maestra busca a las niñas ...
The teacher look-for-3Sg A the girls ...

- All cues favor object: ... porque ellas se esconden.
  ... because they hide-3Pl.
- One cue disfavors the object
  - Pronoun form: ... porque ø se esconden.
    ... because *pro* hide-3Pl.
  - Discourse connective: ... y después ellas se esconden.
    ... and then they hide-3Pl.
  - Agreement morphology: ... porque ella se esconde.
    ... because she hide-3Sg.

Picture 1: Girls hiding
Picture 2: Teacher hiding

3. **Opening**: Ahora, todos bailan.
   Now, everyone dance-3Pl.

   Subject-favoring: Las niñas siguen a la maestra ...
   The girls follow-3Pl A the teacher ...

   - All cues favor subject: ... y después ø siguen ...
     ... and then *pro* follow-3Pl ...
   - One cue disfavors the subject
     - Pronoun form: ... y después ellas siguen ...
       ... and then they follow-3Pl ...
     - Discourse connective: ... porque ø siguen ...
       ... because *pro* follow-3Pl ...
     - Agreement morphology: ... y después ø sigue ...
       ... and then *pro* follow-3Sg ...

   Object-favoring: La maestra sigue a las niñas ...
   The teacher follow-3Sg A the girls ...

   - All cues favor object: ... porque ellas siguen ...
     ... because they follow-3Pl ...
   - One cue disfavors the object
     - Pronoun form: ... porque ø siguen ...
       ... because *pro* follow-3Pl ...
     - Discourse connective: ... y después ellas siguen ...
       ... and then they follow-3Pl ...
     - Agreement morphology: ... porque ella sigue ...
       ... because she follow-3Sg ...

   Closing: ... al líder.
   ... A-the leader.

Picture 1: Girls in a conga line with a man
Picture 2: Teacher in a conga line with a man

4. **Opening**: Ya terminó la clase.
   Already finished-3Sg the class.

   Subject-favoring: Las niñas siguen a la maestra ...
   The girls follow-3Pl A the teacher ...

   - All cues favor subject: ... y después ø salen ...
     ... and then *pro* go-out-3Pl ...
   - One cue disfavors the subject
     - Pronoun form: ... y después ellas salen ...
       ... and then they go-out-3Pl ...
     - Discourse connective: ... porque ø salen ...
       ... because *pro* go-out-3Pl ...
     - Agreement morphology: ... y después ø sale ...
       ... and then *pro* go-out-3Sg ...

   Object-favoring: La maestra sigue a las niñas ...
   The teacher follow-3Sg A the girls ...

   - All cues favor object: ... porque ellas salen ...
     ... because they go-out-3Pl ...
   - One cue disfavors the object
     - Pronoun form: ... porque ø salen ...
       ... because *pro* go-out-3Pl ...
     - Discourse connective: ... y después ellas salen ...
       ... and then they go-out-3Pl ...
     - Agreement morphology: ... porque ella sale ...
       ... because she go-out-3Sg ...

   Closing: ... al recreo.
   ... to-the playground.

   Picture 1: Girls exiting to playground
   Picture 2: Teacher exiting to playground

5. **Opening**: Afuera, están jugando.
   Outside be-3Pl playing.

   Subject-favoring: Las niñas le echan una porra a la maestra ...
   The girls to-her cheer-on-3Pl A the teacher ...

   - All cues favor subject: ... y después ø saltan ...
     ... and then *pro* jump-3Pl ...

- One cue disfavors the subject
  - Pronoun form: ... y después ellas saltan ...
    ... and then they jump-3Pl ...
  - Discourse connective: ... porque ø saltan ...
    ... because *pro* jump-3Pl ...
  - Agreement morphology: ... y después ø salta ...
    ... and then *pro* jump-3Sg ...

Object-favoring: La maestra les echa una porra a las niñas ...
The teacher to-them throw-3Sg a baton to the girls ...

- All cues favor object: ... porque ellas saltan ...
  ... because they jump-3Pl ...
- One cue disfavors the object
  - Pronoun form: ... porque ø saltan ...
    ... because *pro* jump-3Pl ...
  - Discourse connective: ... y después ellas saltan ...
    ... and then they jump-3Pl ...
  - Agreement morphology: ... porque ella salta ...
    ... because she jump-3Sg ...

Closing: ... la cuerda.
... the rope.

Picture 1: Girls jumping rope
Picture 2: Teacher jumping rope

6. **Opening**: Es la tarde y todos tienen sueño.
Be-3Sg the afternoon and all be-3Pl sleepy.

Subject-favoring: Las niñas tapan a la maestra ...
The girls cover-3Pl A the teacher ...

- All cues favor subject: ... y después ø se acuestan ...
  ... and then *pro* go-3Pl ...
- One cue disfavors the subject
  - Pronoun form: ... y después ellas se acuestan ...
    ... and then they go-3Pl ...
  - Discourse connective: ... porque ø se acuestan ...
    ... because *pro* go-3Pl ...
  - Agreement morphology: ... y después ø se acuesta ...
    ... and then *pro* go-3Sg ...

Object-favoring: La maestra tapa a las niñas ...
The teacher cover-3Sg A the girls ...

- All cues favor object: ... porque ellas se acuestan ...
  ... because they go-3Pl ...
- One cue disfavors the object
  - Pronoun form: ... porque ø se acuestan ...
    ... because *pro* go-3Pl ...
  - Discourse connective: ... y después ellas se acuestan ...
    ... and then they go-3Pl ...
  - Agreement morphology: ... porque ella se acuesta ...
    ... because she go-3Sg ...

Closing: ... a dormir.
... A to-sleep.

Picture 1: Girls going to sleep
Picture 2: Teacher going to sleep

7. **Opening**: Hoy celebran una fiesta.
   Today celebrate-3Pl a party.

   Subject-favoring: Las niñas le cantan a la maestra ...
   The girls to-her sing-3Pl to the teacher ...

   - All cues favor subject: ... y después ø sacan ...
     ... and then *pro* take-out-3Pl ...
   - One cue disfavors the subject
     - Pronoun form: ... y después ellas sacan ...
       ... and then they take-out-3Pl ...
     - Discourse connective: ... porque ø sacan ...
       ... because *pro* take-out-3Pl ...
     - Agreement morphology: ... y después ø saca ...
       ... and then *pro* take-out-3Sg ...

   Object-favoring: La maestra les canta a las niñas ...
   The teacher to-them sing-3Sg to the girls ...

   - All cues favor object: ... porque ellas sacan ...
     ... because they take-out-3Pl ...
   - One cue disfavors the object
     - Pronoun form: ... porque ø sacan ...
       ... because *pro* take-out-3Pl ...
     - Discourse connective: ... y después ellas sacan ...
       ... and then they take-out-3Pl ...
     - Agreement morphology: ... porque ella saca ...
       ... because she take-out-3Sg ...

Closing: ... el pastel del refri.
... the cake from-the fridge.

Picture 1: Girls taking out the cake
Picture 2: Teacher taking out the cake

8. **Opening**: Ya terminó el día.
Already finished-3Sg the day.

Subject-favoring: Las niñas le dicen adiós a la maestra ...
The girls to-her say-3Pl goodbye to the teacher ...

- All cues favor subject: ... y después ø se van ...
  ... and then *pro* go-3Pl ...
- One cue disfavors the subject
  - Pronoun form: ... y después ellas se van ...
    ... and then they go-3Pl ...
  - Discourse connective: ... porque ø se van ...
    ... because *pro* go-3Pl ...
  - Agreement morphology: ... y después ø se va ...
    ... and then *pro* go-3Sg ...

Object-favoring: La maestra les dice adiós a las niñas ...
The teacher to-them say-3Sg goodbye to the girls ...

- All cues favor object: ... porque ellas se van ...
  ... because they go-3Pl ...
- One cue disfavors the object
  - Pronoun form: ... porque ø se van ...
    ... because *pro* go-3Pl ...
  - Discourse connective: ... y después ellas se van ...
    ... and then they go-3Pl ...
  - Agreement morphology: ... porque ella se va ...
    ... because she go-3Sg ...

Closing: ... a la casa.
... to the house.

Picture 1: Girls going to the house
Picture 2: Teacher going to the house

C Statistical model selection for behavioral results

We analyzed the data using mixed-effects logistic regressions with the *lme4* package in R (Bates *et al.*, 2015a). Following a data-driven approach that includes both fixed-effect and random-effect terms if they improve the model fit (Bates *et al.*, 2015b; Matuschek *et al.*, 2017; Sonderegger, 2023), we began with a baseline model structure that included fixed effects of pronoun **form** (null=1, overt=0), **conn**ective (*porque*=0, *y después*=1), agreement **morph**ology (object-aligned=0, subject-aligned=1), and **age** (children $\geq$5=0, adults=1). The baseline model also included by-item and by-participant random intercepts.

We then sequentially added theoretically-motivated terms if they significantly improved fit according to likelihood ratio tests ($\alpha \leq .05$). Because prior studies (English: Johnson *et al.* (2005); Legendre *et al.* (2014); Spanish: Pérez-Leroux (2005); Legendre *et al.* (2014); Xhosa: Gxilishe *et al.* (2009); Arabic: Rastegar *et al.* (2012)) found an asymmetry in the use of morphology number (singular vs. plural), we considered morphology **number** as a potential factor.

Including a main effect of morphology number did not significantly improve fit over the baseline model ($\chi^2(1)$ = 0.0849, $p = .771$). Two-way interactions with age group significantly improved fit over the baseline model ($\chi^2(3)$ = 67.27, $p < .001$). Including morphology number (both as a main effect and all two-way interactions) on top of the model with interactions by age group did not significantly improve fit over the model with two-way interactions with age group ($\chi^2(4)$ = 3.234, $p = .520$), although it did improve fit over the baseline model ($\chi^2(7)$ = 70.503, $p < .001$). Given these analyses, we selected the model with two-way interactions by age group only, to reduce model complexity.

Turning to the random effects, by-participant random slopes for agreement morphology significantly improved fit ($\chi^2(2) = 42.303$, $p < .001$). Random slopes for other predictors either failed to converge or did not improve fit. The final model included random intercepts for participants and items, and by-participant random slopes for agreement morphology, and has the form in (5), repeated below in (10).

(10) Model with age interactions and by-participant random slopes for agreement morphology

$$\text{subj-antecedent} \sim \text{form*age + conn*age + morph*age} \\ + (1|\text{item}) + (1 + \text{morph}|\text{participant})$$

D Behavioral results in adults vs children

Here we provide the mixed-effects logistic regression results for each age group reported in the main text: adults and children age five and older ($\geq$5). As with the model reported in the main text, we analyzed the data using mixed-effects logistic regressions with the *lme4* package in R (Bates *et al.*, 2015a). Following the same data-driven approach that includes both fixed-effect and random-effect terms if they improve the model fit (Bates *et al.*, 2015b; Matuschek *et al.*, 2017; Sonderegger, 2023), we began with a baseline model structure that included fixed effects of pronoun **form** (null=1, overt=0), **conn**ective (*porque*=0, *y después*=1), and agreement **morph**ology (object-aligned=0, subject-aligned=1). The baseline model also included by-item and by-participant random intercepts. We then sequentially added theoretically-motivated terms if they significantly improved fit according to likelihood ratio tests ($\alpha \leq .05$).

D.1 Adults

Including morphology number (either alone or as two-way interactions) led to singularity warnings. Given this, we selected the baseline model's fixed effects. Turning to random effects, by-participant random slopes for agreement morphology significantly improved fit ($\chi^2(2) = 28.504$, $p < .001$). Random slopes for other predictors failed to converge. The final model included random intercepts for participants and items, and by-participant random slopes for agreement morphology, and has the form in (11).

(11) Model with by-participant random slopes for agreement morphology

$$\text{subj-antecedent} \sim \text{form + conn + morph} \\ + (1|\text{item}) + (1 + \text{morph}|\text{participant})$$

Fixed Effects	β	SE	z	p	
Intercept	.006	0.15	.041	.97	
Form	0.45	0.15	3.12	.002	**
Connective	0.61	0.15	4.12	<.001	***
Morphology	1.82	0.20	9.07	<.001	***

Random Effects	Variance	SD
Participant: Intercept	0.03	0.16
Participant: Morphology	0.93	0.97
Item: Intercept	0.03	0.18

Note. N = 736 observations from 47 participants across 8 items.
Random effects correlation between participant intercept and morphology slope = -0.09.

Table 8: Mixed-effects logistic regression results for model `subj-antecedent ~ form + conn + morph + (1|item) + (1 + morph|participant)`, including β coefficients, standard deviation (SE), z-values, and p-values. Significant effects are reported at alpha level $p < 0.05$ *, $p < 0.01$ **, or $p < 0.001$ ***.

The analysis in Table 8 suggests that, for adults, all three cues were significant (pronoun form $p = .002$, discourse connective $p < .001$, agreement morphology $p < .001$). Agreement morphology has the strongest effect (β=1.82).

D.2 Children $\geq$5

Including a main effect of morphology number did not significantly improve fit over the baseline model ($\chi^2(1)$ = 0.1283, $p = .72$). Including two-way interactions with morphology number did significantly improve model fit over the baseline model ($\chi^2(4)$ = 30.92, $p < .001$). Because morphology number might reasonably be thought to interact with agreement morphology, we tested including only an interaction between morphology number and agreement morphology. Including only this interaction improved model fit over the baseline model ($\chi^2(2)$ = 28.54, $p < .001$); notably, a model including all two-way morphology number interactions did not significantly improve fit over a model with only the interaction between morphology number and agreement morphology ($\chi^2(2)$ = 2.44, $p = .29$); Given this, we selected the model that included only the two-way interaction between morphology number and agreement morphology.

Turning to random effects, by-participant random slopes for both discourse connective and agreement morphology significantly improved fit over the baseline model (connective $\chi^2(2) = 7.225$, $p = .027$; morph $\chi^2(2) = 12.40$, $p = .002$). Random slopes for other predictors led to singularity warnings. Notably, the AIC, BIC, and log likelihood scores for the model with by-participant random slopes for agreement morphology were better than the equivalent scores for the model with by-participant random slopes for discourse connective. Given this, we selected a final model that included the interaction between morphology number and agreement morphology, and by-participant random slopes for agreement morphology. This model has the form in (12).

(12) Model with the interaction between morphology number and agreement morphology, and by-participant random slopes for agreement morphology

$$\text{subj-antecedent} \sim \text{form + conn + morph*number} \\ + (1|\text{item}) + (1 + \text{morph}|\text{participant})$$

Fixed Effects	β	SE	z	p	
Intercept	-.38	0.19	-2.03	.042	*
Form	-.17	.14	-1.17	.24	
Connective	.32	0.14	2.23	.026	*
Morphology	.85	0.20	4.32	<.001	***
Number	.11	-.64	.12	.88	
Morphology:Number	-.64	0.13	-5.1	<.001	***

Random Effects	Variance	SD
Participant: Intercept	0.23	0.48
Participant: Morphology	0.57	0.75
Item: Intercept	0.08	0.28

Note. N = 445 observations from 29 participants across 8 items.
Random effects correlation between participant intercept and morphology slope = -0.44.

Table 9: Mixed-effects logistic regression results for model `subj-antecedent ~ form + conn + morph:number + (1|item) + (1 + morph|participant)`, including β coefficients, standard deviation (SE), z-values, and p-values. Significant effects are reported at alpha level $p < 0.05$ *, $p < 0.01$ **, or $p < 0.001$ ***.

The analysis in Table 9 suggests that, for children age five and older, the agreement morphology cue has the strongest effect (β=0.85, $p < .001$), and interacts with morphology number of that cue (β=-.64, $p < .001$). This importance of morphology number aligns with both prior research (Johnson *et al.* (2005); Pérez-Leroux (2005); Gxilishe *et al.* (2009); Rastegar *et al.* (2012); Legendre *et al.* (2014) and the child responses separated out by morphology number (shown in Figure 3). In particular, we see in Figure 3 that child responses involving singular morphology seem to pattern differently than child responses involving plural morphology.

The discourse connective cue also has a significant effect on its own (β=.32, $p = .026$), while the pronoun form cue does not ($p = .24$).

E Corpus data used for modeled listener input

To calculate the prior and likelihood probabilities used as input by the modeled listeners, we examined samples of child-directed speech from the Schmitt-Miller corpus (Miller & Schmitt, 2012). For each cue (pronoun form, discourse connective, and agreement morphology), Table 10 reports the following: (i) number of child-directed utterances analyzed, (ii) number of instances of a particular cue value, and (iii) number of instances of a cue value that correspond to a particular antecedent type.

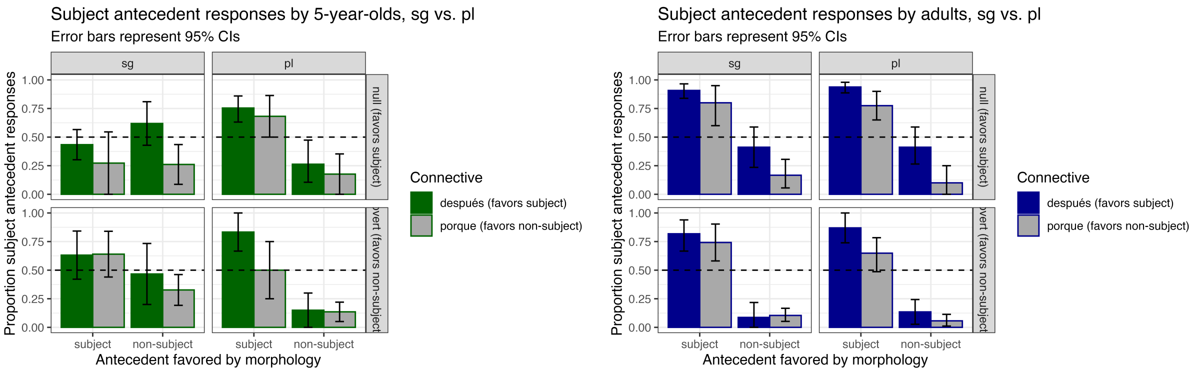


Figure 3: Rate of subject antecedent responses by children (≥5 years old) and adults interpreting Spanish pronouns in utterances where the pronoun form favors the subject (null) or the object (overt), the discourse connective favors the subject (*y después*) or the object (*porque*), agreement morphology (singular or plural) favors the subject or the object, and the morphology number is either singular (sg) or plural (pl). Error bars represent 95% confidence intervals.

For pronoun form (null vs. overt), 7,332 child-directed speech utterances were analyzed. For the purposes of coding whether the subject pronoun referred to the subject antecedent vs. a non-subject antecedent, we followed the methods of Forsythe *et al.* (2022), and included only those instances preceded by at least one other clause in the same speaker turn (2,667 found). A speaker turn was defined as the longest uninterrupted string of speech by the same person. Of these 2,654 null and overt pronoun instances, the following distrubutions were found: (i) antecedents: within the same sentence = 1,341 (.51), extra-sentential = 1,313 (0.49), (ii) person: 981 (0.37) 1st-person, 1,116 (0.42) 2nd-person, and 557 (0.21) 3rd-person, and (iii) gender (of 3rd-person pronouns overtly marked = 35 total): 19 (0.54) feminine, and 16 (0.46) masculine.

For discourse connective, the entire corpus of 65,087 utterances was searched to find all examples of *después* and *porque* (203 found). The subject phrase that immediately followed each connective instance (whether pronoun, noun phrase, or other) was analyzed for its antecedent (subject or non-subject), with the idea that children could learn about how often a discourse connective caused the subsequent subject to refer to the preceding subject, irrespective of the form of the subsequent subject.

For agreement morphology, 13,371 utterances were searched for instances of subject-verb agreement (6,998 found). When a verb was present, its number marking (singular or plural) was compared to the number marking (singular or plural) of the antecedent of the subject. As with the discourse connective, the idea is that children could learn about how often agreement morphology on the subsequent clause's verb aligns with the agreement morphology of the subject's antecedent, irrespective of the form of the subject (i.e., pronoun, noun phrase, or other).

F Inaccurate deployment implementation

Equation 17 shows the implementation of how the modeled listener using selective deployment calculates the posterior probability for a particular experimental item with potential antecedent α. This equation represents a mix of the 16 possibilities for the four information types (prior and three cues) being either used (β) or ignored (1-β).

cue	utt anal	value (instances)	antecedent instances (proportion)
FORM	7,332	null (2,365)	subject: 1,091 (0.46)
			non-subject: 1,274 (0.54)
		overt (289)	subject: 63 (0.22)
			non-subject: 226 (0.78)
CONN	65,087	*después* (54)	subject: 29 (0.54)
			non-subject: 25 (0.46)
		porque (149)	subject: 52 (0.35)
			non-subject: 97 (0.65)
MORPH	13,371	singular (5,662)	singular: 5,655 (>0.99)
			plural: 7 (<0.01)
		plural (1,336)	singular: 9 (<0.01)
			plural: 1,327 (>0.99)

Table 10: Rates of reference to different antecedent types in the presence of different CONNectives, pronoun FORMs, and agreement MORPHology in child-directed Spanish speech samples from the Schmitt-Miller corpus. For each cue type, the number of utterances analyzed (utt anal) is shown, along with the number of subject pronoun instances with a particular cue value found, and both the number and proportion of reference to different antecedent types (antecedent instances).

(17)

$$\begin{aligned}
&p_\beta(\alpha|\text{FORM}, \text{CON}, \text{MOR}, \alpha_{num}, \alpha_{subj?}) = \\
&(\beta_{form})(\beta_{con})(\beta_{mor})(\beta_\alpha) \cdot p(\alpha|\text{FORM}, \text{CON}, \text{MOR}, \alpha_{num}, \alpha_{subj?}) + \\
&(\beta_{form})(\beta_{con})(\beta_{mor})(1-\beta_\alpha) \cdot p_{\text{UNIF}}(\alpha|\text{FORM}, \text{CON}, \text{MOR}, \alpha_{num}, \alpha_{subj?}) + \\
&(1-\beta_{form})(\beta_{con})(\beta_{mor})(\beta_\alpha) \cdot p(\alpha|\text{CON}, \text{MOR}, \alpha_{num}, \alpha_{subj?}) + \\
&(1-\beta_{form})(\beta_{con})(\beta_{mor})(1-\beta_\alpha) \cdot p_{\text{UNIF}}(\alpha|\text{CON}, \text{MOR}, \alpha_{num}, \alpha_{subj?}) + \\
&(\beta_{form})(1-\beta_{con})(\beta_{mor})(\beta_\alpha) \cdot p(\alpha|\text{FORM}, \text{MOR}, \alpha_{num}, \alpha_{subj?}) + \\
&(\beta_{form})(1-\beta_{con})(\beta_{mor})(1-\beta_\alpha) \cdot p_{\text{UNIF}}(\alpha|\text{FORM}, \text{MOR}, \alpha_{num}, \alpha_{subj?}) + \\
&(\beta_{form})(\beta_{con})(1-\beta_{mor})(\beta_\alpha) \cdot p(\alpha|\text{FORM}, \text{CON}, \alpha_{num}, \alpha_{subj?}) + \\
&(\beta_{form})(\beta_{con})(1-\beta_{mor})(1-\beta_\alpha) \cdot p_{\text{UNIF}}(\alpha|\text{FORM}, \text{CON}, \alpha_{num}, \alpha_{subj?}) + \\
&(1-\beta_{form})(1-\beta_{con})(\beta_{mor})(\beta_\alpha) \cdot p(\alpha|\text{MOR}, \alpha_{num}, \alpha_{subj?}) + \\
&(1-\beta_{form})(1-\beta_{con})(\beta_{mor})(1-\beta_\alpha) \cdot p_{\text{UNIF}}(\alpha|\text{MOR}, \alpha_{num}, \alpha_{subj?}) + \\
&(\beta_{form})(1-\beta_{con})(1-\beta_{mor})(\beta_\alpha) \cdot p(\alpha|\text{FORM}, \alpha_{num}, \alpha_{subj?}) + \\
&(\beta_{form})(1-\beta_{con})(1-\beta_{mor})(1-\beta_\alpha) \cdot p_{\text{UNIF}}(\alpha|\text{FORM}, \alpha_{num}, \alpha_{subj?}) + \\
&(1-\beta_{form})(\beta_{con})(1-\beta_{mor})(\beta_\alpha) \cdot p(\alpha|\text{CON}, \alpha_{num}, \alpha_{subj?}) + \\
&(1-\beta_{form})(\beta_{con})(1-\beta_{mor})(1-\beta_\alpha) \cdot p_{\text{UNIF}}(\alpha|\text{CON}, \alpha_{num}, \alpha_{subj?}) + \\
&(1-\beta_{form})(1-\beta_{con})(1-\beta_{mor})(\beta_\alpha) \cdot p(\alpha_{num}, \alpha_{subj?}) + \\
&(1-\beta_{form})(1-\beta_{con})(1-\beta_{mor})(1-\beta_\alpha) \cdot p(\text{UNIF})
\end{aligned}$$

G Best-fitting parameter values for all modeled listeners

Table A1 shows the parameter values for the best-fitting version of each modeled listener type (inaccurate representation, inaccurate deployment, both inaccurate) for each age group whose behavioral results were presented in the main text (≥5 children, adults).

	σ_{for}	σ_{con}	σ_{mor}	σ_{α}	β_{for}	β_{con}	β_{mor}	β_{α}
inaccurate representations								
≥5	0.02	0.28	0.11	0.00	1	1	1	1
adults	0.25	0.33	0.28	0.00	1	1	1	1
inaccurate deployment								
≥5	1	1	1	1	0.00	0.43	0.30	0.00
adults	1	1	1	1	0.24	0.65	0.67	0.00
both inaccurate								
≥5	0.0 *any*	4.00	0.22	0.0 *any*	*any* 0.00	0.25	0.61	*any* 0.00
adults	0.33	0.66	0.29	0.0 *any*	0.78	0.52	1.00	*any* 0.00

Table A1: Best-fitting parameter values for all modeled listeners for age groups with behavioral data presented in the main text. Inaccurate representation listeners have contrast parameter σ, inaccurate deployment listeners have use parameter β, and listeners with both inaccurate representations and inaccurate deployment have both contrast σ and use β parameters. Modeled listeners with accurate deployment have β=1, while modeled listeners with accurate representations have σ=1. For the both-inaccurate type, if σ_x=0.00, any β_x value can be used, and is listed as *any*; if β_x=0.00, any σ_x value can be used, and is listed as *any*.