# Inaccurate representations, inaccurate deployment, or both? Using computational cognitive modeling to investigate the development 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 an inaccurate representation of relevant information, inaccurate deployment of accurate representations, or both inaccurate representations and inaccurate deployment. We show how computational cognitive modeling can be used to identify which options could lead to non-adult-like language behavior, using the case study of Spanish subject pronoun interpretation by typically-developing preschoolers. In a picture-selection task, children interpret subject pronouns differently from adults; modeling results suggest that children are likely either always deploying inaccurate representations of their input, or selectively deploying accurate representations, but not selectively deploying inaccurate representations. In contrast, adults are likely always deploying inaccurate representations. So, to become adult-like, children need to learn how to be inaccurate in the right ways for potentially both representation and deployment of pronoun information. We discuss the promise and the 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, knowledge deployment, knowledge representations, picture-selection task, pronoun interpretation, Spanish**

## 1 Introduction

When children produce and interpret language differently from adults, the underlying cause can be unclear: do they have an immature representation of the target language, or do they simply deploy that representation in an immature way—or perhaps some of both? A traditional way to try separating out the effects of immature deployment is to use experimental designs and techniques that facilitate language deployment. If non-adult 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 than have yet to be discovered–so it's difficult 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.

Here, we show how computational cognitive modeling can be fruitfully used to analyze children's non-adult-like behavior, focusing on the case study of Spanish subject pronoun interpretation by typically-developing preschoolers. We first present a pronoun interpretation task showing that three-, four-, and five-year-old 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 at each age. More specifically, we investigate whether the patterns of behavior observed at ages three, four, and five are more likely caused by children (i) inaccurately representing information available for pronoun interpretation, (ii) deploying accurate representations inaccurately, or (iii) both inaccurately representing pronoun information and inaccurately deploying those inaccurate 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 information and for deployment of 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.

Our modeling results suggest that children's behavior is best captured if children are doing one of two things: (i) they always deploy inaccurate representations of their input, or (ii) they selectively deploy accurate representations. In contrast, adult behavior is best captured if adults are always deploying inaccurate 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 pronoun information.

More generally, this case study demonstrates how computational cognitive modeling complements existing behavioral techniques investigating language development: 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 the 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: (i) the pronoun's form, (ii) the semantic relation between the pronoun's clause and other clauses in the surrounding discourse, and (iii) the pronoun's grammatical features. 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 is more strongly associated than the overt form with the preceding subject antecedent (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(Ø→*Juan*) > p(*él*→*Juan*)
Juan called A Pedro when *pro*/he was in house.
'Juan called Pedro when (he) was at home.'

For semantic 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). 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).

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

a. ...**y después** Ø se va. p(*y después*→*Juan*) > p(*y después*→*Pedro*)
...and then *pro* leaves.

b. ...**porque** Ø se va. p(*porque*→*Pedro*) > p(*porque*→*Juan*)
...because *pro* leaves.

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

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'). In contrast to the probabilistic cues of pronoun form and lexical connective, number agreement is categorical when it comes to its effects on pronoun interpretation. In (3), 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 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 categorical cue of agreement morphology forces the interpretation to be the object antecedent (*la maestra*, 'the teacher').

(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.'

This last example demonstrates one of the main tasks children must learn to solve when interpreting pronouns: what to do when multiple cues are available for how to interpret the pronoun, and those cues conflict. In particular, children must learn two key things. First, they must learn the information each cue carries – we refer to this as the pronoun cue's *representation*. For example, the representation of a Spanish null subject pronoun encodes a bias towards subject antecedents (i.e., p(Ø→subject) > p(overt pronoun *él*→subject)). Second, children must learn how to integrate the information from different cues – 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. It's currently unknown when Spanish-learning children acquire adult-like pronoun cue representations and when they achieve adult-like deployment of these representations.

## 2.2 What we know about when children learn what

All three information sources illustrated above – pronoun form, semantic relations, and grammatical features – have been studied to different degrees on their own. 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, inter-clausal semantic relations, and the grammatical features 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 (4), children begin displaying different antecedent preferences for null and overt pronouns by age four and a half to six in Spanish (Forsythe *et al.*, in press) and by age six to seven in Greek (Papadopoulou *et al.*, 2015). That is, when asked to interpret sentences like (4), 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 real time.

(4) 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 (ii) Pedro leaving

However, in felicity judgment tasks like the one in (5), adult-like preferences are much slower to develop (Shin & Cairns, 2012; Sorace *et al.*, 2009). For instance, in (5), 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 (5a) – María is. Therefore, it's inappropriate for Puppet A to use a null pronoun in (5b), since the null pronoun is biased towards María; so, the child should choose the description provided by Puppet B in (5c). Shin & Cairns (2012) report that up until age eight, children acquiring Mexican Spanish failed to reliably choose the overt form in contexts like (5). 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.

(5) 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).

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 (5) seem much more cognitively demanding than pronoun interpretation tasks like (4); 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 (4) 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).

**Inter-clausal semantic relations.** Pronoun interpretation can also be impacted by the semantic relations between the clause containing the pronoun and other clauses in the discourse. These relations are sometimes called *Coherence Relations* (Kehler, 2002; Kehler *et al.*, 2008) or *Rhetorical Relations* (Asher & Lascarides, 2003), and they can be signaled by a variety of syntactic and lexical cues. When these cues change, the underlying semantic relation between clauses also changes, potentially triggering a change in the preferred pronoun interpretation.

For example, narrative semantic 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 semantic 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 semantic information of event sequences.

Syntactic structure can also serve as an inter-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 (6). 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.

(6) 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). An example is in (7) (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 (7a), 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 (7b), 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*.

(7) 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.

Preschool children can leverage these lexical and syntactic cues to interpret pronouns. For inter-clausal semantic 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 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).

So, children can apply adult-like pronoun interpretation strategies under a variety of discourse conditions. There is less evidence about what they do when inter-clausal discourse cues conflict with each other, as with the syntactic parallel structure cue and the pragmatic real world knowledge cue in (7) above. (Recall in this case that adults override the syntactic cue in favor of the pragmatic cue.) However, Forsythe *et al.* (in press) offers some information about how children resolve conflicts between inter-clausal discourse cues and other cues to pronoun interpretation. In particular, Forsythe *et al.* (in press) manipulated both pronoun form (null vs. overt) and lexical cues to inter-clausal relations; the inter-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, they 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). They 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 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 (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 (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).

One way to view these asymmetries is that children's representations of (some) agreement morphology are fairly adult-like by age two, but their ability to deploy these representations in real time to aid comprehension is still immature. 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 reliable than children's 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 (in press) 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, in press).

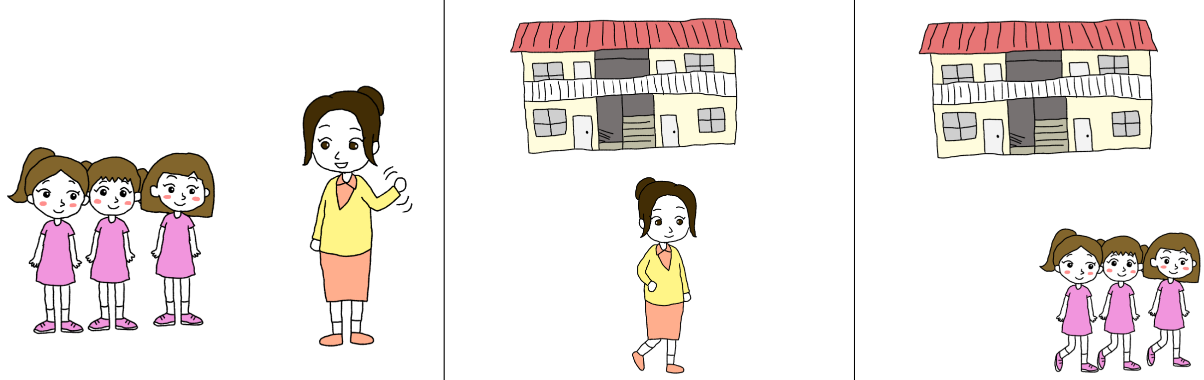
**Summary of what pronoun interpretation cues children know when.** The evidence reviewed above collectively suggests that by age four or five, children are capable of using a variety of cues to interpret pronouns in an adult-like way, including the pronoun's form, the presence of various lexical and syntactic cues that reveal its semantic relation with surrounding clauses, and the 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. This kind of variation in 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 singular and plural 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. 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 or overt), 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 test how strongly each cue affects pronoun interpretation, cues were alternately aligned with each other, as in (8a), or set up so that one was pitted against the others, as in (8b)-(8d).

(8) Utterance opening:



(a) Screen 1:
*La maestra saluda a las niñas...*
The teacher greets A the girls...

(b) Screen 2:
*...y después ø se va*
...and then *pro* leave-3Sg.

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

"La maestra saluda a las niñas ..."
The teacher greets A the girls ...

a. Form → subject, Connective → subject, Agreement Morphology → subject
"...y después ø sale."
...and then *pro* leave-3Sg

b. *Form → object*, Connective → subject, Agreement Morphology → subject
...y después **ella** sale.
...and then **she** leave-3Sg.

c. Form → subject, *Connective → object*, Agreement Morphology → subject
"...**porque** ø sale."
...**because** *pro* leave-3Sg.

d. Form → subject, Connective → subject, *Agreement Morphology → object*
"...y después ø sale**n**."
...and then *pro* leave-**3Pl**.

Recall that two of these cues (pronoun form, lexical connective) are probabilistic, while agreement morphology is categorical. 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 (8a), where all cues signal the subject, and (8d), where agreement morphology signals the object in opposition to the other two cues. However, we would expect a much smaller difference between (8a) and (8b-8c), where only the lexical connective and the pronominal form signal the object, respectively. How and when children differ from adults in these conditions can help us identify which cues children may be heeding at different ages.

### 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. Children were divided into three age groups for analysis: 33 children age three and under ($\leq$3: 1;11-3;10; *M* = 3;3), 35 children age four (4: 4;0-4;11; *M* = 4;5), and 29 children age five and older ($\geq$5: 5;0-6;9; M = 5;8).

### 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 object). That is, in addition to the condition in (8a), where all three cues favor the subject interpretation, and the three conditions in (8b)-(8d) where two of three cues favor the subject, there were also three conditions where two of three cues favored the object, and one condition where all three cues favored the object. 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'). Following Johnson *et al.* (2005), we used /s/-initial predicates (e.g., *sale*) to mask plural /s/ marking on the subject of the second clause; 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)*). 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 A 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 8a) and 8 "incongruent" trials in which agreement morphology favored the opposite antecedent that the pronoun form and connective favored (as in 8d). The same protocol was used for participants assigned to the "pronoun form" and "connective" versions (pronoun form: 8a vs. 8b; connective: 8a vs. 8c).

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 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

Participant performance on both the practice and filler trials suggested that the participants in all age groups understood the task and were capable of completing it as intended. For the practice trials, mean scores in each age group were above chance ($\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, recall that the second clause had an overt description of the subject (i.e., "the boys") rather than a pronoun; so, the target picture corresponding to the correct interpretation showed the individuals described (i.e., the boys). As with the practice trials, mean scores of target picture selection for filler trials in each age group were above chance ($\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).

### 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.

In general, children in all three age groups 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 of all ages prefer the subject less strongly; when adults strongly prefer the object antecedent, children of all ages prefer the object less strongly. So, even up to age five, children appear to have immature pronoun interpretations, whether this is due to 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, a logistic regression was used with one main effect each for pronoun form (null=1, overt =0), connective (*y después*=1, *porque*=0), and agreement morphology

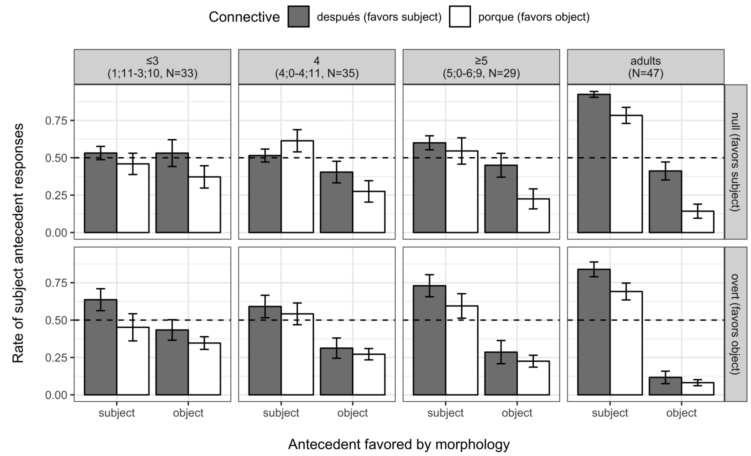



Figure 2: Rate of subject antecedent responses by children (ages ≤3 years old, 4 years old, ≥5 years old) and adults interpreting Spanish pronouns in utterances where the pronoun form favors the subject (null) or the object (overt), the discourse connective favors the subject (*y después*) or the object (*porque*), and disambiguating agreement morphology favors the subject or the object. Error bars represent +/-1SE participant means.

(subject-disambiguated=1, object-disambiguated=0) within each age group. The maximal random effects structure that produced convergence was used (see Table 1).[1]

| age group | form | connective | agreement | random effects |
|---|---|---|---|---|
| ≤ 3 | -0.01 (0.21) | **0.52 (0.22)*** | *0.41 (0.22)*$^{(0.06)}$ | (1\|item)+(1+agr+conn\|ptcpt) |
| 4 | 0.03 (0.20) | 0.08 (0.20) | **1.00 (0.20)***** | (1\|item)+(1\|ptcpt) |
| ≥ 5 | -0.18 (0.28) | **0.63 (0.28)*** | **1.52 (0.43)***** | (1\|item)+(1+agr\|ptcpt) |
| adults | **0.93 (0.27)***** | **1.81 (0.43)***** | **4.05 (0.48)***** | (1\|item)+(1+agr+conn\|ptcpt) |

Table 1: $\beta$ coefficients (SE) and maximal converging random effects structure for a binary logistic regression model of the form `subj.antecedent~form + connective + agreement`. Bold indicates a significant effect (at alpha level $p<0.05$ *, $p<0.01$ **, or $p<0.001$ ***); italics indicate a marginally significant effect ($p<0.1$).

This analysis suggests that children age three and younger (≤3) were influenced by discourse connectives, but not necessarily by pronoun form or agreement morphology. Children age four (4) were influenced by agreement morphology but not by pronoun form or discourse connectives. Children age five and older (≥5) were influenced by both agreement morphology and discourse connectives, but not by pronoun form. In contrast to all children, adults were influenced by all three cues.

It also seems that children rely on the cues they use (connectives and agreement morphology) less than adults do, as indicated by the estimated coefficients. For instance, if we consider connectives, the estimated coefficient for adults using the connective cue is 1.81 (SE = 0.43); this means

[1]Interactions were excluded from this analysis for two reasons. First, interactions increase the complexity of the model, requiring higher power than available within each age group. Second, we focused on whether a given cue has any effect at all, regardless of whether that effect is modulated by an interaction.

that the subject interpretation is $e^{1.81}$=6.1 times more likely when the subject-favoring connective *y después* is used compared to when the object-favoring connective *porque* is used. In contrast, for children, the subject interpretation is only $e^{0.52}$=1.7 (age $\leq$3) or $e^{0.63}$=1.9 (age $\geq$5) times more likely when the subject-favoring connective is used. So, these results suggest that connectives can influence children's pronoun interpretations in the same direction as adults, just not to the same degree.

We see the same pattern for agreement morphology: the estimated coefficient for adults is 4.05 (SE = 0.48) vs. children's estimated coefficient of 1.00-1.52 (SE = 0.20-0.43). So, for adults, the subject interpretation is $e^{4.05}$=57.4 times more likely when subject-disambiguating agreement morphology is used; for children, the subject interpretation is only $e^{1.00}$=2.7 (age 4) or $e^{1.52}$=4.6 times (age $\geq$5) more likely when subject-disambiguating morphology is used. Taken together, this analysis suggests that child pronoun interpretations can be more strongly influenced by agreement morphology than by discourse connectives, but not nearly to the same degree as adult pronoun interpretations.

## 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: pronoun form, discourse connectives, and agreement morphology. Interestingly, children in the age groups tested here don't seem to rely on pronoun form, 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.*, in press) 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 at some ages ($\leq$3 and $\geq$5, but not age 4) do seem to rely on the discourse connective to guide their pronoun interpretation, and they do so in the same direction as adults (e.g., favoring the subject when *y después* is used). These results align with Forsythe *et al.* (in press), who also found sensitivity to the cue of discourse connectives in children younger than four and a half – in particular, younger children use that cue while older children appear to ignore it. We speculate that this may be due to cue integration in both cases (i.e., a deployment issue). Here, children seem to temporarily stop relying on discourse connectives when they begin to rely on agreement morphology at age four; in Forsythe *et al.* (in press), older children stopped relying on discourse connectives when they began relying on pronoun form at age four and a half.

For agreement morphology, children's reliance on this cue seems to grow steadily over time, from not relying on it at age three, to relying on it more and more at ages four and five. Still, children don't seem to rely on agreement morphology 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 agreement morphology. One possibility is that children have an inaccurate representation of agreement morphology. For example, perhaps children age three and younger view subject-favoring morphology as being less indicative of the subject antecedent than it truly is. That is, these children view subject-favoring morphology as more probabilistic than categorical. Suppose also that these children believe that subject-favoring morphology favors the subject less than other object-favoring cues favor the object (like the pronoun form and discourse connective); then, when agreement morphology signals the subject antecedent while the pronoun form or discourse connective signals the object antecedent, these children would favor the object antecedent (unlike adults). This non-adult-like behavior would be due to the inaccurate representation of agreement morphology information.

Another possibility for children's non-adult-like behavior with agreement morphology is that they deploy its information inaccurately. For instance, perhaps children age three and younger fail to perceive the verb's ending in the moment (e.g., singular *sale* vs. plural *salen*), and so lose access to the agreement morphology indicating whether the pronoun is singular or plural. Or, perhaps these children correctly perceive the agreement morphology, but the number information it carries decays too much in short-term memory before they can use it. If the information provided by agreement morphology is lost, then it can't be deployed in the moment and integrated with the information from other cues. So, a child age three and younger could have an adult-like representation of agreement morphology, where subject-favoring morphology categorically 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 object antecedent, despite having an adult-like representation of subject-favoring agreement morphology.

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, in press). Here, we can use computational cognitive modeling to simulate how listener 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 information for interpreting pronouns on the basis of the input encountered (i.e., the representation of pronoun information), as well as how they use those cue representations in real time to identify a pronoun's interpretation in context (i.e., the deployment of those representations).

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 development; 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 (i.e., the modeled listener will choose whether the pronoun refers to the subject antecedent or the object antecedent for a particular experimental item).

When making this inference, the modeled listener will (i) use either accurate or inaccurate representations of the available information, and (ii) either accurately or inaccurately deploy those representations. In particular, we compare four versions of this modeled listener: (i) a **baseline** model with accurate representations and accurate deployment, (ii) an **inaccurate representation** model with inaccurate representations but accurate deployment, (iii) an **inaccurate deployment** model with accurate representations but inaccurate deployment, and (iv) a model with **both inaccurate** representations and inaccurate deployment.[2]

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 child pronoun interpretation behavior for children age three and younger ($\leq$3), four (4), and five and older ($\geq$5), compared with adult controls.

[2]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.

## 3.1 Bayesian inference for pronoun interpretation

A Bayesian modeled listener calculates the probability of a hypothesis $h \in H$, given some data $D$ (the posterior probability $p(h|D)$), as shown in (9). 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 $p(D|h)$).

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

For pronoun interpretation, the hypotheses might be that the pronoun refers to each of the available antecedents $\alpha$ 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 ($H = \{\alpha_{subj.\text{SG}}, \alpha_{obj.\text{PL}}\}$. 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 $p(\alpha_{obj.\text{PL}}|D)$=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% of the time and the object 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∈{ø, overt}, discourse connective CON∈{*y después*, *porque*}, and agreement morphology MOR∈{SG, PL}). So, we can calculate the posterior $p(\alpha_{subj.\text{SG}}|\ \text{FORM}, \text{CON}, \text{MOR})$, as in (10). If we assume each cue is independent[3], then the likelihood calculation can be further divided as in (11).

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

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

In a specific experimental context, the pronoun 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 (8), with a singular subject antecedent (*la maestra*) and a plural object antecedent (*las niñas*). Suppose then that the speaker uses the cues in (8a): a null pronoun (FORM=ø), the discourse connective *y después* (CON=*y después*), and singular agreement morphology (MOR=SG). The likelihoods then reflect 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) suggests the following values: $p(\text{FORM=ø}|\alpha_{subj.\text{SG}})$=0.938, $p(\text{CON=}\textit{y después}|\alpha_{subj.\text{SG}})$=0.324, and $p(\text{MOR=SG}|\alpha_{subj.\text{SG}})$=0.998.

[3] We discuss the implications of this idealizing assumption in section 4.2.

To calculate the probability that the pronoun refers to the subject versus the object antecedent in this specific experimental context, we use (11) to calculate the posteriors: $p(\alpha_{subj.\text{SG}}|\text{FORM},$ $\text{CON}, \text{MOR})$ for the singular subject antecedent and $p(\alpha_{obj.\text{PL}}|\text{FORM}, \text{CON}, \text{MOR})$ for the plural object antecedent. Let's again consider the experimental context of (8a). To calculate the posterior probability $p(\alpha_{subj.\text{SG}}|\text{FORM=}\emptyset, \text{CON=}$*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 (11) (repeated below as (12)), 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 (8a) was the plural object *las niñas*. Our corpus analysis of child-directed speech suggests the prior is $p(\alpha_{obj.\text{PL}})$=0.129, and the likelihoods are $p(\text{FORM=}\emptyset|\alpha_{obj.\text{PL}})$=0.959, $p(\text{CON=}$*y después*$|\alpha_{obj.\text{PL}})$=0.394, and $p(\text{MOR=SG}|\alpha_{obj.\text{PL}})$=0.005. We then arrive at the posterior in (13), which is even lower.

(12)
$$\begin{aligned} p(\alpha_{subj.\text{SG}}|\text{FORM}, \text{CON}, \text{MOR}) &\propto p(\alpha_{subj.\text{SG}}) \cdot p(\text{FORM}|\alpha_{subj.\text{SG}}) \cdot p(\text{CON}|\alpha_{subj.\text{SG}}) \cdot p(\text{MOR}|\alpha_{subj.\text{SG}}) \\ &\propto 0.362 \cdot 0.938 \cdot 0.324 \cdot 0.998 \\ &\propto 0.110 \end{aligned}$$

(13)
$$\begin{aligned} p(\alpha_{obj.\text{PL}}|\text{FORM}, \text{CON}, \text{MOR}) &\propto p(\alpha_{obj.\text{PL}}) \cdot p(\text{FORM}|\alpha_{obj.\text{PL}}) \cdot p(\text{CON}|\alpha_{obj.\text{PL}}) \cdot p(\text{MOR}|\alpha_{obj.\text{PL}}) \\ &\propto 0.129 \cdot 0.959 \cdot 0.394 \cdot 0.005 \\ &\propto 0.000244 \end{aligned}$$

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 listener model. It assumes accurate representation of the information about the pronoun form, connective, and agreement morphology, as based on the likelihoods, as well as accurate representation of antecedent information, based on the prior for that antecedent type; this implementation also assumes accurate deployment of those 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.[4]

[4]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.

## 3.2 Modeled listener input

To estimate the input probabilities used for the priors and likelihoods in our modeled Bayesian listeners, we used 54,757 utterances of child-directed speech from the Schmitt–Miller corpus (Miller & Schmitt, 2012). These data were taken from spontaneous interactions between caretakers and their children (ages 1;6-5;11) born and raised in Mexico City, Mexico. Caretakers and children were recorded during two to four free-play sessions lasting around 30 minutes each, as well as one approximately 30-minute session chatting with another adult. These probabilities are shown in Table 2, 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).

| | | prior | likelihoods | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | $p(\alpha)$ | $p(\text{FORM}\|\alpha)$ | | $p(\text{CON}\|\alpha)$ | | $p(\text{MOR}\|\alpha)$ | |
| antecedent type | | | ø | overt | *y después* | *porque* | 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 2: Priors and likelihoods for different types of antecedents (subject (SUBJ) or non-subject (¬SUBJ), singular (SG) or plural (PL)) estimated from the prononun information distributions in naturalistic child-directed speech.

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).

## 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 representations only, inaccurate deployment only, or both inaccurate 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 information available (about cues or antecedents), or noise in the modeled listener's ability to reliably use that information in novel situations, such as an experimental task. So, the baseline Bayesian listener is adapted to incorporate noise in the representations (inaccurate representations), noise in the integration of information (inaccurate deployment), or both.[5]

### 3.3.1 Baseline Bayesian listener incorporating antecedent type

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

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

Because priors and likelihoods vary depending on the number and position of the antecedent under consideration, (14) 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 ($\alpha_{num}\in\{\text{SG}, \text{PL}\}$), 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 $\alpha_{num}$=SG and $\alpha_{subj?}$=SUBJ. A plural object antecedent ($\alpha_{obj.\text{PL}}$) would have $\alpha_{num}$=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 2, which also varied by these properties of the 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 2. 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 representations to do so. These representations correspond to the modeled listener's priors and the 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 priors.** To incorporate noise into the priors, our modeled listener uses the softmax function ($e^{\sigma ln(prior)}$=$prior^{\sigma}$) as shown in (15), where $\sigma$ serves as a "contrast" parameter. When $\sigma$=1, probabilities are kept the same (i.e., equal to the values in Table 2); 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,

[5]Python code implementation of all modeled learners described here is available at https://github.com/lisapearl/pronoun-interpretation.

our modeled learner can decrease the contrast between relative probabilities (e.g., using $\sigma$=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 $\sigma$=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 $0.0 \leq \sigma \leq 4.0$. The prior has a single contrast parameter value ($\sigma_\alpha$), which holds for all antecedents (i.e., $\sigma_\alpha$ is the same for all $\alpha_{num}$ 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.

**Inaccurate likelihoods.** The same approach can be used to incorporate noise into the likelihoods, using the softmax function ($e^{\sigma ln(likelihood)} = likelihood^{\sigma}$) as shown in (15), where $\sigma$ again serves as a contrast parameter. We investigate contrast parameter values $0.0 \leq \sigma \leq 4.0$. Each pronoun cue has its own contrast parameter value (i.e., $\sigma_{form}$, $\sigma_{con}$, $\sigma_{mor}$), representing (potentially) different noise levels associated with each cue representation.

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

We can then determine the contrast parameter values that allow the modeled Bayesian listener to best match the child and adult control behavior from our experiment. By doing so, we can get a sense of how the true representations would need to be distorted in order for this modeled listener to reproduce the kind of non-adult-like behavior we observe at different ages. We note that if the best-fitting value is $\sigma$=1 for any information type, then this implies that the accurate representation for that 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 representations. More generally, we can also determine if a better fit occurs when relative probability differences are decreased ($\sigma$<1), increased ($\sigma$>1), or left undistorted ($\sigma$=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 of priors.** Our implementation of a listener with inaccurate deployment involves the listener having accurate representations, but not always using them. For information about the priors over possible antecedents, when the listener calculates the posterior probabilities, the listener has available an accurate representation of possible antecedents with the corresponding number and position information (i.e., the prior). However, the listener simply "misses" the 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 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 $p_{\text{UNIF}}$). An example is shown in (16) 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.

(16)
$$\begin{aligned} p_{\text{UNIF}}(\alpha_{num}, \alpha_{subj?}|\text{FORM}, \text{CON}, \text{MOR}) \propto\ & p(\text{UNIF}) \\ & \cdot p(\text{FORM}|\alpha_{num}, \alpha_{subj?}) \\ & \cdot p(\text{CON}|\alpha_{num}, \alpha_{subj?}) \\ & \cdot p(\text{MOR}|\alpha_{num}, \alpha_{subj?}) \end{aligned}$$

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 information coming from the pronoun form, the discourse connective, and the agreement morphology, represented by the likelihoods.

**Inaccurate deployment of likelihoods.** Similarly, our implementation of a listener with inaccurate cue deployment involves the listener calculating the posterior probabilities while having accurate cue representations (i.e., likelihoods) available. However, the listener simply "misses" the 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 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 (17), where only the information from the pronoun form and the discourse connective are used, along with the information from the prior.

(17)
$$\begin{aligned} p(\alpha_{num}, \alpha_{subj?}|\text{FORM}, \text{CON}) \propto\ & p(\alpha_{num}, \alpha_{subj?}) \\ & \cdot p(\text{FORM}|\alpha_{num}, \alpha_{subj?}) \\ & \cdot p(\text{CON}|\alpha_{num}, \alpha_{subj?}) \end{aligned}$$

In this way, the accurate morphology cue information 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 information (and may even use it for other experimental items). In this way, inaccurate deployment is again implemented via a selective dropout of information, this time from one or more cues.

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

Because the prior could be used ($\beta_\alpha$) or ignored (1-$\beta_\alpha$), 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 is used ($\beta_{form}$) or ignored (1-$\beta_{form}$); the discourse connective is used ($\beta_{con}$) or ignored (1-$\beta_{con}$); the agreement morphology is used ($\beta_{mor}$) or ignored (1-$\beta_{mor}$). So, the posterior for a particular experimental item is a mix of these 2 x 8 possibilities, as in (18). 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.

(18)
$$\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}$$

We can then determine the use parameter $\beta$ 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 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 $\beta$=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 adults and children of different ages (e.g., if discourse connective information is used more often than morphology information at certain ages).

### 3.3.4 Bayesian listener with both inaccurate representations and inaccurate deployment

Our implementation of a listener with both inaccurate representations and inaccurate deployment of those representations combines the implementations of the listeners with only inaccurate representations and the listeners with only inaccurate deployment. More specifically, the softmax function with contrast parameter $\sigma$ is used for the representations of the prior and likelihood probabilities, as shown in $p_\sigma$ in (15), using $\sigma_\alpha$, $\sigma_{form}$, $\sigma_{con}$, and $\sigma_{mor}$; the selective dropout function with use parameter $\beta$ is used for the deployment of the prior and likelihood probabiltiies, as shown in $p_\beta$ in (18), using $\beta_\alpha$, $\beta_{form}$, $\beta_{con}$, and $\beta_{mor}$. More concretely, the both-inaccurate listener calculates the posterior $p_{\sigma,\beta}$, which uses the inaccurate representation posterior calculation $p_\sigma$ within the mixture model defined for the inaccurate deployment posterior calculation $p_\beta$. So, $p_{\sigma,\beta}$ looks identical to $p_\beta$, except that $p_\sigma(\alpha|$cues, $\alpha_{num}$,$\alpha_{subj?})$ is used in place of $p(\alpha|$cues, $\alpha_{num}$,$\alpha_{subj?})$ in (18); similarly, $p_{\text{UNIF},\sigma}(\alpha_{num}$,$\alpha_{subj?}|$cues) is used in place of $p_{\text{UNIF}}(\alpha_{num}$,$\alpha_{subj?}|$cues), with (19) showing this implementation when all three cues are used.

(19)
$$\begin{aligned} p_{\text{UNIF},\sigma}(\alpha_{num}, \alpha_{subj?}|\text{FORM}, \text{CON}, \text{MOR}) \propto\ & p(\text{UNIF}) \\ & * p(\text{FORM}|\alpha_{num}, \alpha_{subj?})^{\sigma_{form}} \\ & * p(\text{CON}|\alpha_{num}, \alpha_{subj?})^{\sigma_{con}} \\ & * p(\text{MOR}|\alpha_{num}, \alpha_{subj?})^{\sigma_{mor}} \end{aligned}$$

### 3.3.5 Comparing Bayesian listeners

To sum up, we defined four modeled listener types implemented as Bayesian listeners: the baseline listener, the listener with inaccurate representations of information about potential antecedents and cues, the listener with inaccurate deployment of accurate information, and the listener with both inaccurate representations and inaccurate deployment. We can now compare these modeled listeners on their ability to best capture the child and adult control data from the pronoun interpretation task. This will allow us to evaluate which hypothesis provides the best explanation for children's observed behavior at each age and how this compares to the best explanation for adults' observed behavior. We can also look within the best-fitting model to find the parameter values that yield the best fit. This will give us an estimate of exactly how inaccurate participants' representations are (i.e., how much $\sigma$ differs from 1) and/or how inaccurate their deployment is (i.e., how much $\beta$ differs from 1), and whether this changes over time.

## 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), and chose the combination with the best match to the observed participant responses 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 (20); 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 $\sigma$ parameters and 4 $\beta$ parameters, for a total of 8 parameters) has to have a substantially better fit than the inaccurate representation listener (with 4 $\sigma$ parameters) and the inaccurate deployment listener (with 4 $\beta$ 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.

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

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.[6] The data in this case are the number of subject and object responses provided by participants in each condition; so, we calculate $P$(all participant responses across all conditions | best-fitting parameter values) for each modeled listener type, and take the log. A better log likelihood score is closer to 0.[7] This in turn means that 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 3.

| age | baseline | inacc rep | inacc deploy | both inacc |
|---|---|---|---|---|
| $\leq 3$ | 2913.00 | **735.99** | **735.17** | 758.11 |
| 4 | 2784.61 | **752.89** | **754.06** | 773.65 |
| $\geq 5$ | 1931.33 | **590.04** | **590.70** | 607.12 |
| adults | 1323.82 | **642.48** | 646.96 | 668.58 |

Table 3: 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).

We see that the baseline model, with both accurate representations and deployment, fares the worst; this result suggests that both children and adults are using either inaccurate representations, inaccurate deployment of accurate representations, or both inaccurate representations and inaccurate deployment. That is, becoming adult-like in pronoun interpretation doesn't mean that children should represent and deploy available pronoun information with total accuracy. Rather, children have to become more adult-like in how they inaccurately do these things.

We can see in Table 3 that children's behavior is best captured by the modeled listener types that use either inaccurate representations or inaccurate deployment, but not both. So, these results

[6]Appendix B 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 for each age group.

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

suggest that children either (i) have inaccurate representations, but deploy them accurately, or (ii) have accurate representations, but deploy them inaccurately. Table 3 also suggests that adult behavior is best captured by the modeled listener type using inaccurate representations, but deploying those representations accurately.

So, it could be that children are basically doing the same thing as adults: accurately deploying inaccurate representations. In this case, children wouldn't need to qualitatively shift their approach to interpreting pronouns. However, our results are also compatible with children doing something fundamentally different from adults, by inaccurately deploying accurate 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 representations of available pronoun information, but accurately deploy those distorted representations in real time.

### 3.4.2 What's changing?

To better understand the potential change between children and adults better, we can examine the best-fitting model parameters that yielded the BIC scores in Table 3. In particular, we can look at the best-fitting $\sigma$ contrast values for the inaccurate representations listener to understand what children and adults may be doing, as well as the best-fitting $\beta$ use values for the inaccurate deployment listener to understand what else children may be doing. These parameter values are shown in Table 4.[8]

| | $\sigma_{for}$ | $\sigma_{con}$ | $\sigma_{mor}$ | $\sigma_{\alpha}$ | $\beta_{for}$ | $\beta_{con}$ | $\beta_{mor}$ | $\beta_{\alpha}$ |
|---|---|---|---|---|---|---|---|---|
| **children: inaccurate representations** | | | | | | | | |
| ≤3 | 0.00 | 0.11 | 0.04 | 0.00 | 1 | 1 | 1 | 1 |
| 4 | 0.00 | 0.01 | 0.09 | 0.00 | 1 | 1 | 1 | 1 |
| ≥5 | 0.02 | 0.28 | 0.11 | 0.00 | 1 | 1 | 1 | 1 |
| **children: inaccurate deployment** | | | | | | | | |
| ≤3 | 1 | 1 | 1 | 1 | 0.00 | 0.18 | 0.10 | 0.00 |
| 4 | 1 | 1 | 1 | 1 | 0.00 | 0.08 | 0.24 | 0.00 |
| ≥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 4: Best-fitting parameter values for the modeled listeners that best capture the pronoun interpretation behavior of children and adults, based on BIC scores. $\sigma$ values are used for inaccurate representations, while $\beta$ values are used for inaccurate deployment of representations. Accurate deployment in the inaccurate representations learner uses $\beta$ values of 1; accurate representations in the inaccurate deployment learner uses $\sigma$ values of 1.

**Inaccurate representations in adults and possibly in children.** Adult behavior is best captured by a modeled listener with inaccurate representations, and this is one of the two modeled listener types that children's behavior is also best captured by. In Table 4, we see that the way in which

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

these representations are inaccurate is qualitatively the same between adults and children. In particular, the best-fitting $\sigma$ 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 $\sigma$ value of 0.33 (the highest among the adults' $\sigma$ values) would take a probability distribution of 0.75 vs. 0.25, and transform it into 0.59 vs. 0.41. Lower $\sigma$ values have even stronger smoothing effects: a $\sigma$ value of 0.28 (the highest among the children's $\sigma$ values) transforms 0.75 vs. 0.25 into 0.58 vs. 0.42; $\sigma$=0.11 transforms 0.75 vs. 0.25 into 0.53 vs. 0.47; $\sigma$=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 $\sigma$=0.00 (adults and all children: $\sigma_\alpha$; children 4 and younger: $\sigma_{for}$). This can be interpreted as participants completely ignoring this information, as a $\sigma$ 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 $\sigma_{for}$), children would need to smooth away less of the probability contrast available from the input (adults: 0.25; children: 0.00-0.02). Similarly, children would need to smooth away less of the probability contrasts available for the information about discourse connectives and agreement morphology (adult $\sigma_{con}$: 0.33; child $\sigma_{con}$: 0.01-0.28; adult $\sigma_{mor}$: 0.28, child $\sigma_{mor}$: 0.04-0.11). More generally, under this view, development involves tuning how much pronoun 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 information, but are inaccurately deploying them. In particular, children would completely ignore information about the prior (i.e., children use only information from the current pronoun context) and completely ignore information about the pronoun form, as $\beta_\alpha$=$\beta_{for}$=0.00. The other two pronoun information types of discourse connectives and agreement morphology are ignored less, with the amount varying by age ($\beta_{for}$=0.08-0.43, $\beta_{mor}$=0.10-0.30). In contrast, adults would be deploying their representations accurately (all $\beta$s=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 representations. We do note that adult-like smoothing of representations involves completely smoothing away information about the prior distribution over antecedents ($\sigma_{alpha}$=0.00); this total smoothing is behaviorally equivalent to never using that information in the moment ($\beta_\alpha$=0.00), because it means children completely ignore the information from the input. So, for information about prior antecedents, children wouldn't necessarily need to qualitatively change what they end up doing, which is to completely ignore that 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 information available (inaccurate representations) or they're selectively using accurate representations of the pronoun information available (inaccurate deployment). Notably, the best-fitting inaccurate representation listener for children completely smoothed away all information for the prior distribution over antecedents ($\sigma_{\alpha}$=0.00) and nearly always did so for the pronoun form ($\sigma_{for}$=0.00-0.02); smoothing away all information in an inaccurate representations learner yields an equivalent result to the best-fitting inaccurate deployment learner that never deploys these information types ($\beta_{\alpha}$=$\beta_{for}$=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 ($\sigma_{\alpha}$=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 representation less (to achieve adult-like $\sigma_{for}$=0.25), or making a qualitative shift to always deploy (i.e., $\beta_{for}$=1) a representation that's inaccurate from being smoothed (i.e., $\sigma_{for}$=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 representations less or by making a qualitative shift to always using a smoothed 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 discourse connectives mattered for children age three and younger, agreement morphology mattered for children age four, 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 age group.

In particular, if we look at the modeling results for children, the result is the same for all three age groups: the best-fitting modeled listeners were either relying on inaccurate representations that were smoothed or inaccurately deploying accurate representations. Either option results in children ignoring specific types of information, in line with the logistic regression analysis from before. For children age three and younger, the discourse connective probabilities were either less smoothed than the other information types ($\sigma$=0.11 vs. 0.00-0.04) or deployed more often ($\beta$=0.18 vs. 0.00-0.10). For children age four, the agreement morphology probabilities were either less smoothed than other information types ($\sigma$=0.09 vs. 0.00-0.01) or deployed more often ($\beta$=0.24 vs. 0.00-0.08). For children age five and older, both the discourse connective probabilities and the agreement morphology probabilities were either less smoothed ($\sigma$=0.11-0.28 vs. 0.02) or deployed more often ($\beta$=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 more 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 of the probability contrasts available in the input for the pronoun cues. All cue representations (pronoun form, discourse connectives, agreement morphology) were smoothed around the same amount ($\sigma$=0.25-0.33). So, the modeled learner aligns with the behavioral results by showing that for adults, all cue information mattered, and in fact was equally important.

For both children and adults, what the modeling results additionally uncover is exactly how the information that mattered might actually matter. For children, either the representations of 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 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 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. In particular, the modeling allowed us to identify potential differences in child and adult representations of information relevant for interpreting pronouns as well as potential differences in child and adult deployment of that 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 developmental 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 information about pronoun cues (they learn to smooth away less information), or (ii) both the way they represent information about pronoun cues (they learn to smooth some information away) and the way they deploy that information (they always deploy it).

More specifically, our results suggested one of two options for children's observed behavior. First, children could have representations that are both overly-smoothed and unequally-smoothed for pronoun cue information, though they 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 information: they should smooth these 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 adults: smoothed representations of pronoun cue information that are always deployed. What changes is simply how much these representations are smoothed. More generally, what changes is not the fact that the representations are inaccurate, but the way that the representations are inaccurate.

The other option is for children to have accurate representations that are inaccurately deployed, 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 cue representations inaccurate by smoothing away relative differences the way adults do; second, they would need to make their deployment accurate, by always using these smoothed 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 represent 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 information for interpreting pronouns. Below, we discuss the plausibility of this finding in the context of human cognition more generally, as well as for 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?

When it comes to how (adult) humans 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). What's notable is that optimal isn't necessarily the same as accurate. 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.

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.

Within language development, 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)). Here, being resource-optimal could well lead adults to inaccurate representations of relevant pronoun information.

Still, what about the specific ways that our modeling results suggest that adults are inaccurate? Are these plausible, especially if we consider them from a resource-optimal perspective? For the representation of pronoun cue information, our results suggested that adults had fairly smoothed representations, with $\sigma$=0.25-0.33 for the information about pronoun form, discourse connectives, and agreement morphology.[9] Within decision-making 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). This 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 5). For example, agreement morphology probabilities ($p(\textsc{mor}|\alpha)$) of 0.995-0.998 are smoothed to 0.815-0.851 with a $\sigma_{mor}$=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 5, where $\sigma_\alpha$=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

[9] Adults also had complete smoothing of the prior distribution over possible antecedents, $\sigma_\alpha$=0.00. We discuss this more below.

| **original input probabilities** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | prior | likelihoods | | | | | |
| | | $p(\alpha)$ | $p(\text{FORM}\|\alpha)$ | | $p(\text{CON}\|\alpha)$ | | $p(\text{MOR}\|\alpha)$ | |
| antecedent type | | | ø | overt | *después* | *porque* | 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 |
| **smoothed input probabilities** | | | | | | | | |
| adult $\sigma$ | | 0.00 | 0.25 | | 0.33 | | 0.28 | |
| | | prior | likelihoods | | | | | |
| | | $p(\alpha)$ | $p(\text{FORM}\|\alpha)$ | | $p(\text{CON}\|\alpha)$ | | $p(\text{MOR}\|\alpha)$ | |
| antecedent type | | | ø | overt | *después* | *porque* | SG | 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 | 0.190 | 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 5: 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 $\sigma$ estimated for the best-fitting modeled listener for adult pronoun interpretation.

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" 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 developmental 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 2) 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 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.

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.

We note here that this may not be implausible 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. Relatedly, 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.

# 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 developmental 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 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), and so children must learn not to become accurate, but rather how to become usefully 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 development actually is in any particular domain. We hope to have done that here for pronoun interpretation.

# References

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.

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.

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.

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.

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.

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. in press. Considering the whole paradigm: Preschoolers' comprehension of agreement is not uniformly late.

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. in press. After the Null Subject Parameter: Acquisition of the Null-Overt Contrast in Spanish. *Language Learning and Development*.

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.

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).

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.

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.

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*.

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.

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. in press. Modeling syntactic acquisition. *In:* 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.

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.

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.

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.

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 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'. 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
   Everyone greet-3Pl. The girls greet-3Pl A the boys, and the boys greet-3Pl A the
   maestra.
   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.

   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
   Oh no, the classroom got-3Sg dirty. There-is mud on the floor, and the boys stay-3Pl A
   limpiar.
   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,
The-walk to-the park, take-3Pl the food in a restaurant. Eat-3Pl soup of noodles,
y los niños piden postre.
and the 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
Already go-3Sg A to-start the class. Enter-3Pl to-the living room, and the boys
se sientan.
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
Oh no, forgot-3Sg something. Left-3Pl the lights lit, and the boys turn-off-3Pl the
luces.
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
    After of the party must to-put-in-order. Put-away-3Pl the dishes, and the boys
    limpian la mesa.
    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

# B 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 (children: $\leq$3, 4, $\geq$5; adults).

| | $\sigma_{for}$ | $\sigma_{con}$ | $\sigma_{mor}$ | $\sigma_{\alpha}$ | $\beta_{for}$ | $\beta_{con}$ | $\beta_{mor}$ | $\beta_{\alpha}$ |
|---|---|---|---|---|---|---|---|---|
| **inaccurate representations** | | | | | | | | |
| $\leq$3 | 0.00 | 0.11 | 0.04 | 0.00 | 1 | 1 | 1 | 1 |
| 4 | 0.00 | 0.01 | 0.09 | 0.00 | 1 | 1 | 1 | 1 |
| $\geq$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** | | | | | | | | |
| $\leq$3 | 1 | 1 | 1 | 1 | 0.00 | 0.18 | 0.10 | 0.00 |
| 4 | 1 | 1 | 1 | 1 | 0.00 | 0.08 | 0.24 | 0.00 |
| $\geq$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** | | | | | | | | |
| $\leq$3 | 0.0<br>*any* | 4.00 | 0.26 | 0.0<br>*any* | *any*<br>0.00 | 0.12 | 0.18 | *any*<br>0.00 |
| 4 | 0.0<br>*any* | 4.00 | 0.24 | 0.0<br>*any* | *any*<br>0.00 | 0.10 | 0.43 | *any*<br>0.00 |
| $\geq$5 | 0.0<br>*any* | 4.00 | 0.22 | 0.0<br>*any* | *any*<br>0.00 | 0.25 | 0.61 | *any*<br>0.00 |
| adults | 0.33 | 0.66 | 0.29 | 0.0<br>*any* | 0.78 | 0.52 | 1.00 | *any*<br>0.00 |

Table A1: Best-fitting parameter values for all modeled listeners for all age groups. Inaccurate representation listeners have contrast parameter $\sigma$, inaccurate deployment listeners have use parameter $\beta$, and listeners with both inaccurate representations and inaccurate deployment have both contrast $\sigma$ and use $\beta$ parameters. Modeled listeners with accurate deployment have $\beta$=1, while modeled listeners with accurate representations have $\sigma$=1. For the both-inaccurate type, if $\sigma_x$=0.00, any $\beta_x$ value can be used, and is listed as *any*; if $\beta_x$=0.00, any $\sigma_x$ value can be used, and is listed as *any*.