Top-Down Learning in the Acquisition of Pronouns

Hannah Forsythe

The ability to use language to refer to people and things is one of the most basic skills that human children must acquire in the first years of life. How they learn to do this given their limited cognitive resources and how the learning process varies across languages is one of the great puzzles of acquisition research. From this perspective, personal pronouns are a particularly interesting case study. Not only are pronouns extremely common cross-linguistically, but understanding what they refer to hinges on the ability to integrate information from multiple sources. Consider the interpretation of she in (1) below. The child must extract the pronoun-internal cues of gender, number, and person in order to begin the search for an appropriate antecedent. Often, however, more than one such antecedent exists (ex. Maria, Ana), and the child must therefore also extract information from pronoun- or even sentence-external sources to decide among these antecedents.

(1) Maria greeted Ana and Juan. Now she is leaving.

Studies on adult pronoun resolution have identified a large number of discourse factors that raise the salience of certain antecedents over others. All else being equal, adults show a preference for antecedents that are mentioned first (Järvikivi et al. 2005, Arnold et al. 2007), agitative (Pykkonen & Jarvikivi 2010), introduced in subject position (Gordon et al. 1993, a.o) and fulfilling a parallel grammatical function to the pronoun (Smyth 1994, Chambers & Smyth 1998). Pronominal form also matters, as the more reduced a referring expression is (ex. unstressed pronouns, null pronouns, clitics) the stronger its tendency to pick out a highly salient antecedent (Gundel et al. 1993, Ariel 2001, Carminati 2002, Alonso-Ovalle et al. 2010). These discourse-level, probabilistic cues contrast sharply with the more statistically reliable, pronoun-internal grammatical cues of person, number, and gender, yet both cue types are necessary for interpretation. Therefore, we ask the question: How do children learn to integrate grammatical and discourse cues together in order to interpret pronouns?

*Michigan State University, forsy40@msu.edu. This research was partially supported by the MSU College of Arts and Letters Dissertation Completion Fellowship. Many thanks to the teachers, children, and administrators at SEDI, Mexico City, in particular Patricia de la Fuente and Beti López Juárez. Thanks to Cristina Schmitt, Alan Munn, Ni La Le, and the rest of the MSU language acquisition lab for insightful feedback.

If statistical reliability is any guide, then the ideal learner should proceed bottom-up, relying first and foremost on grammatical cues and later on discourse cues only if necessary. However, real learners don’t always behave like ideal learners (Gagliardi et al. 2016), and the literature indicates that this may be one of those cases. Therefore this paper will pursue the top-down hypothesis, namely, that children rely on discourse cues at least as early as grammatical cues—even in languages like Spanish, where grammatical cues are abundant.

The paper begins with a review of the relevant experimental literature motivating the top-down hypothesis, as well as our theoretical assumptions concerning the nature of discourse cues. Next, we report three studies addressing Mexican Spanish-speaking preschoolers’ sensitivity to discourse cues in isolation (Study 1) and in combination with grammatical person (Study 2) and number (Study 3) cues.

1. Acquisition background

English-speaking children show early sensitivity to many of the same discourse factors that guide adult pronoun resolution. Eye-tracking studies have shown a preference for agentive, first-mentioned referents in subject position by age 5 (Song & Fisher 2005, Harthorne et.al. 2015) and some even as early as 3 (Song & Fisher 2007, Pyykkönen et al. 2010). Act-out tasks (Maratsos 1973) indicate that 3-year-olds, like adults, have a preference for parallel antecedents when pronouns and their potential antecedents are embedded in structurally and semantically parallel contexts (ex. John hit Harry. Then Sara hit him. him = Harry). Regarding sensitivity to the form of referring expressions, the same study showed that by 5, children can use pronominal stress to reverse the parallel preference. And an elicited production study by Campbell, Brooks & Tomasello (2000) showed that 2-3-year-olds’ choice of referring expression is sensitive to the presence or absence of an explicit antecedent within the elicitation question (ex. “What did X do?” vs. “What happened?”).

In Spanish, evidence is really available only for one discourse cue: the null/overt distinction. Shin & Cairns (2012) found that 8-9-year-olds, like adults, preferred null subjects when referring to the preceding subject antecedent, but that even adolescents failed to show a preference for overt subjects when referring to non-subjects. Both findings suggest late development; however, this may have been due in part to the authors’ use of a metalinguistic judgment task, something that can be very difficult for younger children. Study 1 of this paper therefore probes the null/overt contrast using a forced-choice picture-selection paradigm instead.

At the same time that children show sensitivity to various discourse cues, comprehension studies show them ignoring relevant grammatical cues until surprisingly late in development. For example, in picture-selection tasks in both English and Spanish, children fail to use number agreement to disambiguate null and/or masked 3rd person singular and plural subjects until about 6 (Johnson et al. 2005, Pérez-Leroux 2005). This, despite the fact that their production is
adult-like from much earlier on (Phillips 1995). And in French, children reportedly produce 1st, 2nd, and 3rd person singular pronouns at around the same time early in their second year, yet their ability to deploy this information in comprehension tasks varies by person (with 3rd person suffering the longest delay, see Legendre & Smolensky 2012 and sources therein) and by pronominal form (with clitic and strong pronouns suffering the longest delay, see Pirvulescu & Strik 2014).

This comprehension-production asymmetry suggests that something prevents children from deploying their knowledge of grammatical cues in comprehension. How do children overcome this challenge? One study suggests that discourse plays an important role. Forsythe (2015) used a picture-selection task to test comprehension of Spanish agreement and clitics, which inflect for both person and number. For 1st and 2nd person singular and plural forms, comprehension was adult-like, while for 3rd person singular and plural forms, children tended to select the most recently mentioned antecedent from the preceding discourse—even if this antecedent clashed with that form’s person and number features. That is, even in a morphologically rich language like Spanish, children seem to pay at least as much attention to discourse salience as to grammatical person and number cues. This raises the possibility that children can use the former to facilitate deployment of the latter, a hypothesis that we will now test.

2. Linguistic assumptions

In order to test child sensitivity to discourse cues, we must make a theoretical commitment as to the nature of these cues in the adult grammar. As mentioned above, adults show a variety of disparate, sometimes contradictory pronoun resolution strategies, including strategies directed at the preceding subject and/or first-mentioned antecedent, as well as pragmatically plausible and structurally parallel antecedents. Kehler et al. (2008) argue that what dictates adults’ choice of strategy is the semantic relation that holds between the clause containing the pronoun and the clause(s) containing its potential antecedents. These relations may be overtly signaled via connectors like too, then, therefore, etc., or they may be inferred.

For example, then is a marker of the OCCASION relation, which presupposes that the related events are ordered in time, while so marks RESULT, which assumes a cause-effect relation. As illustrated by the contrast in (2) and (3), different relations trigger different pronoun resolution strategies. The OCCASION relation encourages a topic-continuation strategy, which in the default case ends up being a subject or first-mention strategy. A RESULT relation, on the other hand, encourages a pragmatic strategy guided by real-world knowledge of causes and effects. In addition to these relations, we will also examine children’s sensitivity to the PARALLEL relation, illustrated in (4), which juxtaposes similar events and thus triggers a parallel resolution strategy.
3. Hypotheses and predictions

Given this experimental and theoretical background, we now have the terminology to formulate a version of the top-down hypothesis. The hypothesis comes in two parts: a processing effect and a developmental effect:

**H1**: Children use higher-level discourse relations to facilitate processing of lower-level grammatical cues.

**H2**: Sensitivity to discourse relations leads to better *overall* deployment of children’s knowledge of grammatical cues.

Studies 2 and 3 test this two-part hypothesis by pitting discourse relations against grammatical person and number cues. When discourse and grammatical cues disagree (incongruent conditions), the adult-like response is to rely on grammatical cues, though performance may suffer a little due to the processing load that this conflict imposes. When discourse and phi-features agree, performance should be at ceiling. For children, our two-part hypothesis makes the following two predictions:

**Prediction 1**: Children sensitive to discourse relations will show better comprehension of grammatical cues in congruent relative to incongruent conditions.

**Prediction 2**: Sensitivity to discourse relations will trigger better overall use of grammatical cues in comprehension (i.e., emergence of the congruence effect will precede or coincide with above-chance performance overall).

Before testing these predictions, we will take a moment to first confirm that the early sensitivity to discourse cues found among English-speaking children is also characteristic of children acquiring Mexican Spanish.

4. Study 1: Sensitivity to Discourse relations and pronominal form

The literature has shown early sensitivity to discourse cues among English-speaking children interpreting grammatically ambiguous pronouns. We now test this among Spanish-speaking children, focusing on two discourse cues: (i) the contrast between discourse relations OCCASION and RESULT, and (ii) the contrast between null and overt pronominal subjects. Using a forced-choice picture-selection paradigm, we examine the interpretation of 3rd person singular pronouns with two grammatically compatible potential antecedents, as in (5)-(6) below.
If children are sensitive to discourse relations, then the OCCASION condition should elicit more subject-antecedent responses relative to the RESULT condition, which is designed to be pragmatically biased towards a non-subject interpretation. If children are sensitive to form of referring expression, then null subjects should elicit more subject responses relative to overt pronominal subjects.

4.1. Methods

4.1.1. Subjects

40 adults (34 women) and 73 children (39 girls) ages 2;11 to 6;4 (mean: 4;6, sd: 11.5 months) completed the task, with 3 exclusions because of failure to learn the names of the characters in the pictures.

4.1.2. Design, procedure, and exclusions

Subjects participated in a 16-prompt forced-choice picture selection task in a 2 (null, overt) x 2 (Occasion, Result) within-subjects design. Items were created by crossing each condition with eight verb-phrase pairs (alegrar-aplaudir: ‘cheer up – applaud,’ cantar para-bailar: ‘sing for – dance,’ perseguir-cansarse: ‘chase – get tired,’ asustar-gritar: ‘scare – yell,’ pegarle-irse: ‘hit – leave,’ pelearse con-llorar: ‘quarrel with – cry,’ tocar-reírse: ‘poke – laugh,’ and hablar con-sonreír: ‘speak to – smile’). Each subject saw every item in two conditions. Prompts were blocked by condition, and blocks were separated by items from a separate study that used some of the same characters. Order of presentation was randomized within each block, ordering of blocks was counterbalanced across subjects, and the position of the first-mentioned character (left or right side) was counterbalanced across verb-phrase pairs.

Pictures were presented on a computer screen using Psychopy version 1.82.01 (Pierce 2007). Children were read the prompts by a native speaker who recorded their responses on the computer, while adults listened to pre-recorded prompts and entered their responses for themselves. Prior to beginning, subjects were taught the names of the characters in the pictures (María, Sara, Juan, Pedro) and given a 4-item recognition task. Subjects providing fewer than 2 correct answers were excluded (3 children). The remaining children gave an average of 3.37 correct responses out of 4 (SD: 0.73) and adults an average of 3.6 out of 4 (SD: 0.5).
4.2. Results and Discussion

The proportion of subject responses in each condition is shown in Fig. 1. For all three studies reported here, children were divided into the same age groups: those under 4;6 and those at least 4;6. Separate mixed effects logistic regression models were fit to each age group, with Relation (OCCASION, RESULT) and pronominal form (null, overt) as level-1 fixed effects and items and subjects as level-2 random intercepts. For adults, both effects were significant (relation: \( \beta = -1.1, p < 0.0001 \); pronominal form: \( \beta = -0.88, p = 0.001 \)). For younger children (\( N = 40, M = 3;9 \)), relation was significant (\( \beta = -0.36, p = 0.03 \)), and for older children (\( N = 33, M = 5;5 \)), Pronominal form was significant (\( \beta = -0.58, p < 0.01 \)).

![Figure 1](image)

**Figure 1**: Interpretation of grammatically ambiguous null and overt subjects in Spanish

Contra Shin & Cairns, and consistent with findings in English, these results indicate early sensitivity to both discourse relations and pronominal form, although children may weight these cues differently at different ages. The next two studies test the top-down hypothesis more directly by pitting discourse relations against grammatical person and number. In order to separate the effects of discourse cues from the potential effects of pronominal form, we test not only subject but also object pronouns, which do not present the null/overt alternation.

5. Study 2: Pitting discourse relation OCCASION against grammatical PERSON

This study examines children’s integration of cues marking the OCCASION relation, which triggers a topic-driven pronoun resolution strategy, with cues marking 1st and 3rd person. In this task, OCCASION is overtly marked by the connective *y ahora* ‘and now,’ plus a switch from preterit to simple present. Person is either marked via agreement with a null subject (7) or encoded in the form of an accusative clitic (8).
Ana bailó una cumbia conmigo y ahora canta o
Ana danced a cumbia with-me and now sing-3S [congruent]
sing-1S [incongruent]

Ana llegó a la casa conmigo y ahora Chicho está saludándola/me
Ana arrives home with-me and now Chicho is greeting-3S [congruent]
-1S [incongruent]

For half of the items, these two cues will be congruent with each other and for the other half they will be in opposition. The intended antecedent of the pronoun can be determined by attending only to person cues. However, if children rely on discourse relations to facilitate the processing of grammatical cues, then their rate of target answers should be sensitive to this manipulation. Moreover, if this facilitation plays a role in development, then the age at which the congruence effect appears should precede or coincide with an overall increase in target responses.

5.1. Methods
5.1.1. Subjects

42 adults (35 women) and 82 children (43 girls, ages 2;11-6;5, \( M = 4;6, \) SD = 12.1 months) completed the task, with no exclusions.

5.1.2. Design, procedure, and exclusions

Subjects participated in a 16-prompt forced-choice picture selection task in a 2 (congruent, incongruent) x 2 (3rd person, 1st person) x 2 (subject, object) within-subjects design. Items were created by crossing congruence and person with four verb-phrase pairs in the subject condition (tener hambre-comer: ‘be hungry – eat,’ cantar-bailar: ‘sing-dance,’ bailar-cantar: ‘dance-sing,’ and pintar-dibujar: ‘color – draw’) and four pairs in the object condition (tener frío-tar: ‘be cold – cover,’ llegar-saludar: ‘arrive-greet,’ embarrarse-lavar: ‘get dirty – wash,’ and hablar-escuchar: ‘talk – listen’). Each subject saw these verb pairs in exactly two out of four possible conditions. Items were blocked by condition and blocks were separated by prompts from a separate study.

Experiment procedures and counterbalancing were carried out in the same manner as in Study 1. Characters included the experimenter and an adult named Ana who was present during the study but did not interact with the subjects. Prior to beginning, subjects were asked to identify these people in real life as well as on the screen and were given a 7-item recognition task to verify that they had learned who was who. Adults scored an average of 5.64 out of 7 (SD = 1.6) and children an average of 5.10 (SD = 1.4). No subjects were excluded.
5.2. Results and Discussion

The proportion of target responses for adults (N = 36), children under 4;6 (N = 37; M=3;9) and children at least 4;6 (N = 36; M=5;6) is shown in figure 2, collapsed across subject and object conditions.

To test Prediction 1, we fit a multilevel mixed effects logistic regression model to each age group, using the predictors of congruence, pronoun position (subject, object), and person (3\textsuperscript{rd}, 1\textsuperscript{st}) as level-1 fixed effects and subjects and items as level-2 random intercepts. For adults, there was a significant though small effect of congruence on the proportion of target answers (β=2.1, p<0.004). For older children there was also a significant effect of congruence (β=2.4, p<0.0001). For younger children, person (β=1.7, p<0.0001) and pronoun position (β=0.7, p=0.01) were the only significant effects.

To test our Prediction 2, two-tailed t-tests were used to compare the proportion of target answers to chance for each age group in each condition. All age groups performed well above chance in 1\textsuperscript{st} person conditions, congruent and incongruent alike (all p < 0.003). For the 3\textsuperscript{rd} person, adults performed above chance in all conditions (all p < 0.001), as did older children (all p < 0.001). Younger children performed no better than chance in any of the 3\textsuperscript{rd} person conditions (all t < 0.73; all p > 0.24).

These results are consistent with both parts of the top-down hypothesis. A grammatical cue that children have shown consistent difficulty using (3\textsuperscript{rd} person marking) is facilitated by the presence of discourse cues, and this coincides with better overall use of this cue. Next, we will pit a different discourse relation against a different grammatical feature.

6. Study 3: Pitting discourse relation PARALLEL against grammatical NUMBER

This study examines children’s integration of cues marking the PARALLEL relation, which triggers a parallel pronoun resolution strategy, with cues marking
singular and plural. In this task, PARALLEL is overtly marked by the connective también ‘too,’ plus semantic parallelism between the two clauses. Number is either marked via agreement with a null subject (9) or encoded in the form of an accusative clitic (10). These cues are alternately placed in congruence and in opposition to each other.

(9) La maestra abraza a las niñas y abraza/n a los niños también.
The teacher hugs the students and hug-3S the boys, too  [congruent]
-3P                [incongruent]

(10) La maestra abraza a las niñas y María la/las abraza también
The teacher hugs the students and Maria 3S hugs, too.   [congruent]
3P                [incongruent]

Once again, the intended antecedent of the pronoun can be determined by attending only to number marking. However, if children use this discourse relation to facilitate the processing of this grammatical cue, then their rate of target answers should be sensitive to the congruence manipulation. And if this facilitation plays a role in development, then the age at which the congruence effect arises should precede or coincide with an overall increase in target responses.

6.1. Methods
6.1.1. Subjects

23 adults (22 women) participated, with no exclusions. 44 children (25 girls) ages 2;11-5;10 (M = 4;5, SD = 10.8 months) completed the task, with 4 exclusions because of failure to learn character names.

6.1.2. Design and procedure

Subjects participated in a 16-prompt forced-choice picture selection task, blocked by condition. Congruence (congruent, incongruent) and number (singular, plural) were manipulated within-subjects and pronoun position (subject, object) was manipulated between-subjects. Items were created by crossing each condition with each of four verbs (abrazar: ‘hug,’ mirar: ‘look at,’ perseguir: ‘chase,’ and tapar: ‘cover’). Each subject saw every item in four conditions.

Experiment procedures and counterbalancing were carried out in the same manner as in Study 1. Characters included one singular antecedent (a teacher), one plural antecedent (a group of girls), and a third character that varied from item to item, including either Sara, María, or a group of boys. Prior to beginning, subjects were asked to identify these characters in a 5-item recognition task, and any subject that failed to produce at least 3 out of 5 correct responses was excluded (N = 4 children). The remaining children
answered correctly an average of 4.4 out of 5 (SD 0.74) items, and adults an average of 4.7 out of 5 (SD 0.5).

Finally, blocks were separated by 12 filler items, which were used to check for general comprehension of the task.

6.2. Results and Discussion

The proportion of target responses for adults, younger children (N = 23; 14 girls; M = 3;9, SD 5.9), and older children (N = 17; 11 girls; M = 5;3; SD 4.7) is shown in figure 3, collapsed across singular and plural responses. All age groups scored above chance in filler items (younger children: M = 0.73, t(22) = 5.24, p < 0.001, older children: M = 0.81, t(16) = 8.52, p < 0.001, adults: M = 0.92, t(22) = 11.25, p < 0.001).

To test Prediction 1, we fit a multilevel mixed effects logistic regression model to each age group, using the predictors of congruence, pronoun position (subject, object), and number (singular, plural) as level-1 fixed effects and subjects and items as level-2 random intercepts. For adults, there were significant effects of congruence (β=2.0, p<0.001), and pronoun position (β=2.8, p<0.001), both of which appear to be driven by a very low proportion of target responses in the incongruent subject condition. (See discussion below.) For older children, the same effects reached significance (congruence: β=0.8, p=0.003, pronoun position: β=0.6, p=0.02). For younger children, there were no significant effects (all p > 0.16).

To test Prediction 2, two-tailed t-tests were used to compare the proportion of target answers to chance for each age group in each condition. Adults and older children both performed above chance in all conditions (adults: all p < 0.001; older children: all p < 0.053) except for the incongruent subject condition. Younger children performed above chance in the congruent subject condition only (M = 0.66, t(21) = 2.47, p = 0.022).

![Figure 3](image-url)
The overall results of this experiment are for the most part consistent with the top-down hypothesis. A grammatical cue that children have shown difficulty using in other tasks (number marking) is facilitated by the presence of discourse cues, and this leads to better overall use of this cue in most conditions.

Two results were unanticipated, however. First, younger children showed above-chance performance in one condition (congruent subject) despite their lack of sensitivity to the effect of congruence. This may indicate that discourse is not as instrumental in children’s developing ability to use grammatical cues in real time as predicted. Second, adults fail to perform above chance in one condition (incongruent subject). This result is reminiscent of cases of agreement attraction, in which perfectly competent speakers fail to perceive agreement markers (Wagers et al. 2009, a.o.); nevertheless it still calls the rest of the subject condition into question.

However, a different explanation could potentially explain these two effects. Notice that unlike the object condition, the subject condition displays perfect syntactic parallelism. Thus, it could be that syntactic priming alone is responsible for adult and older children’s unexpectedly poor performance in the incongruent subject condition and younger children’s unexpectedly high performance in the congruent subject position. If this is true, it means that only the object condition accurately measures participants’ sensitivity to the PARALLEL discourse relation and therefore only that subset of the data should be used to test our hypothesis. And indeed, in this subset of the data, the top-down hypothesis is upheld.

7. General Discussion

Together, these three studies add to the growing literature on children’s discourse sensitivity, as well as the already substantial literature on the acquisition of person and number features. Study 1 shows that, like English-speaking children, Spanish-speaking children use discourse cues in comprehension tasks at least as early as they have been found to use person and number cues. Studies 2 and 3 show that after approximately 4 ½, they are sensitive to overt markers of OCCASION and PARALLEL discourse relations. And these studies also suggest that such discourse cues facilitate the processing of person and number cues, and that this facilitatory effect plays a role in children’s growing ability to deploy their knowledge of grammatical features in real time.

References


Gundel, Jeanette, Nancy Hedberg and Ron Jacharski (1993)


