

Piecewise Structural Equation Modeling of the collective implicature in child language

John Grinstead, Ramón Padilla-Reyes, Melissa Nieves-Rivera & Morgan Oates

To cite this article: John Grinstead, Ramón Padilla-Reyes, Melissa Nieves-Rivera & Morgan Oates (2023): Piecewise Structural Equation Modeling of the collective implicature in child language, *Language Acquisition*, DOI: [10.1080/10489223.2023.2185523](https://doi.org/10.1080/10489223.2023.2185523)

To link to this article: <https://doi.org/10.1080/10489223.2023.2185523>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 11 Apr 2023.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Piecewise Structural Equation Modeling of the collective implicature in child language

John Grinstead^a, Ramón Padilla-Reyes^b, Melissa Nieves-Rivera^a, and Morgan Oates^a

^aThe Ohio State University; ^bReality Labs, Meta Platforms, Inc.

ABSTRACT

We test children's distributive and collective sentence interpretations and the variables that predict them. In our first experiment, we establish that adult English collective sentences with *the* or *some* in the subject are categorically collective in their interpretations. We further demonstrate that children's collective and distributive interpretations are predicted by an independent measure of lexical growth, consistent with the lexical refraction hypothesis, and that their collective interpretations are predicted by their distributive interpretations, consistent with the pragmatic scale hypothesis. Furthermore, the distributive interpretations produce complete mediation between lexicon and collective interpretations in a mediation analysis. In our second experiment, we take independent measures of Spanish-speaking children's knowledge of the Approximate Number System, inhibition, lexicon and phrasal syntax. We then fit a Piecewise Structural Equation Model with these variables, with high statistical power and showing high goodness-of-fit. We consider the contrast between this model of collective interpretations, which are putatively conversational implicatures, and quantity implicature models from previous studies.



ARTICLE HISTORY

Received 17 January 2022

Accepted 24 January 2023

1. Modeling implicature interpretations

Comprehending natural language utterances in a given pragmatic context is a task that draws on multiple areas of cognition, including lexicon and syntax. When the utterance contains morphemes expressing quantity information, as most finite sentences do (e.g., singular vs. plural distinctions), some sort of basic numerical ability must also be engaged in the process. Because most utterances are inherently susceptible to many different interpretations, the executive function abilities that are thought to help us inhibit irrelevant interpretations, attend to relevant interpretations, and hold in memory important aspects of the linguistic signal are also implicated. Given the complexity of this system of variables, it is impressive that adults, much less children, are ever able to draw interpretations of given utterances reliably. As we will see, children take a relatively long time to interpret sentences with collective and/or distributive interpretations in an adult-like fashion. What is the cause of this delay? To gain insight into the cause of this delay, we explore the system that yields these interpretations. To do so, we measure an array of linguistic and non-linguistic cognitive variables and attempt to model them, using a statistical technique borrowed from the life sciences, known as Piecewise Structural Equation Modeling (PSEM). Before we turn to this larger study of children's developing interpretations in Spanish, we will attempt to replicate in child English earlier findings from child Spanish.

CONTACT John Grinstead  grinstead.11@osu.edu  The Ohio State University, 298 Hagerty Hall – 1775 College Road, Columbus, OH 43214

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

2. Collective and distributive interpretations

Children across an array of languages appear to be delayed in their interpretations of distributive sentences. In child English (e.g., Hanlon 1986, Brooks & Braine 1996, Musolino 2009, Syrett & Musolino 2013), Spanish (Grinstead et al. 1998), Dutch (e.g., de Koster et al. 2017), Italian (Pagliarini et al. 2012), and Mandarin (Brooks et al. 1998), children do not show adult-like behavior until they are in middle-to-late elementary school, around 10 or 11 years of age. In particular, up until this relatively advanced age for first language acquisition, Spanish-speaking-children will accept the following distributive sentence type in collective contexts, such as the one depicted in Figure 1. Adults will only accept it in a distributive context, such as the one given in Figure 2.

1. Cada minion movió una piedra.
Each minion moved a rock.

Intriguingly, Pagliarini et al. (2012) showed that children's rates of acceptance of collective sentences with a plural definite subject in child Italian, marked with *i* or *le*, which occurred in a distributive context, could be predicted by their rates of accepting distributive sentences with subjects modified by distributive *ciascun* (each) in collective contexts. That is, interpretations of the collective sentences, in a cross-sectional sample of Italian-speaking children, appeared linked to the development of children's interpretations of distributive sentences. This behavior seems consistent with the authors' hypothesis (call it the Pragmatic Scale Hypothesis) that the quantifiers were developing *in tandem* as part of a collective-distributive pragmatic scale.¹

In follow-up work in child Spanish, by Grinstead et al. (2021), in a cross-sectional sample of Spanish-speaking children, the linked development of the distributive (*cada*) and the definite plural collective (*los*) was confirmed. Furthermore, the distributive also predicted the collective interpretations of sentences with subjects modified by an indefinite plural quantifier (*unos*). Novel findings in



Figure 1. Image of the last scene of a collective scenario, in which all three minions push a single rock.

¹See Padilla-Reyes (2018) for an in-depth discussion of prominent formal semantic theories of collectivity and distributivity, including Link (1983), Landman (1989), Schwarzschild (1996), Roberts (1987), and others. As in Grinstead et al. (2021), we take the position that the Pragmatic Scale Hypothesis is the only psycholinguistically testable option among these theories, with empirical consequences that are obvious to us, at least. We recognize that not all formal semanticists share our mentalist goals in explaining these phenomena.



Figure 2. Image of the last scene of a distributive scenario, in which each of three minions pushes its own rock.

this study included the fact that both collective and distributive interpretations in children could be predicted from general lexical development. Further, the authors showed through a mediation analysis that distributive interpretations of *cada* in collective contexts mediated between lexicon and collective interpretations of both *los* and *unos* in distributive contexts. That is, the predictive power of lexicon on collective sentences with *unos* or *los* was reduced to non-significance when acceptance of distributive *cada* was added to the equation. This is consistent with Dotlačil's (2010) claim that the plural quantifiers form a distributive-collective pragmatic scale and that, with the notable exception of *each*, which appears to derive its distributive interpretation from an entailment, the remaining plural quantifiers are underspecified as their collective vs. distributive interpretations.² When the remaining plural quantifiers are interpreted as collective, the argument is that this interpretation results from a Gricean informativeness calculation, because *each* is a non-ambiguous means of expressing distributivity. Also of interest in this sample, it was found, in contrast to the adult Italian data of Pagliarini et al. (2012), that collective sentences with *los* or *unos* in the subject position were categorically (near 100%) rejected by adult Spanish-speakers in distributive contexts. In Pagliarini's data, acceptance of sentences with (putatively) collective plural definite *i* and *le* subjects, paired with distributive pictures, was roughly 50%. This raises the question of whether there could be cross-linguistic differences or whether methodological differences led to the different acceptance rates.

In addition to the empirical generalization that the lexicon predicts distributive-collective interpretations and that adult interpretations of collectives were categorical, the authors proposed what they refer to as the Lexical Refraction Hypothesis. This hypothesis claims that the meanings of the plural quantifiers on the Distributive-Collective Scale (following Dotlačil 2010) become clearer as more lexical items are added to the child's general lexicon, via contrast. Furthermore, it claims that quantifier-internally, development occurs in children's Approximate Number System (ANS) development, which is argued to have its own developmental trajectory, because it provides ever-more-

²To see that *cada/each* is an entailment, we can use the "in fact" test. While the collective interpretation of the first clause of the following sentence can be canceled and replaced by the distributive entailment in the second clause, the reverse is not possible, arguably because the distributive entailment of *each* cannot be canceled.

- (i) The minions pushed a rock. In fact, each minion pushed a rock.
- (ii) # Each minion pushed a rock. In fact, the minions pushed a rock.

precise quantity representations to the human lexicon that are then refracted as number-related morphemes. This is particularly striking in Spanish, a language in which nouns, verbs, adjectives, determiners, quantifiers—almost all grammatical categories—mark number. This type of relationship suggests that general lexical development, and not just the development of quantifiers, is likely to be predicted by the development of ANS. The second part of this hypothesis, regarding ANS, was not explicitly tested with an independent measure of ANS in this work, however. We will do so in Experiment 2 in the following sections.

3. Piecewise Structural Equation Modeling

Before turning to our experiments, however, we would like to provide some background on the statistical model we will be using for the data produced by Experiment 2, PSEM. As previously alluded to, the fact that children do not show adult-like comprehension of distributive and collective sentences until they are in 5th or 6th grade (10 or 11 years-old) is startling for most developmental linguists. Unraveling the mystery of why these constructions should take so long to develop drives us to consider multiple domains of cognition that could be the sources of such a protracted developmental trajectory. A type of model that allows for representing a complex, acyclic network of predictor variables on an outcome variable is Structural Equation Modeling (SEM). Specifically, SEM not only allows for the estimation of the impact of multiple predictor variables on a single outcome variable, as in multiple regression, but also allows those predictor variables to be evaluated simultaneously as outcome variables, in their own right. This could be helpful in understanding the development of collective-distributive interpretations because we do not know, a priori, which domains of cognition are predictive of the children's interpretations. Thus, there are potentially multiple causes of the long developmental curve. The advantage of this approach is that we are not limited to descriptive analyses of the ages at which children attain mastery of the constructions or to minor manipulations of the linguistic characteristics of the linguistic stimuli (e.g., active vs. passive sentences), interesting and valuable though this information is. Rather, we are able to say, to begin with, what domains of cognition are significantly predictive of collective-distributive interpretations and how these domains interact, or do not interact, with one another in development.

For example, in recent work, Grinstead et al. (2022) used the version of SEM we will be concerned with here, PSEM, to illustrate how a set of linguistic and non-linguistic variables could be modeled to understand children's quantity implicature interpretations of sentences in Spanish using the quantifier *algunos* (some) in subject position, such as 2. Figure 3 illustrates the model.

In Figure 3, we see at the rightmost edge of the path diagram, that measures of phrasal syntax (ARG-I), inhibition (the Flanker Task) and lexicon (*Adivinanzas* "Riddles") are each significantly predictive, and account for unique variance, in children's interpretations of sentences with *algunos* implicature interpretations. The sentences included subject noun phrases that include the quantifier *algunos*, such as 2, which are produced after viewing a video-recorded Truth Value Judgment Task (TVJT) in which all of the children (not just some) went down a slide.

2. Algunos niños bajaron la resbaladilla.
Some children went down the slide.

Following our preceding discussion, however, we see that one of these predictor variables, lexicon, is not only a predictor variable, but also simultaneously an outcome variable, which is significantly predicted by the executive function ability, inhibition, and by the non-linguistic ANS, consistent with the Lexical Refraction Hypothesis proposed by the authors. Finally, we note that inhibition is able to play a role in this model in not only the interpretation of the full sentence in 2, when it occurs in a video-recorded pragmatic context, as assumed in natural language processing (see, Novick et al. 2005 for a review), but also plays its well-known role in lexical development (see, e.g., Gangopadhyay et al.

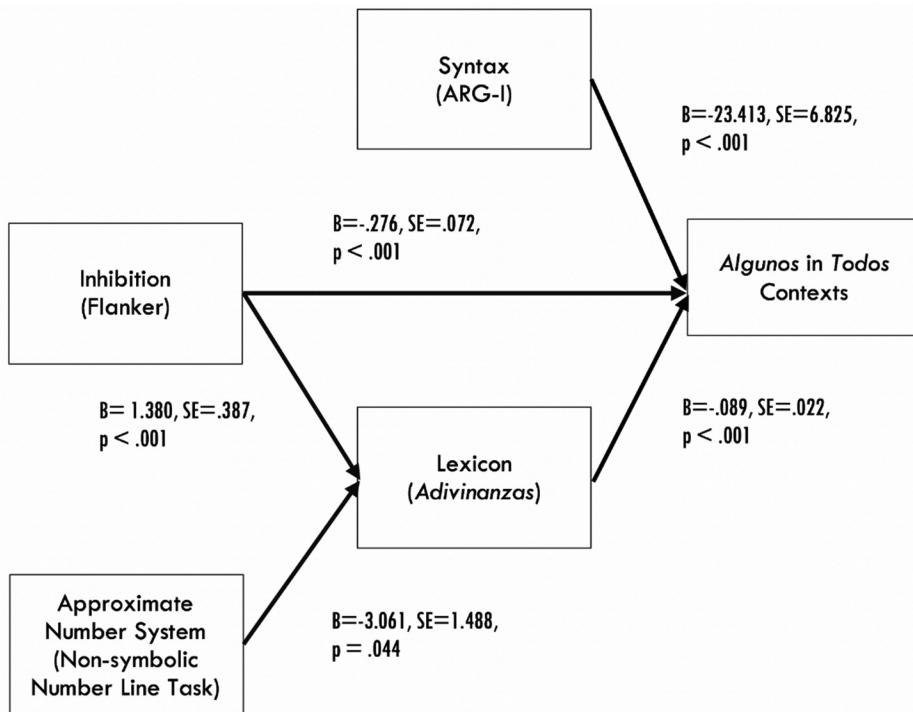


Figure 3. Syntax, lexicon and inhibition predict *algunos* interpretations in contexts where *todos* (“all”) would be correct; inhibition and approximate number predict lexicon: Fisher’s $C = 2.783$, $p = .595$, $df = 4$, $AIC = 18.783$. (Source: Grinstead et al. 2022, Figure 8. Reprinted by Permission of SAGE Publications).

2019, Larson et al. 2020). The theoretical advantage of using this kind of model is that it permits us to explicitly measure not only comprehension of the final, composed sentence in context—difficult enough, in and of itself—but also to gain insight into the relative contributions, and non-contributions, of a range of potentially relevant cognitive variables to this ultimate, pragmatically-situated interpretation.

In what way does PSEM allow us to make claims about “non-contributions” of potentially relevant variables? The answer to this question lies in the PSEM technique, which was created by the ecologist and statistician William Shipley (2000, 2002, 2009) as a means of modeling ecological variables that impact a particular dimension of an ecosystem, but in a structured, hierarchical fashion. This type of modeling intuitively seems very parallel to the kinds of questions we ask in developmental linguistics. Shipley’s technique, further developed by Grace et al. 2015; Lefcheck, 2016, 2019; Schoolmaster et al. 2020; and others, includes a set of potential paths in the model that are tested for significance, referred to as Tests of Directed Separation. For the PSEM model in Figure 3, for example, these include the two paths given in Table 1, neither of which are significant, which supports the plausibility of the model.

Thus, the Tests of Directed Separation consider the 2 outcome variables in the model (lexicon and the implicature interpretation associated with *algunos*) and test to see if they are significantly predicted

Table 1. Non-significant tests of directed separation.

Independence claim	<i>df</i>	Critical value	<i>p</i> -value
Lexicon ~ syntax + inhibition + ANS	58	.680	.499
<i>algunos</i> ~ syntax + inhibition + ANS + lexicon	57	-.677	.498

ANS = Approximate Number System. Source: Grinstead et al. 2022, Table 5. Reprinted by permission of SAGE Publications.

by combinations of predictor variables, other than those under consideration in the model. The non-significant p -values in Table 1 suggests that these paths are not valid competitors for the single-outcome, multivariate causal hypotheses, expressed by the model.

In contrast to the PSEM approach, traditional structural equation models calculate all of the regressions in a model simultaneously, that is, they use a single variance-covariance matrix, which is evaluated using a set of goodness-of-fit indices, including chi-squared statistics, to evaluate how well the model explains the data. PSEM instead replicates this global function of traditional SEMs using the Fisher's C Statistic. Shipley (2002) shows that the result of the Fisher's C statistic has a chi-squared distribution, with $2k$ degrees of freedom, as explained in the work of Lefcheck (2016). The technical formulation of the C statistic, as given in Lefcheck (2016), includes the following formula.

$$C = -2 \sum_{i=1}^k \ln(p_i)$$

The formula includes k , which is the number of independence claims, which is 2 in Figure 3. The notation i means that the first independence claim is 1 and sums up to the k th independence claim in a possible path. In Figure 3, to calculate the C statistic, you take the natural log of the p -value of each independence claim and sum them. You then multiply them by -2 , which yields a C statistic of 2.783, that is, $-2 \times (\ln(.499) + \ln(.498)) = 2.783$. The degrees of freedom in this case, following Shipley (2009), are calculated multiplying k by 2, which equals 4. Given the 4 degrees of freedom, and this chi-squared value, we find a model-wide p -value is .595. If this model-wide p -value were significant, then the causal model being hypothesized should have been rejected, because the null hypothesis for this statistic (as with chi-squared statistics, in general) is that the model fits the data. The path-analytic structure being tested does not fit the data if the p -value is significant. Because the model-wide p -statistic here is not significant, our causal model is not rejected.³

Why would ecologists go to all the trouble of developing PSEMs, if they could simply use traditional SEMs? The answer lies in the scarcity of data provided by particular ecosystems for the hypotheses that scientists want to investigate. Sometimes ecosystems simply do not present the number of exemplars of each of the variable types that would be necessary to adequately power the variance-covariance matrix required for a traditional, globally-computed SEM. In PSEM, in contrast, the statistical power lies in each individual regression in the model and the global goodness-of-fit measure does not depend on there being a single variance-covariance matrix, but rather depends on the C statistic and its relationship to the chi-squared distribution. What this means for behavioral scientists studying phenomena such as the interpretations modeled in Figure 3, is that comparatively greater power (much greater power) can be achieved with a sample of 64 children using PSEM than could have been achieved using traditional SEM, as illustrated in Table 2.

As behavioral scientists, we face a similar challenge to the one faced by ecologists, in the sense that it may simply not be tractable to measure all of the variables that potentially relate to an outcome variable of interest using a sample of hundreds of children, as required for traditional SEM. Notice that especially if we were fortunate enough to have the resources necessary to invest in testing the numbers of children required for a traditional SEM, we might be hesitant to test hypotheses that did not have a high probability of success. This dynamic unfortunately seems likely to depress innovative theorizing and experimentation, which could produce a concomitant depression in the generation of insight into

³Our chi-squared test is a model fit statistic that is assumed to be chi-squared distributed under the null hypothesis. This is common in structural equation modeling context (e.g., Kline 1998). This is not to be confused with the likelihood ratio (LR) test of nested models, as the reviewer correctly points out. LR tests also happen to assume a chi-squared distribution under the null hypothesis (e.g., Agresti 2013), but is a different test. In our setting, a significant test means that our model does not fit the data. A significant LR test means the more complicated model is preferred. This strikes some as counter-intuitive, because we are most frequently in search of significant p -values to make our claims.

Table 2. Power comparison ($n = 64$) of piecewise (locally estimated) and traditional (globally estimated) SEM of the relationships depicted in Figure 3 predicting implicature generation with *algunos*.

	Locally estimated	Globally estimated	
	Lexicon Regression (inhibition, ANS)	<i>algunos</i> Regression (syntax, lexicon, inhibition)	Entire Model
Power with $n = 64$.98	.99	.21

ANS = Approximate Number System. Source: Grinstead et al. 2022, Table 6. Reprinted by Permission of SAGE Publications.

our behavioral scientific questions.⁴ Having laid this foundation for the analytical tool we will use in Experiment 2, let us now turn back to the questions we raised earlier regarding collective-distributive interpretations in child and adult English, their relationships to the lexicon and the degree to which distributive interpretations mediate between the lexicon and collective interpretations.

4. Experiment 1

As we previously alluded to, in Grinstead et al. (2021), both a general measure of lexical development (the Test de Vocabulario en Imágenes Peabody; Dunn et al. 1986) and children's distributive interpretations with *cada* (each) correlated with child Spanish-speakers' collective interpretations with *unos* (some) and *los* (the). However, they further showed that the predictive power of the general lexical measure disappeared when children's interpretations of *cada* (each) were included in the equation. A means of describing this type of statistical relationship is Mediation Analysis. That is, a lexical measure (x) predicts collective interpretations (y). Distributive interpretations (m) also predict y and are predicted by x . The simplest form that a mediation analysis can take is a regression that begins with 1 predictor (x) and 1 outcome (y). Add the second variable (m) as a predictor to create a multiple regression and now, if mediation has occurred, x is no longer predictive. This is the simple version. Grinstead et al. (2021) adopt the more widely used Preacher & Hayes (2008) model, which uses slightly different math, but the intuitive outcome is the same as a multiple regression. There is either a significant mediation effect (i.e., 1 of the predictors is no longer significant when you add a second predictor) or not (both predictors are in the model and each accounts for unique variance). If similar relationships are found for child English, a mediation analysis of this type would be appropriate.

Returning to the findings in adult and child Italian and Spanish, we find a number of lacunae in what is known, which we will attempt to fill by asking the following research questions with respect to another language, English:

- (1) Do collective interpretations in adult English of sentences with definite plural subjects, modified by *the*, and of indefinite plurals modified by *some*, show the 50% rejection in distributive contexts pattern, reported in Pagliarini et al. (2012) for adult Italian, or the near-100% rejection pattern reported in Grinstead et al. (2021) for adult Spanish?
- (2) Does a measure of general lexical development in child English predict collective and distributive interpretations, as it did in Spanish?
- (3) Are collective interpretations of *some* and *the* predicted by interpretations of *each* in collective contexts, as predicted by the Pragmatic Scale Hypothesis?
- (4) If so, does *each* mediate between the lexicon and the collective interpretations of *some* and *the*, as it appears to in child Spanish?

⁴For more on PSEM in the social and behavioral sciences, see Grinstead et al. (2022) and for more on the technical details of PSEM, see Shipley (2000), Lefcheck (2016), and the references previously given. See Lefcheck (2019) for an R package that implements PSEM for different regression types.

4.1. Methods

4.1.1. Participants

A total of 29 monolingual, typically-developing child English-speakers (mean age = 95.4 months [7.8 years-old], SD = 7.65 months) and 22 adult English-speaking university students participated in our protocol and passed our filler items, described in the following section. A parent or guardian of each child participant, and each adult participant, signed a university institutional review board-approved consent form before participating in the protocol.

4.1.2. Procedures

For adults and children, a video-recorded, stop-motion TVJT, following the design essentials of Crain & McKee (1985), including plausible dissent, rich pragmatic context and following the Question-Answer Requirement of Gualmini et al. (2008), was used to measure interpretations of collective and distributive sentences. 6 sentences with each quantifier (one distributive—*each*, and two collective—*some* and *the*) were paired with both a collective and a distributive pragmatic context, represented in the stop-motion videos by actions carried out by the Minions from the movie *Despicable Me*.

The TVJT is an excellent way to measure distributive and collective interpretations because participants' attention can be called to an action in process, which is either carried out distributively, with multiple agents affecting multiple objects, or collectively, with multiple agents affecting one object. This type of action can be harder to represent in static pictures, which may have led to difficulty in interpretation in some of the existing literature. This is the same experiment given in Padilla-Reyes (2018) and Grinstead et al. (2021), with the exception that the audio files were English translations of the Spanish sentences given originally, recorded with the voice of a female native-speaker of U.S. English. The sentences were presented in each of three random orders. There was no significant effect of order of presentation ($p > .05$). Predicates can be inherently distributive, or at least tend towards distributivity, (e.g., The basketball players chose an opponent to guard.) and the same is true of collective predicates (e.g., The movers picked up a piano.). We piloted and chose predicates that were equally acceptable to adult native speakers as distributives, with a distributive quantifier in subject position (*Each minion*), and as collectives, with a collective quantifier in subject position (*Some minions/The minions*):

2. Catch a rooster.
3. Carry a bag.
4. Find a goose.
5. Carry a rooster.
6. Move a rock.
7. Plant a tree.

These 18 sentences (the predicates in 2-7, with *Each minion*, *Some minions* or *The minions* as a subject) were presented in video contexts that were either distributive, as in Figure 1, or collective, as in Figure 2, for a total of 36 experimental sentences.

Filler items were also included to ensure that participants understood the task and were paying attention. There were 12 such sentences, which included subjects modified by *one* or by *no*, as in the following, each of which was presented in contexts in which the sentences were congruent with the contexts and also in contexts that were incongruent with the sentences:

8. One minion could open the door.
9. No minion could open the door.
10. One minion could find a pig.
11. No minion could find a pig.
12. One minion could climb a rock.
13. No minion could climb a rock.

The quantifiers *one* and *no* were chosen for fillers because in our pilot work, even our very youngest children were able to interpret them consistently. To be included in the sample, participants had to answer the filler items at above chance levels, which for 12 items is 10, 11 or 12 correct. 1 adult and 3 child participants were removed for failing to meet this criterion, leaving us with the sample previously described.

There were 4 warm-up items given, which followed the same pattern as the filler items, except that explicit feedback was given after participants responded to orient them to the nature and goals of the task. No feedback was given thereafter.

For child participants, in addition to the TVJT, a standardized English receptive lexical measure, the Peabody Picture Vocabulary Test – 4 (Dunn & Dunn 2007) was given.

4.2. Results

4.2.1. Descriptive statistics

Adults, on average, accepted *some* and *the* in collective contexts almost always (more than 5 out of 6 times, on average) and almost never in distributive contexts (less than 1 out of 6 times, on average), as illustrated in Figure 4. Judgements appear categorical. Similarly, *each* was almost always accepted in distributive contexts and rejected in collective contexts.

Children in this English-speaking sample are less categorical than are adults, as illustrated in Figure 5, where we see that the collectives *some* and *the* in distributive contexts, and distributive *each* in collective contexts, were accepted between 1 and 2 times, out of 6, on average.

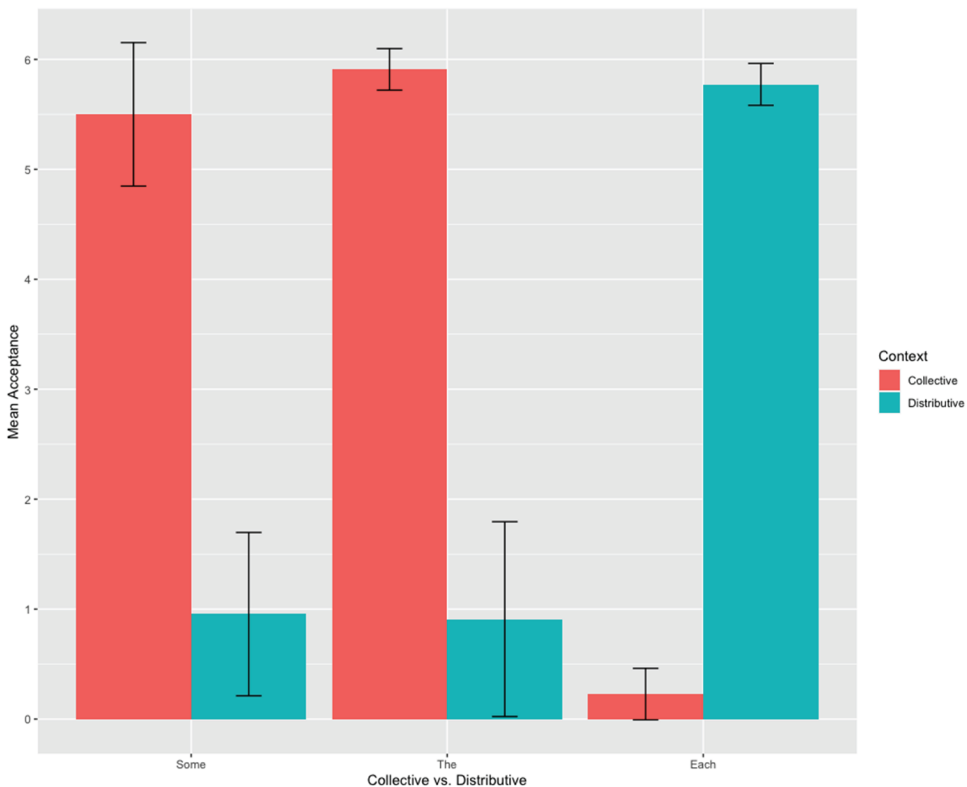


Figure 4. Adult English-speaker acceptance (out of 6 opportunities) of *some*, *the*, and *each* quantifiers in collective and distributive contexts.

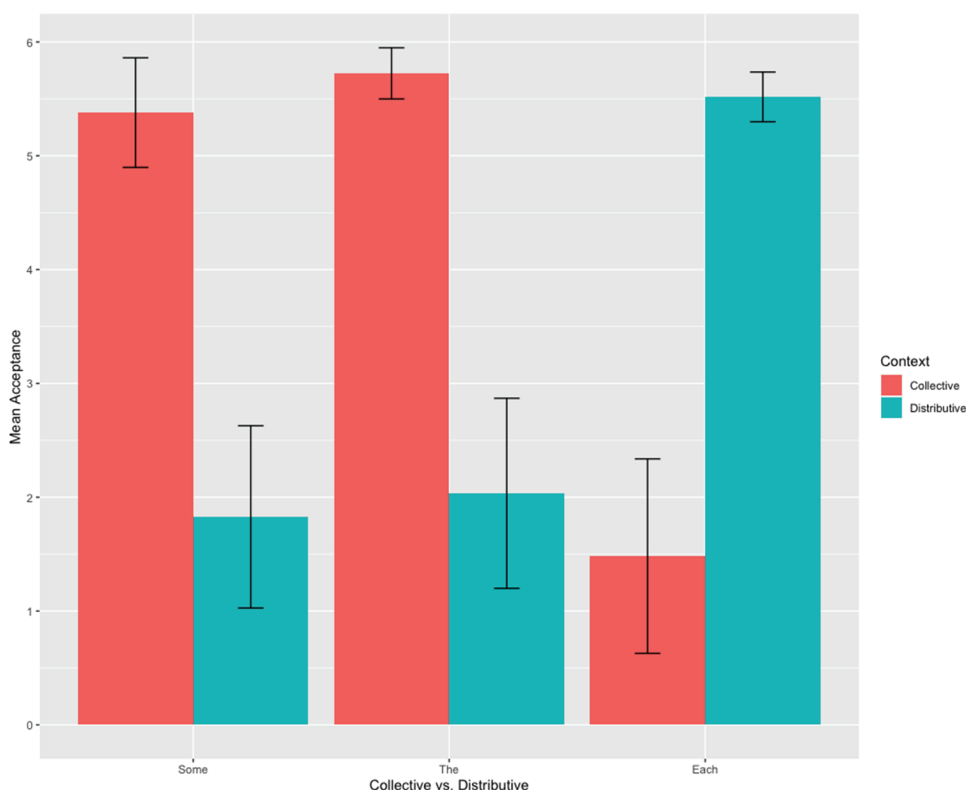


Figure 5. Child English-speaker acceptance (out of 6 opportunities) of *some*, *the*, and *each* quantifiers in collective and distributive contexts.

Finally, children's mean Peabody Picture Vocabulary Test raw scores = 147.97, with a standard deviation of 14.95.

4.2.2. Inferential statistics

To begin with, Table 3 illustrates that our lexical measure primarily associates with quantifier interpretations in contexts that are incongruent (e.g., a distributive quantifier in collective contexts), as reported for child Spanish in Grinstead et al. (2021). Specifically, the PPVT scores are negatively correlated with acceptance. Requiring rejection of a sentence as a characterization of pragmatic context to demonstrate knowledge of the interpretation is one of the primary advantageous design features of TVJTs, in that both children and adults are more predisposed generally to saying “yes” than “no”. Thus, if children are willing to say “no”, this likely arises from strong conviction.

With respect to the mediating role played by judgments of *each* in collective contexts, between the lexicon and the collective judgments of *some* and *the* in distributive contexts, we fit two mediation analyses, following Preacher & Hayes (2008). The rationale for this test is that the distributive entailment of *each* may be driven by the increasing precision of its denotation, which is a function of overall lexical growth, as well as increasingly precise ANS representations, to which we turn in the next experiment. The lack of ambiguity in this distributive entailment, in turn, hypothetically drives the implicature interpretations of the ambiguous quantifiers *the* and *some*, via a pragmatic informativeness inference.

In Figure 6, we see the linear regression showing that (a) lexicon PPVT significantly predicts acceptance of *each* in (incongruent) collective contexts ($B = -.093$, $SE = .024$, p

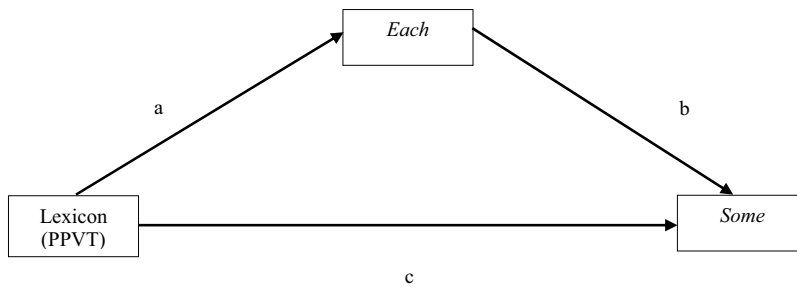


Figure 6. Mediation analysis of interpretations of *each* in collective contexts between the lexicon and implicature interpretations of *some* in distributive contexts.

Table 3. Correlations of Lexicon (PPVT) and Acceptance of Distributive-Collective Quantifiers in Incongruent Contexts.

	Lexicon (PPVT)	<i>Some</i> Distributive	<i>The</i> Distributive	<i>Each</i> Collective	<i>Some</i> Collective	<i>The</i> Collective	<i>Each</i> Distributive
Lexicon (PPVT)		-.597**	-.591**	-.565**	-.356	-.094	-.156
<i>Some</i> Dist.			.890**	.872**	.240	-.097	.313
<i>The</i> Dist.				.902**	.239	.008	.410*
<i>Each</i> Coll.					.222	-.111	.298
<i>Some</i> Coll						.193	.261
<i>The</i> Coll							.330

**Correlation is significant at the .01 level. *Correlation is significant at the .05 level (2-tailed).

< .001). Next, (b) in a multiple regression including PPVT, acceptance of *each* in incongruent contexts significantly predicts acceptance of *some* in (incongruent) distributive contexts ($B = .733$, $SE = .113$, $p < .001$). Following Preacher & Hayes (2008), the product of (a) x (b) (the Indirect Effect) is significant ($B = -.068$, $SE = .019$, $p < .001$). This significance was determined using a percentile bootstrap estimation approach with 5000 samples, executed using the PROCESS macro, Version 4 (Hayes 2018). This indicates that after the mediator, interpretations of sentences with *each* in subject position in collective contexts, is controlled for, our general lexical measure, the PPVT, is no longer a significant predictor of *some* in distributive contexts. The Percent Mediated, or percentage of the total effect accounted for by the mediator is 75.5% and approximately 39% of the total variance in the interpretation of *some* in distributive contexts is accounted for by the predictors ($r^2 = .394$).

In Figure 7, similarly, we see the linear regression showing that (a) lexicon PPVT significantly predicts acceptance of *each* in (incongruent) collective contexts ($B = -.093$, $SE = .024$, $p < .001$). Next, (b) in a multiple regression including PPVT, acceptance of *each* in incongruent contexts significantly predicts acceptance of *the* in (incongruent) distributive contexts ($B = .812$, $SE = .105$, $p < .001$). Following Preacher & Hayes (2008), the product of (a) x (b) (the Indirect Effect) is significant ($B = -.075$, $SE = .021$, $p < .001$). This significance was determined using a percentile bootstrap estimation approach with 5000 samples, executed using the PROCESS macro, Version 4 (Hayes 2018). This indicates that after the mediator, interpretations of sentences with *each* in subject position in collective contexts, is controlled for, our general lexical measure, the PPVT, is no longer a significant predictor of *the* sentences in distributive contexts. The Percent Mediated, or percentage of the total effect accounted for by the mediator is 80.6% and approximately 39% of the total variance in the interpretation of *the* in distributive contexts is accounted for by the predictors ($r^2 = .391$).

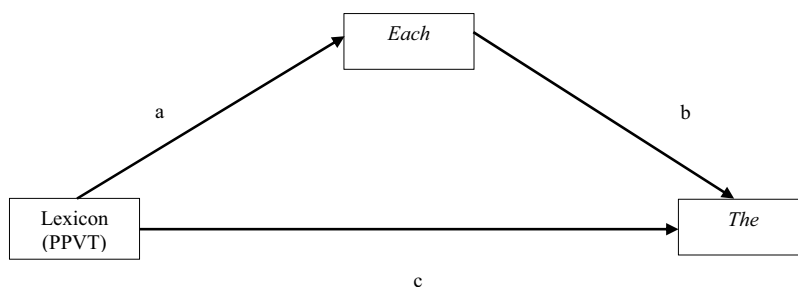


Figure 7. Mediation analysis of interpretations of *each* in collective contexts between the lexicon and implicature interpretations of *the* in distributive contexts.

4.3. Discussion

Returning to our research questions, we see that adult English-speakers, at least on our TVJT measure, appear to have categorical judgments of not only definite plural *the* as collective, as in Spanish, and in contrast to the 50% acceptance results in Italian of Pagliarini et al. (2012), but also have categorical, near 100% collective interpretations of *some*. We hasten to point out that the difference is not the predicates we used, which are ambiguous between collectivity and distributivity, as evidenced by the fact that, in our within-subjects design, participants judged the exact same predicates to be categorically distributive when the subject noun phrase (NP) was modified by the distributive quantifier *each*.

Next, we see that a general measure of lexical development, the PPVT, was predictive of both collective *the* and *some* sentences, presented in distributive contexts, and also of distributive *each* sentences, presented in collective contexts. This appears consistent with the Lexical Refraction Hypothesis in that general lexical development drives greater precision of denotative content as more lexical items are added to the overall lexicon.

In addition, we see in Table 3 that interpretations of *each* in incongruent collective contexts were highly correlated with interpretations of *the* (.87) and of *some* (.90) in incongruent distributive contexts. We take the closeness of these statistical relationships to be strongly supportive of the Pragmatic Scale Hypothesis.

Finally, we see that though general lexical development appeared to be driving the collective interpretations of *some* and *the*, in fact, they were really driving development of the distributive entailment of *cada*, which drained the predictive power of the PPVT from the equation predicting the *some* and *the* implicature.

In sum, the results from Spanish reported in Grinstead et al. (2021) were not unique. Adult English also appears to have categorical judgments of sentences with *the* and *some* in subject position. Furthermore, the English lexicon also appears to have relationships among its quantifiers such that they are structured into a pragmatic scale of distributivity-collectivity, with the unambiguous, distributive extreme of this scale being anchored by the distributive entailment of *each*. The interpretations of at least the plural generalized quantifiers *the* and *some*, though they are potentially underspecified between distributivity and collectivity, come to have collective interpretations via a Gricean informativeness judgment, to the effect that if a cooperative conversation participant wanted to be as informative as necessary, they would choose to use *each* to express distributivity, as it is most informative by virtue of being most specified.

5. Experiment 2

In experiment 1, we saw a replication of existing work in other languages on the development of collective-distributive interpretations, showing that lexical development predicts collective and

distributive interpretations and that distributive interpretations mediate between lexical development and collective interpretations. Also, our sample of 7-year-olds was not adult-like in their interpretations, returning us to one of the core mysteries of collective-distributive interpretations, namely, why it is that they take so long to develop. Clearly lexical development plays a role, but that cannot be all there is to the story. The sentences that carry the quantifiers must be syntactically composed. Is syntactic development also part of what slows down development, or is it at least relevant? What about executive function? Does development in this domain of cognition somehow relate to collective interpretations? Finally, does the non-linguistic number ability relate to development of these quantificational expressions? It seems a priori likely to matter. Perhaps one of these domains of cognition is especially problematic or slow to develop. Though we will not be able to say much about the rate of development of each of these subdomains of cognition in our cross-sectional study, we can at least, in principle, determine whether all of these variables matter, which brings us to SEM. Given the PSEM analytical tool we previously described, we are in a position to measure all of these abilities in one sample of children and to test whether a model of sentence interpretation involving them shows goodness of fit, statistically. We now turn to a consideration of each of these sub-domains of cognition and their relationship to collective-distributive interpretations.

5.1. Lexical refraction and ANS

The Lexical Refraction Hypothesis, previously alluded to, claims that natural language quantifiers, and in fact all number-related morphemes in natural language, draw their fundamental sense from ANS representations that are then refracted in language-particular ways and in morpheme-specific ways to convey distinct numerical meanings. One way to illustrate the non-lexical versus the lexical use of this quantity information is with two of the Number Line Estimation tasks of Siegler & Opfer (2003). These are tasks that are used to measure what children know about both approximate estimates of number and exact, cardinal estimates of number. Specifically, children are presented with a horizontal line, one extreme of which has, for example, the Arabic numeral 0 at the left edge and at the other edge, the Arabic numeral 30. In the middle of the screen, another Arabic numeral appears, say 8, and children are asked to click with a mouse where on the line they believe 8 belongs between 0 and 30. Siegler & Opfer (2003) refer to this as the “Bounded Symbolic, Number-to-Point” version of their task, which is illustrated in Figure 8.

Siegler & Opfer (2003) show that as children age, their estimates move from being a natural log function of the number they are asked to situate on the number line, to being a linear function of this estimate. More linear estimates of these numbers correlate with results of the Give-a-Number task of Wynn (1990, 1992), which is standardly assumed to be evidence that children have grasped the Cardinality Principle (Gelman & Gallistel 1978), which is the property of the counting process by which the last number in the count routine reflects the exact cardinality of the set of items that has been counted, up to that point (Opfer et al. 2019).

The Arabic number, graphically given to children to situate on the number line is not any kind of direct, visual representation of quantity, but rather is a grapheme that corresponds to the phonological representation of a lexical item in the child’s lexicon. What is the information that serves as the denotative content of these exact numeral quantifiers? It seems conceptually inevitable that the

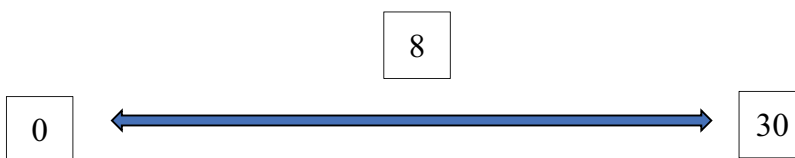


Figure 8. A schematic example of the number-to-point, bounded, symbolic number line estimate task (Source: Siegler & Opfer 2003).

representations that feed the lexicon must come from the part of the mind that our species shares with most other species, known as the ANS, which allows the representation of numerosities and magnitudes (e.g., Agrillo et al. 2012, Dehaene 1997, Feigenson et al. 2004, Gallistel 2011, Xu & Spelke 2000). Siegler & Opfer (2003) have another version of their number-to-point, bounded number line estimation task, (the non-symbolic version) which does not use Arabic numerals, which is to say, it lacks critical use of the natural language lexicon. Instead of Arabic numerals, a box, with a number of dots in it, is used to represent each of the extremes of the number line, as well as the magnitude to be situated on the number line, as in Figure 9. Notice that the only difference between these two tasks is the use of lexicon-based Arabic numerals in Figure 8. Thus, whatever differences there may be between a single child's performance on these two tasks, must be, at least in large part, due to the differences between ANS representations of quantity and lexical representations of exact cardinality.

In recent work, Grinstead, Nieves-Rivera et al. (2020) show that though a sample of children showed a slight correlation between the results of these two tasks ($r = .315$, $p = .002$, $n = 97$), this correlation disappeared in a mediation analysis that included the Peabody lexical measure, which produced a significant indirect effect and 75.5% percent mediation. These results are partially corroborated by Negen & Sarnecka's (2012) demonstration that cardinality, measured by the Give-a-Number task, correlated with both expressive and receptive measures of lexicon. In this way, their study replicates the second half of our mediation equation, so to speak, between lexicon and cardinality.

We take Siegler & Opfer's (2003) findings to mean that the non-species-specific ANS ability produces a type of quantity sensitivity that develops and becomes more linear as children mature. This sensitivity feeds the semantic number dimension of morphemes in the natural language lexicon, which includes cardinal numerical quantifiers, such as those linked to a phonological representation that is in turn linked to a graphical representation used in the Symbolic Number Line Estimation task. We see no reason to assume that this same ANS representation should not also underly all other numerical morphemes in the natural language lexicon, including number marking morphemes on nouns, adjectives, determiners, and verbs in Spanish, and the universal and existential quantifiers, both distributive and collective, as well as intersective existential quantifiers, such as *algunos* (some) in Spanish.

5.2. PSEM and the quantity implicature

Recent work on the interpretation of *algunos* (Grinstead et al. 2022), claims that ANS, as measured by the Non-symbolic Number Line Task, predicts a lexical measure, together with the executive function ability of inhibition (more on this in a subsequent section) and then, together with a measure of syntax and a second use of inhibition, predicts child Spanish-speakers' "some, but not all" Quantity Implicature interpretations of sentences with subjects modified by *algunos*. Because simple mediation was not sufficient to account for this many variables, which the authors argue are important for understanding how the linguistic and cognitive system yields these interpretations, the authors use what they refer to as "Piecewise Structural Equation Model." The principal virtue of this statistical measure, which stems from work in the Life Sciences field of Ecology, is that statistical power is calculated for each local regression and not for the global variance-covariance matrix, as it is in conventional SEM (see, e.g., Wolf et al. 2013 and Kline 1998). This fact implies that smaller samples ($n = 64$ in the study referred to above) can still yield high statistical power (.98 and .99, according to the authors, Table 2) as opposed to what a globally-calculated SEM would have yielded (.21). The model fit in this report is given in Figure 10.

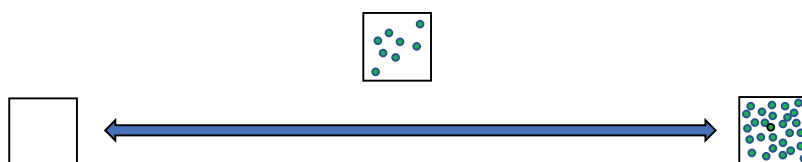


Figure 9. A schematic example of the number-to-point, bounded, non-symbolic number line estimate task (Siegler & Opfer 2003).

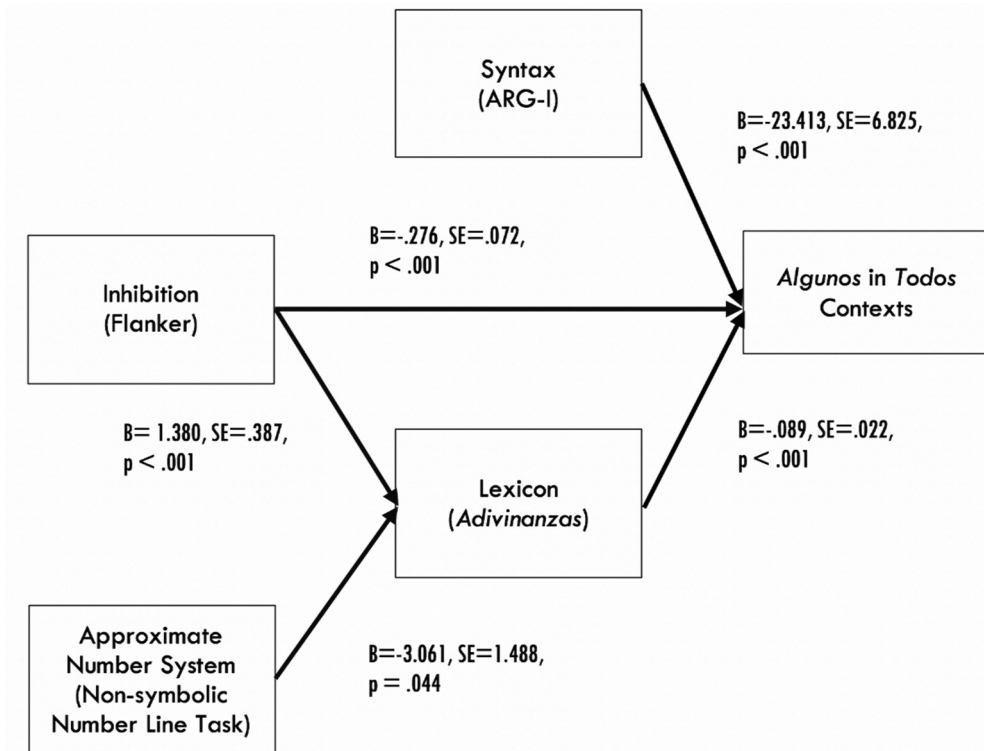


Figure 10. Syntax, lexicon, and inhibition predict *algunos* interpretations in contexts where *todos* (“all”) would be correct; inhibition and approximate number predict lexicon: Fisher’s $C = 2.783$, $p = .595$, $df = 4$, $AIC = 18.783$. (Source: Grinstead et al. 2022, Figure 8, Reprinted by Permission of SAGE Publications.)

Of course, the Quantity Implicature differs from the Collective Implicature considered here, in that the 4- and 5-year-old children whose data is modeled in Figure 10 already had adult-like representations of the meaning of *todos* (“all”), which were in fact required as filler items. In contrast, the children in the English language study we have reported in Experiment 1, did not yet have adult-like representations of *cada/each*, which serves a similar scale-anchoring function to *todos/all* on the Pragmatic Scale Hypothesis view. This fact is important because in the case of the *algunos* implicature modeled in Figure 10, there is no specific variable that shows the impact of children’s knowledge of *todos*, at the extreme edge of the Quantity Scale. In contrast, we have just modeled for child English that knowledge of the distributive entailment of *each* is not adult-like and that children’s knowledge of the collectives *the* and *some* can be directly calculated as a function of their *each* knowledge. The SEM model we propose to account for distributives and collectives will need to include a measure of children’s grasp of *cada/each* in collective contexts to give a full account of the phenomenon.

5.3. Inhibition

Figure 10 shows for the *algunos* implicature that inhibition plays a role, both at the level of disambiguating the sentence carrying *algunos*, as well as at the level of predicting lexical development. We follow Miyake et al. (2000) in assuming that there are at least 3 major subdomains of executive function, including attention, auditory working memory⁵ and inhibition. One very well-studied dimension of domain-general, inhibitory cognition is its role in the development of the lexicon

⁵See Shah & Miyake (1996) for convincing work, distinguishing visual and auditory working memory, storage, and processing.

(Caramazza 1997, Luce & Pisoni 1998, Costa et al. 1999).⁶ An array of developmental studies demonstrate that inhibition and lexicon are statistically associated in development, following the conceptual argument that the more competitors there are in a child's lexicon to express or comprehend a concept, the more inhibition will be required to suppress the incorrect competitors (e.g., Gangopadhyay et al. 2019, Larson et al. 2020). Thus, it seems clear that inhibition at least plays a role in lexical development.

Another role played by inhibition in studies of natural language processing (see, Novick et al. 2005, for a review) is in the suppression of irrelevant interpretations of entire sentences that are ambiguous. Figure 10 illustrates that a simultaneous role for inhibition is possible in the interpretation of sentences in which listeners have to choose from various interpretations, such as those carrying a quantity implicature. Notice that a similar type of ambiguity is present in collective implicature sentences and thus, it is possible that inhibition could play a role in the interpretation of these sentences, as well.

5.4. Syntax

A fundamental distinction between the cardinal number quantifiers used in the count routine and quantifiers discussed thus far for collective-distributive interpretations (*each, some, the*) and for partitive interpretations (*some, all*) is that the latter quantifiers have been tested in sentences, while numerical quantifiers have been primarily studied, at least in developmental psychology, as isolated elements in the count routine. Because counting requires a unit set, or thing counted, we assume that the numbers in the count routine are non-clausal noun phrases or determiner phrases that include both an expressed, or unexpressed noun, corresponding to the unit set, as well as a degree quantifier (following Kennedy & Syrett 2022). This turns out to be an important distinction to make because quantified noun phrases in the count routine have surprisingly different semantic properties than do quantified noun phrases in sentences (Grinstead et al. 1998). Compare the identical noun phrases in bold in 14 and 15.

14. one penguin...two penguins... **three penguins**...
15. I see **three penguins** on the ice flow.

Quantified NPs in sentences, such as *three penguins* in 15 are classic examples of indefinite, existential NPs. They are indefinite because 15 is compatible with a situation in which there are many more than three penguins, but the speaker only commits to seeing three of them. They are taken to be specific to the speaker only. In contrast, in 14, by virtue of the One-to-one Principle (Gelman & Gallistel 1978), having counted 3 penguins implies having matched an individual penguin to each step in the count routine and having partitioned it off from re-counting, usually with some sort of deictic gesture. The act of counting conceptually requires that each step in verbal counting refer to a unique, countable thing, making each step correspond to not an indefinite, but rather to a definite noun, in the spirit of Russell's (1905) Uniqueness Presupposition. Furthermore, while the penguins in 15 are also classical examples of existential quantification, in that they are exemplars of a domain that may or may not include all penguins under consideration, *three penguins* in 14 are all of the penguins under consideration at that step of the counting process, which makes them not existential, but rather universal in their quantification, as argued for definite NPs marked with *the* by Chomsky (1975). Thus, numerically quantified NPs are existential indefinites in the clause, but universal definites in the count routine. An obvious difference between these two types of uses of numerically-quantified NPs is the use of syntax in 15 to concatenate the NP with the verb and with a prepositional phrase "on the ice flow" to create a verb phrase, and finally to merge this verb phrase with the subject NP "I" to create a full sentence.

⁶For a review of the role of auditory working memory in lexical development, see Houston et al. (2020).

Given the critical role played by syntax in composing distributive and collective interpretations from the assembled lexical items, it is important to have an independent measure of syntax to include in our model. The most well-known developmental measures of morphosyntax, Mean Length of Utterance and Mean Length of T-Unit, incorporate both syntax and lexicon. This inclusion of lexicon in syntactic measures is very difficult to avoid, but is conceptually possible, if we abstract away from measures that depend upon the number of morphemes or lexical items per utterance/sentence for their validity. In recent work, Grinstead et al. (2021) showed that the traditional Subordination Index (LaBrant 1933, Loban 1963), which takes the total number of subordinate clauses in a spontaneous production sample over the total number of clauses, could be further divided into an Adjunct Index and an Argument Index. The Adjunct Index takes the total number of adjunct subordinate clauses over the total number of clauses and the Argument Index takes the total number of argument subordinate clauses over the total number of clauses. The authors give evidence that each index correlates with measures of lexicon and morphosyntax, but not the same measures and not with each other. They also show that the Argument Index correlates with auditory working memory and the Adjunct Index does not.

5.5. Summary and research questions

In summary, there is evidence that general lexical development is predictive of distributive *each/cada* interpretations, which are in turn predictive of collective *some/unos* interpretations. Furthermore, ANS and inhibition appear predictive of general lexical development. This fact is consistent with the Lexical Refraction Hypothesis and with what appears to be the consensus concerning the association of inhibition and lexicon in development. Syntax would appear to play a conceptually necessary role in conjunction with the lexicon in composing collective-distributive sentences, though it is not known how this will play out with *each/cada* interpretations in the model. Finally, inhibition appears to play a dual role in *algunos* interpretations in Quantity Implicature sentences, predicting both lexical development and the resolution of ambiguous sentences. It might play a similar role with respect to the ambiguity inherent in collective sentences. These considerations lead us to the following research questions:

1. What relationship does inhibition have to lexicon and *unos* interpretations in collective and distributive contexts?
2. Does syntactic development play a significant role in these interpretations?
3. Does ANS predict lexical development, as in *algunos* interpretations, following Lexical Refraction?

5.6. Methods

5.6.1. Participants

A total of 57 monolingual Spanish-speaking children (mean age = 80.0 months [6.66-year-olds], SD = 14.2, range = 50–101 months) participated in our protocol and passed our filler items. Filler items included the Spanish version of the same sentences, 8–13, in Experiment 1, with *todos* (all) and *ningún* (none) quantifiers used in subject position, presented in both congruent and incongruent contexts. Participants had to answer 10, 11, or 12 of the 12 items correctly to be statistically above chance and remain in our experimental group. A total of 26 children and no adults were removed from the sample for this reason. The children removed were primarily in the 4-year-old group. We attribute this attrition to the length of the experiment, but because the construction does not become adult-like until well-after even the age of our oldest participants, we were not concerned with the appropriateness of the task for the rest of the children. Rather, we were primarily concerned with the variance in their scores and its relation to the additional variables we tested.

5.6.2. *Procedures*

Participants took the TVJT, as previously described; however, this time audio was recorded by a female native-speaker of the variety of Mexico City Spanish spoken by the children in the sample. Also, in contrast to the TVJT as previously described for English, this experiment only used the *cada* (each) condition and the *unos* (some) condition, to make it shorter to allow time for our other measures. Our measure of ANS was the Number-to-Point, Bounded, Non-symbolic Number Line Estimation Task of Siegler & Opfer (2003), as previously described. Our lexical measure was the Number of Different Words (NDW) measure (Miller 1991), which is a calculation, from spontaneous production data, of the number of unique words used by a child during that session. To make these most comparable, the spontaneous production data was gathered using the Frog Story paradigm, during which children narrate the *Frog, Where Are You?* book, by Mercer Mayer (1969). Each Frog Story speech sample was transcribed by native-speakers of the Mexico City dialect of Spanish and was checked for reliability iteratively, following Castilla-Earls et al. (2015). This sample was used as the basis for the NDW calculation. The same iterative procedure was followed for transcribing and coding a second spontaneous speech sample that was minimally structured and was produced in response to questions about children’s school day routine, their vacations, friends, and family, roughly following the sociolinguistic interview guidelines of Labov (1984). This sample was used as the basis for our syntax measure, the Argument Index, as previously described. To measure inhibition, children were given the pediatric version of the Flanker Task (Eriksen & Eriksen 1974), implemented as part of the EXAMINER Battery (Kramer et al. 2014).

5.7. *Results*

5.7.1. *Descriptive statistics*

Table 4 gives the means and standard deviations for each measure for our sample.

5.7.2. *Inferential statistics*

The results of the children in our sample with respect to acceptance of collective-distributive sentences parallel those in Grinstead et al. (2021), which was done with monolingual Spanish-speaking children in Puerto Rico. Table 5 shows the Pearson Product Moment Correlations of our study variables.

Table 4. Means and standard deviations of protocol measures.

	Syntax (Arg. Index)	Lexicon (NDW)	ANS (Number Line)	Inhibition (Flanker)	<i>cada</i> Coll.	<i>cada</i> Dist.	<i>unos</i> Coll.	<i>unos</i> Dist
Mean	.036	113.719	.715	5.846	1.211	5.439	3.579	2.561
SD	.022	18.589	.414	0.771	1.645	0.982	2.244	1.861

Table 5. Pearson correlations of study variables.

	<i>unos</i> Dist.	Lexicon (NDW)	Syntax (Arg. Index)	ANS (Number Line)	Inhibition (Flanker)
<i>cada</i> Coll.	.410**	-.441**	-.328*	.214	-.303*
<i>unos</i> Dist.		-.402*	-.039	.133	-.095
Lexicon (NDW)			.249	-.317	.459**
Syntax (Arg. Index)				-.294*	.115
ANS (Number Line)					-.221

NDW = Number of Different Words; ANS = Approximate Number System. **Correlation is significant at the .01 level. *Correlation is significant at the .05 level.

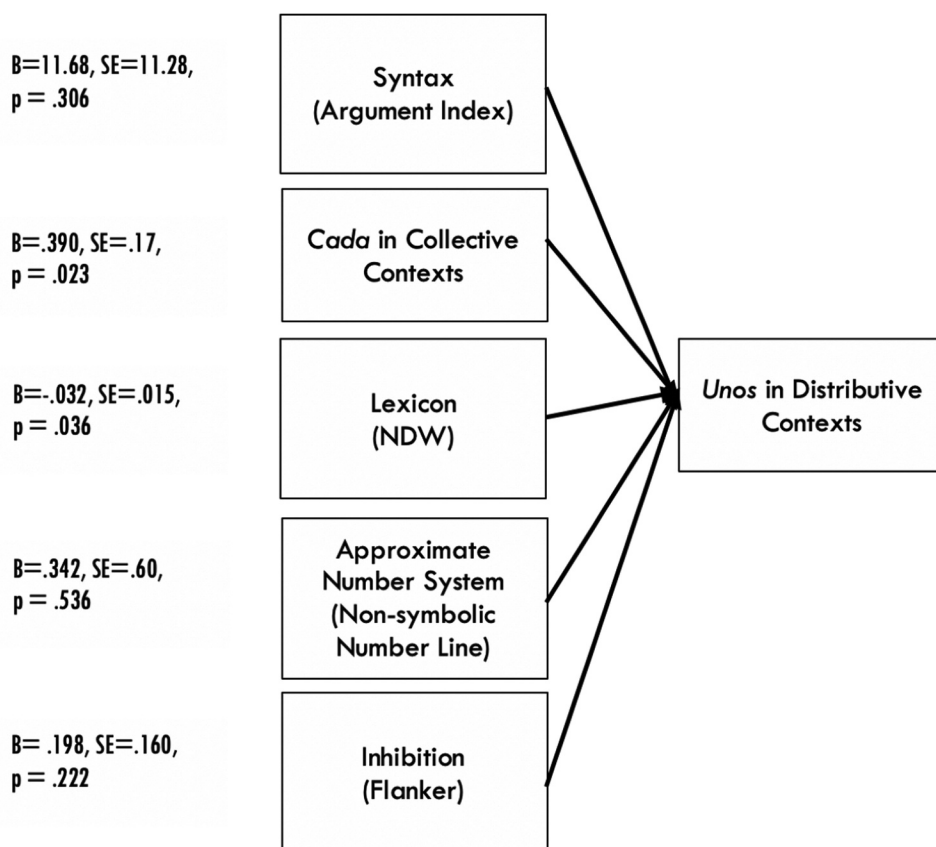


Figure 11. Flat-structured multiple regression model of collective implicature interpretations. Only *cada* ($B = .195$, $SE = .082$, $p = .039$) and Lexicon ($B = -.032$, $SE = .01$, $p = .036$) are significant. AIC = 158.35.

A standard follow-up analysis to the correlation is to sort out collinearity through a flat-structured multiple regression, such as the one depicted in Figure 11.

As we can see in Figure 11, there appears to be no significant role played by ANS, inhibition, or syntax in the interpretation of *unos* in distributive contexts, contrary to what one might assume to be necessary conceptually. In contrast, in the following PSEM, given in Figure 12, we see that each of these variables can be shown to be playing a significant role in the ultimate judgment of *unos* (“some”) in subject position in collective sentences, presented in distributive pragmatic contexts, once they are situated in a theoretically sensible relationship to the other relevant variables.

This model was calculated using the “PieewiseSEM” package in R, developed by Lefcheck (2019), following Shipley (2000, 2002, 2009).⁷ The results show a C-statistic of 9.38, with a p-value of .67 and on 12 degrees of freedom, based on the three independent regression equations that compose this structural equation model.⁸ The regressions inside the model follow our intuitions. The negative coefficient between Lexicon and *unos* means that the greater one’s vocabulary, the less likely one is to accept a collective quantifier (*unos*) in a distributive sentence. Similarly, the coefficient between *cada*

⁷See Lefcheck (2016) and Grinstead et al. (2022) for more on pieewise structural equation modeling in the social and behavioral sciences.

⁸An anonymous reviewer pointed out that the *unos* and *cada* outcome variables have 7 possible outcomes each, which they argued should not be treated as a continuous variable. In fact, with this many items, categorical variables can function indistinguishably from continuous variables, even in structural equation models, as demonstrated by Rhemtulla et al. (2012). In the Appendix, we provide a PSEM with categorical (Poisson) regressions used for the *cada* and the *unos* outcome variables, which yields very similar goodness of fit measures and significance values. The model-wide Akaike’s Information Criterion values are also virtually identical: 32.12 for Poisson vs. 33.38 for linear.

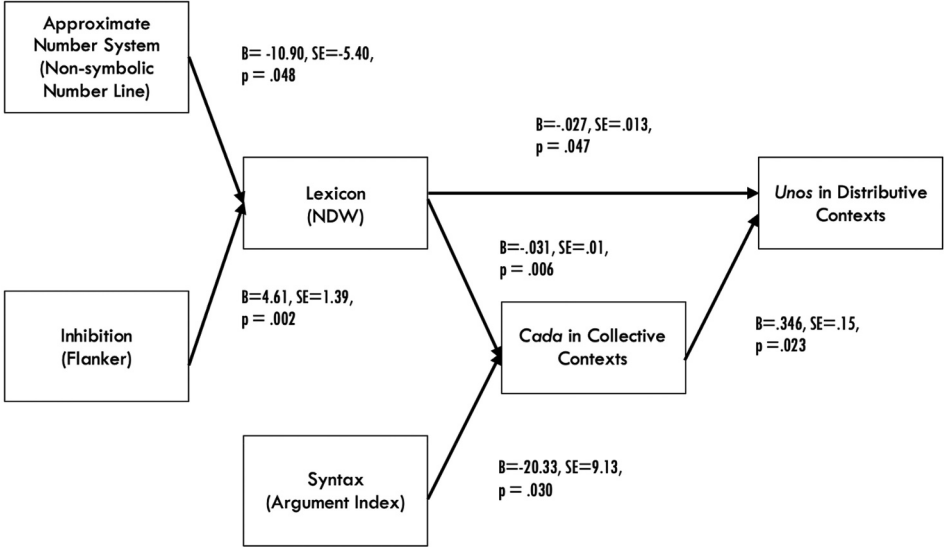


Figure 12. Piecewise structural equation model: Fisher’s C-statistic = 9.38, $p = .67$, 12 degrees of freedom. No significant alternative endogenous variable-predicted paths. AIC = 33.380. A data file (.csv) and an R script to reproduce the results in Figure 12 are available here: <https://osf.io/xv2pj/>.

and *unos* is positive, which means that the less likely one is to accept a distributive quantifier (*cada*) in collective contexts, the less likely one is to accept a collective quantifier (*unos*) in distributive contexts, as predicted by the Pragmatic Scale Hypothesis, and consistent with previous findings in the literature. The negative coefficients of the regression of Lexicon and Syntax on *cada* mean that the larger one’s vocabulary and the larger one’s ability to use argumental subordinate clauses, the less likely one is to accept a distributive quantifier in collective contexts. Finally, the positive coefficient of Inhibition on Lexicon means that the greater one’s ability to ignore irrelevant, but competitive, words for specific concepts, the greater one’s lexicon. Also, the negative coefficient between ANS and Lexicon means that the lower one’s score on the Nonsymbolic Number Line Task, the greater one’s vocabulary. This is because a lower score on Siegler & Opfer’s (2003) Lambda Score is a more adult-like (more linear and less logarithmic) estimation of non-discrete quantity.

Returning to the C-statistic, we remind readers that the null hypothesis for this test is that there are no other significant configurations of the endogenous variables tested. If there were such configurations, we would expect a lower C-statistic and eventually a significant p-value, indicating a distinct

Table 6. Tests of directed separation: Tests whether 3 endogenous variables (Lexicon, CadaC, and UnosD) can be predicted by alternative paths.

Independence claim	DF	Coefficient	p-value
Lexicon ~ Syntax + Inhibition + ANS	52	0.2966	0.7679
unosD ~ Syntax + Lexicon + cadaC	53	0.9378	0.3526
cadaC ~ Inhibition + Syntax + Lexicon	53	-1.0033	0.3203
unosD ~ Inhibition + Lexicon + cadaC	53	1.1930	0.2382
cadaC ~ ANS + Syntax + Lexicon	52	0.5879	0.5591
unosD ~ ANS + Lexicon + cadaC	52	0.2607	0.7954

ANS = Approximate Number System. All Are non-significant ($> .05$).

Table 7. Power analysis comparison of each local regression in the piecewise structural equation model vs. a global SEM of collective implicatures.

Locally-estimated:			Globally-estimated
<i>unos</i> Regression (<i>cada</i> + Lexicon)	<i>cada</i> Regression (Lexicon + Syntax)	Lexicon regression (Inhibition + ANS)	Entire model
n = 57	n = 57	n = 56	n = 56
.92	.96	.98	.10

SEM = structural equation modeling; ANS = Approximate Number System.

causal structure from the one hypothesized. As it stands, the alternative configurations tested (the Tests of Directed Separation, given in Table 6) are non-significant, consistent with the hypothesis that this causal model, and not others, is explanatory.

To emphasize the utility of this test, in comparison with globally-computed SEMs that employ a single variance-covariance matrix, Table 7 gives the comparative statistical power for the effect we have found, given our sample size and degrees of freedom.

5.8. Discussion

What does all this mean? The model in Figure 12 implies that an array of linguistic and non-linguistic abilities, all tested in individual children, plays a role in the interpretation of what is hypothesized to be a collective implicature. The good fit of this model is consistent with Dotlačil's (2010) hypothesis that the implicature interpretation is drawn because a more informative lexical item exists in the lexicon for conveying the concept of universal quantification, separated out distributively, from the predicate to the atomic members of the set represented by the subject.

5.8.1. Utterance-level informativeness

The model's significance also helps us to evaluate other hypotheses. To begin with, neither syntax nor inhibition appear to directly play a role in the composition and disambiguation of *unos* collective sentences in distributive contexts, as it did with *algunos* in *todos*-type sentences. This can be seen in the second and fourth independence claims in Table 6, which include syntax and inhibition, respectively, as predictors of *unos*. At one level, this is surprising because the constructions seem parallel in the relationships of the implicated interpretations of one of the underspecified quantifiers on the scale (*algunos* and *unos*) to the entailed meaning of the scale anchor quantifiers (*todos* and *cada*).

At another level, perhaps more surprisingly, the model appears to imply that the *unos* implicature is not the product of the interaction of the lexicon and syntax, but rather of the interaction of the lexicon with children's interpretations of fully-specified distributive sentences. Syntax is predictive of the distributive *cada* sentences and then the *cada* sentences + the lexicon yields the interpretation of *unos* collective sentences. One way to make sense of this result is to conceive of elements on a lexical scale, independent of syntax, providing information to be judged as pragmatically felicitous or infelicitous. How can an element on a lexical scale that is not composed by syntax—at least not statistically significantly so—constitute a sentence, capable of serving as a pragmatically evaluable utterance? This latter point is particularly intriguing when considering the mainstream generative grammar syntax-lexicon dichotomy we have assumed, throughout our project, and demonstrated in the non-correlation between our syntax and lexical variables. It seems odd to think about a lexical element, without the involvement of syntax, competing with a syntactically-composed sentence for informativeness. Yet this seems to be the conclusion to which this model drives us. This conundrum leads naturally to the question of how much of a sentence can be lexically stored.

Notice that other generative frameworks, including Head-Driven Phrase Structure Grammar (e.g., Pollard & Sag 1994), Simpler Syntax (Culicover & Jackendoff 2005), among others, assume that there

are ontological linguistic categories in between strictly freely-combining syntactic constructions and lexically-stored, unitary lexical items. Among them, we find inheritance hierarchies, constructional idioms and others, which have some systematic freedom to combine with other lexical items and some fixed properties, phonological, or semantic, that may not vary at all. A conjecture we would like to now entertain is that it is at least conceivable that entire distributive-collective constructional idioms, as opposed to simple quantifiers, could form a scale, from which the collective implicature interpretations is drawn, as a function of Gricean informativeness. Following Culicover & Jackendoff (2005), for example, constituents of the syntactic form in 16, which schematize the sentence in 17, would have variable slots for 2 countable nouns and a transitive, inflected verb.

16. $[_{IP} [_{NP} \text{unos } [_N]] [_{VP} [_V] [_{NP} \text{un/a } [_N]]]]$
 17. Unos minions movieron una piedra.
 Some minions moved a rock.

This collective constructional idiom would carry the plural existential quantifier in subject position and the singular indefinite quantifier in object position. The collective interpretation of this constructional idiom would be drawn by its Gricean informativeness relationship to a freely-composed distributive sentence with the distributive *cada* quantifier in subject position and an indefinite quantifier in object position. The Gricean notion of “pragmatic scale” is not traditionally conceived of as expressing a relationship between distinct ontological categories, such as syntactically-composed sentences (the *cada* distributives) and non-syntactically-composed sentences, such as the constructional idiom *unos* collective sentences we are contemplating here. Is there a way to reconcile this syntax vs. lexicon view?

5.8.2. Syntactic composition and lexical storage

What if the distributive *cada* sentences are *both* syntactically-composed *and* also residents of the lexicon as a constructional idiom, as we have proposed for the collective sentences? The representation would be parallel to the one in 16, as in 18.

18. $[_{IP} [_{NP} \text{cada } [_N]] [_{VP} [_V] [_{NP} \text{un/a } [_N]]]]$
 19. Cada minion movió una piedra.
 Each minion moved a rock.

On this speculative account, there would be both a syntactically-composed version of 19, as well as a constructional idiom version of it. Following at least some thinking about constructions at the morphological level (e.g., Baayen et al. 2001), we would not want to claim that *cada* can never be syntactically merged into an NP that acts as a subject, direct object, object of a preposition, among others, but rather speculate that it may be used compositionally, as well as constructionally. The constructional use we have discussed here may be more prominent or higher frequency than other uses of *cada*, which is relevant in Baayen et al.’s (2001) thinking about “base rate” and “surface rate” reaction times to Dutch inflection. The sentence types in 17 and 19 have certainly been the most-studied, on our reading of the literature, though we have no independent evidence of its frequency or prominence. These and other empirical consequences of our speculation will await further investigation.

But isn’t such an explanation ad hoc? One reason for thinking that it is not has already been presented in the form of the significant goodness-of-fit of the causal model represented in Figure 12. A model more like the compositional *algunos* version in Figure 10 showed poor fit. Perhaps more interesting is the fact that something like a constructional account has always been needed of this phenomenon, for independent reasons. Specifically, sentences such as 1 (repeated here as 19) can only have a distributive meaning, not a collective one. This obligatory distributive interpretation of such sentences has always been problematic for compositional generative accounts of collectivity and

distributivity in terms of a “Scope Analysis.” The generative Quantifier Raising (May 1985, Chomsky 1995) scope analysis of the collective and the distributive interpretations of the sentence in 20 claims that the two interpretations are a function of whether syntax has continued to operate after phonology. On this account, the indefinite direct object, *un pupazzo di neve* (a snowman), may syntactically move above the subject, after the moment in the syntactic derivation that phonology has expressed/received the phonological representation. If it does so, then the quantified object has “wide-scope” relative to the subject and the resulting meaning is collective (e.g., “There is one snowman, such that the girls are building it”). In contrast, if the direct object indefinite remains in situ, then it is said that it has “narrow scope” with respect to the subject, in which case the interpretation is taken to be distributive (“For each girl there is one snowman, such that the girl is building it”).

20. Le bambine costruiscono un pupazzo di neve.

The girls are building a snowman.

The problem is that if Quantifier Raising is the core syntactic process that yields collective vs. distributive readings of sentences, then it is unclear why it cannot apply to the sentence with *cada* in 19. That is, Quantifier Raising accounts have always had to make a lexical exception to sentences that have *cada/each* in subject position, which disallow a collective interpretation, which is why 19 cannot have a collective meaning. Such constructions have been referred to as examples of “Scope Freezing.”⁹ We see, then, that an idiosyncratic, lexical exception must be made to the Quantifier Raising analysis to allow it to account for the distributive-only interpretations of sentences such as 19. Since we inevitably must make a lexical exception to explain 19 and since language independently has to have constructional idioms to account for the broad range of phenomena accounted for in the literature, why not assume that the lexical item in question is not the single distributive quantifier, but rather the distributive and collective constructions themselves?

Returning to Figure 12, it is worth noting that we also find commonalities with the PSEM predicting *algunos* interpretations in Figure 10. Namely, the results showing inhibition and ANS predicting the lexicon are consistent with the Lexical Refraction Hypothesis. This is particularly noteworthy because the lexical measure used here is Number of Different Words, calculated from a Frog Story transcript, and not the standardized, elicited production *Adivinanzas* measure used for the *algunos* model. The fact that two distinct measures of lexicon, across distinct models, were each predicted by ANS and inhibition reduces the likelihood that this result is somehow epiphenomenal of the task demands of each measure. Rather, it seems that multiple measures of lexicon do in fact measure general lexical size as a function of how much approximate numerical magnitude information a child can manipulate and how well they are able to inhibit potential lexical competitors for concepts. As we said earlier, the ANS dimension of the general lexicon in this model seems particularly intuitive for Spanish, in which most words have some type of number information encoded on them.

5.8.3. Combinatorics of multi-quantifier sentence interpretation

We began this report by referring to the relatively advanced age at which children appear to approach adult-like interpretations of collective and distributive interpretations. Here in Figure 13 (from Grinstead et al. 2021), we see that 10-year-old Puerto Rican Spanish-speaking children have adult-like interpretations of *cada* and that before that age, they do not, consistent with English findings in, for example, Brooks & Braine (1996).

Similar results have been reported across a typologically diverse array of languages. The most obvious puzzle to be solved regarding this phenomenon is why it takes so long to acquire. Comparison with the development of *algunos* interpretations can serve as a reference point, in this regard. Some 4- and 5-year-olds in Grinstead et al. (2022) are adult-like in their *algunos* interpretations and virtually all 6-, 7-, and 8-year-olds are, as well. The roles of inhibition and ANS in feeding the lexicon appears the

⁹See Larson et al. (2019) for a discussion of Scope Freezing.

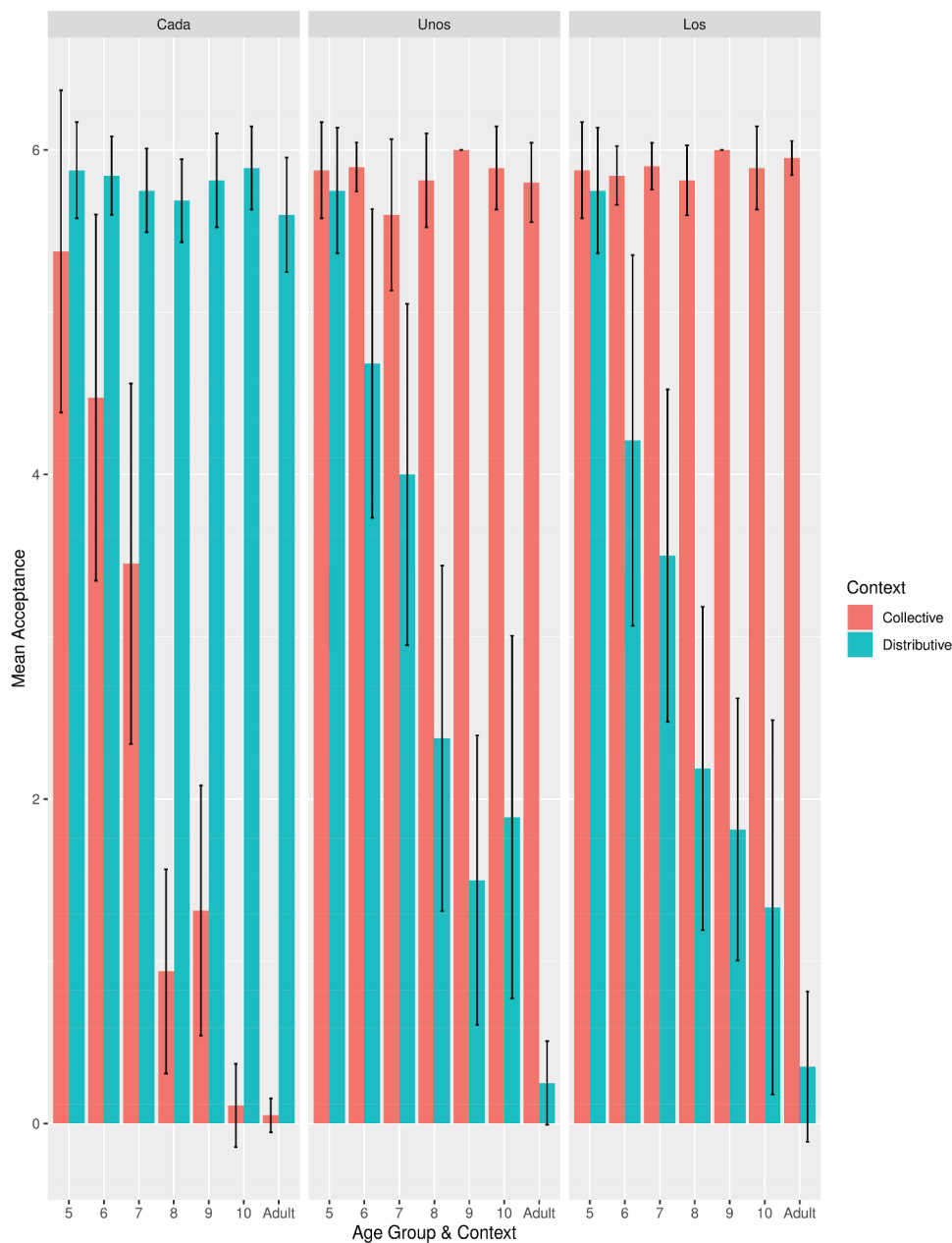


Figure 13. Child Spanish-speakers' acceptance of collective (*unos, los*) and distributive (*cada*) quantifiers in distributive and collective pragmatic contexts (Source: Grinstead et al. 2021, p. 60, Figure 3, reprinted by permission of Informa UK Limited, trading as Taylor & Francis Group, www.tandfonline.com).

same in both models and the roles of syntax and lexicon in feeding at least one of the relevant sentence types (*algunos* sentences in Figure 10 and *cada* sentences here) also appears very similar. Of course, these parallel interactions do not mean that the same exact information is being used for the two interpretation types, but rather only that the same domains of cognition are playing some role in each interpretation type. Where, then, is the factor that takes so long to develop?

Of course, there could be something inherently difficult about acquiring constructional idioms, but even if that speculation is not on the right track, there are other potential sources of difficulty as well. In previous work, we have speculated that the combinatorial nature of distributive-collective sentences is an order of magnitude greater than that of sentences with simple Quantity Implicatures in them. That is, the two critical quantities that must be grasped in order for a “some, but not all” interpretation to be drawn are *all*, which is a quantifier that ranges over all elements of a set, and *some*, which holds of “more than exactly 1” element of a set. To have an adult interpretation of the sentence in 21, speakers must decide how to restrict the plural indefinite *some children*. Child speakers seem able to overcome this obstacle by the end of preschool.

21. Some children will go to bed hungry tonight.

That is, the set of children under consideration in a sentence such as 21 can either be the total of all children, if the speaker is making a statement about world hunger, or it can be a statement that applies to a group of children that are prominent in some discourse, if the speaker is discussing whether the children under consideration will eat their dinner. In contrast to this level of complexity, consider our original stimulus sentence in 1, repeated here.

22. Cada minion movió una piedra.
Each minion moved a rock.

Notice that this sentence can be true in a range of pragmatic contexts other than the canonical distributive context we provide in our experiment, in which there is one rock per minion. Here is where the complexity enters the linguistic-pragmatic computation. For each rock that is added to the set of rocks to be pushed, there is a corresponding increase in the possible allowable combinations of individual minions with rocks. Each of these pragmatic context options is a potentially felicitous sentence-context pairing. A glimpse of this complexity is presented in Musolino (2009) where the author considers sentences such as 23 (p. 27, sentence 5).

23. Three men own four houses.

As it happens, there is a mathematical formula from the field of combinatorics that computes this complexity: a Stirling Number of the Second Kind: $S(n,k)$. What this number-type means for us is that as the number of agents in the subject referent in context grows and the number of different object referents with which they can be paired in context grows (or “... the number of partitions of n things into r non-empty sets” [Rennie & Dobson 1969, p. 116] grows), so does the number of possible combinations of them. Using cardinal numbers, as Musolino does, the combinations are calculable from the sentence.

24. Five minions move three rocks.

For sentence 24, $n = 5$, $k = 3$ and the Stirling number = 25 possible combinations. With 6 minions and 3 rocks, the number jumps to 90 possible combinations, with 7 minions, 301 combinations. The 7 objects paired into sets of 3 does not seem like a pragmatic situation that falls outside of a child’s experience, which leads to the question of how much of a learning obstacle this complexity represents.

Note, in particular, that with a universal quantifier such as *cada/each* in subject position and an indefinite object modified by *una/a*, all subjects and all objects in a pragmatic context must be calculated in all of the possible pairings to exhaustively evaluate the linguistic expression as true or false. While it seems unlikely that actual language use involves such exhaustive computations being used in evaluating the felicity of every distributive sentence, the point is that distributive sentences carry an inherently great potential complexity, which children’s higher-order reasoning may simply not be ready to grapple with, in

service of matching up the linguistic expression to the numbers of objects that present themselves in a given pragmatic context. One might imagine that this is at least an important aspect of the learning problem children face, if not the core.

While the combinatorics speculation is one possible source of the slow development of the construction, it is also possible that the independent developmental trajectories of the other cognitive capacities modeled here could also conspire to slow things down. The fact that an integrated whole, from subject to predicate, must be grasped in order for adult-like interpretations to be drawn, is different from *algunos*-type implicatures, which may occur with a relatively unrestricted set of sentence-types, and depend much more the acquisition of the NP quantifiers themselves. Distributive-collective interpretations may simply require larger, more integrated cognitive representations. This is the core of what is referred to as pluractionality.

5.8.4. *Pluractionality and development*

The difficulty for pluractionality is likely to be more about the fact that the subject is a distributive universal that pairs with individuals denoted by an ambiguous indefinite/numeral. The distributive-collective scale matters for this, but is a sub-part of the bigger learning challenge of the pluractional construction. This is fundamentally different from learning that an existential in subject position is part of the Quantity Scale and gets a partitive interpretation, as in 26.

25. Cada minion movió una piedra.
Each minion moved a rock.
26. Algunos niños bajaron la resbaladilla.
Some children went down the slide.
27. La maestra les dio un regalo a cada estudiante.
The teacher gave a present to each student.
28. El regalo era para cada estudiante.
The present was for each student.
29. La maestra abrazó a cada estudiante.
The teacher hugged each student.

For *cada* sentences, the evaluation always has to be about the distributive and any indefinite with which it can establish a pluractional relationship, as in the subject-object NPs in 25 and the direct object-indirect object NPs in 27. Otherwise, *cada* simply acts as a non-pluractional universal quantifier, as in 28 and 29. The pragmatics of matching pluractional sentences like these to the contexts in which they occur and the act of evaluating their felicity may simply require greater linguistic and cognitive non-linguistic resources than children are equipped with until 10 or 11 years of age. The cross-linguistic nature of this lateness would seem to support such a speculation. That is, there is no evidence, as-yet, of cross-linguistic means of expressing this type of pluractionality that is markedly earlier acquired in one language versus another.

In sum, we have attempted to use a statistical technique from the Life Sciences, PSEM, which calculates a Fisher's C-statistic, with a chi-squared-type distribution, to evaluate the goodness-of-fit of a conceptually plausible system of linguistic and non-linguistic cognitive variables and the links among them. The causal model we constructed was different from the model constructed in previous work in the literature, using the same technique, to model *algunos* implicature interpretations, in that collective implicatures appeared to be drawn at the utterance level, while *algunos* implicature interpretations are drawn at the NP quantifier level for the Quantity Implicature. Common to both models, inhibition and the ANS predict lexical development, and sentences (at least the *algunos* and the *cada* sentences) are predicted by a measure of phrasal syntax (the Argument Index) and a measure of lexicon (the standardized *Adivinanzas* measure

for *algunos*, and the Number of Different Words spontaneous measure for *cada*). Future work should consider similar systemic modeling for other types of comprehension outcome variables, perhaps including measures of literacy and math. Although our PSEM was an appropriate statistical technique for our data, with high statistical power for our sample size, future work could make the same comparison with conventional SEMs, calculated on a larger 200-300 child sample size to further explore the utility and validity of this technique.

Acknowledgments

The authors are grateful to David Melamed, Jonathan Lefcheck, Laura Wagner, Peter Culicover, Micha Elsner, John Opfer, Pedro Ortiz-Ramírez, Ana Arrieta-Zamudio, Ximena Carreto-Guadarrama, Jhovana Sandoval-Estrada, Michell Zúñiga-Espinosa, Alejandro López-Ramos, Debi Gómez, Olivia Mancera-Unikel, Eva Sifuentes-Monroy, Pedro Ruiz-Curcó, Mary Carmen Gasca, Zhiguo Xie, Melvin González-Rivera, Dorian González-Bonilla, Ana Teresa Pérez-Leroux, Cristina Schmitt, Rosa Guzzardo, Armando Betancourt, Melissa Mercado, and Leonor Vega. A special thank-you to Charles Grinstead for help with the combinatorics.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This project was funded by a grant from the National Science Foundation (NSF-1551903).

References

- Agresti, Alan. 2013. *Categorical data analysis* (3rd ed.). Hoboken, NJ: Wiley.
- Agrillo, Christian, Laura Piffer, Angelo Bisazza & Brian Butterworth. 2012. Evidence for two numerical systems that are similar in humans and guppies. *PLoS One* 7(2). <http://dx.doi.org.proxy.lib.ohio-state.edu/10.1371/journal.pone.0031923>
- Baayen, Harald, Robert Schreuder, Nivya De Jong & Andrea Krott. 2001. Dutch inflection: The rules that prove the exception. In S. Nootboom, F. Weerman & F. Wijnen (eds.), *Storage and computation in the language faculty*, 41–73. Dordrecht: Kluwer.
- Brooks, Patricia J. & Martin D.S. Braine. 1996. What do children know about the universal quantifiers all and each? *Cognition* 60(3). 235–268.
- Brooks, Patricia J., Xiangdong Jia, Martin D.S. Braine & Maria Da Graca Dias. 1998. A cross-linguistic study of children's comprehension of universal quantifiers: A Comparison of Mandarin Chinese, Portuguese and English. *First Language* 18(1(52)). 33–79.
- Caramazza, Alfonso. 1997. How many levels of processing are there in lexical access? *Cognitive Neuropsychology* 14(1). 177–208.
- Castilla-Earls, Anny, Douglas Petersen, Trina Spencer & Krista Hammer. 2015. Narrative development in monolingual Spanish-speaking preschool children. *Early Education and Development* 5(8). 1166–1186.
- Chomsky, Noam. 1975. Questions of form and interpretation. *Linguistic Analysis* 1(1). 75–109.
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: Massachusetts Institute of Technology.
- Costa, Albert, Michele Miozzo & Alfonso Caramazza. 1999. Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language* 41(3). 365–397.
- Crain, Stephen & Cecile McKee. 1985. The acquisition of structural restrictions on anaphora. In S. Berman, J. Choe & J. McDonough (eds.), *Proceedings of NELS 15*, 94–110. GLSA, Amherst, University of Massachusetts.
- Culicover, Peter & Ray Jackendoff. 2005. *Simpler syntax*. New York: Oxford University Press.
- de Koster, Anna, Jakub Dotlačil & Jennifer Spenader. 2017. Children's understanding of distributivity and adjectives of comparison. 41st Annual Boston University Conference on Language Development, Boston University.
- Dehaene, Stanislaus. 1997. *The number sense*. New York: Oxford University Press.
- Dotlačil, Jakub. 2010. *Anaphora and distributivity: A study of same, different, reciprocals and others*. Utrecht Institute for Linguistics, OTS, LOT Series, Utrecht doctoral dissertation.
- Dunn, Lloyd & Douglas Dunn. 2007. *Peabody Picture Vocabulary Test - 4th Edition*. San Antonio, TX: Pearson - PsychCorp.

- Dunn, Lloyd M., Delia E. Lugo, Eligio R. Padilla & Leota M. Dunn. 1986. *Test de Vocabulario en Imágenes Peabody* [Peabody Picture Vocabulary Test]. Circle Pines, MN: American Guidance Services.
- Eriksen, Barbara A., & Charles W. Eriksen. 1974. Effects of noise letters upon identification of a target letter in a non-search task. *Perception and Psychophysics* 16. 43–149.
- Feigenson, Lisa, Stanislaus Dehaene & Elizabeth S. Spelke. 2004. Core systems of number. *Trends in Cognitive Science* 8. 307–314.
- Gallistel, C.R. 2011. Mental magnitudes. In S. Dehaene & L. Brannon (eds.), *Space, time and number in the brain: Searching for the foundations of mathematical thought*, 3–12. New York: Elsevier.
- Gangopadhyay, Ishanti, Susan Ellis Weismer & Margarita Kaushanskaya. 2019. Domain-general inhibition and lexical processing in monolingual T and bilingual children: A longitudinal approach. *Cognitive Development* 49. 68–80.
- Gelman, Rochel & C.R. Gallistel. 1978. *The child's understanding of number*. Cambridge, MA: Harvard University Press.
- Grace, James B., Samuel M. Scheiner & Donald R. Schoolmaster. 2015. Structural equation modeling: Building and evaluating causal models. In Gordon A. Fox, Simoneta Negrete-Yankelevich & Vinicio J. Sosa (eds.), *Ecological statistics: Contemporary theory and application*, 168–199. Oxford: Oxford University Press.
- Grinstead, John, Jeff MacSwan, Susan Curtiss & Rochel Gelman. 1998. The independence of language and number. *Proceedings of the Annual Boston University Conference on Language Development* 22(1). 303–313.
- Grinstead, John, Melissa Nieves Rivera & John Opfer. 2020. Lexicon, ANS and symbolic estimation. Boston University Conference on Language Development, Boston University.
- Grinstead, John, Pedro Ortiz-Ramírez, Ximena Carreto-Guadarrama, Ana Arrieta-Zamudio, Amy Pratt, Myriam Cantú-Sánchez, Jonathan S. Lefcheck & David Melamed. 2022. Piecewise structural equation modeling of the quantity implicature in child language. *Language and Speech* 66(1). 35–67. <https://doi.org/10.1177/00238309211066086>
- Grinstead, John, Ramón Padilla-Reyes & Melissa Nieves-Rivera. 2021. A collective-distributive pragmatic scale and the developing lexicon. *Language Learning and Development* 17(1). 48–66. <https://doi.org/10.1080/15475441.2020.1863808>
- Grinstead, John, Jhovana Sandoval-Estrada & Michell Zúñiga-Espinosa. 2020. Working Memory and Subordination Indices. Paper presented at the Speer Lab Colloquium.
- Gualmini, Andrea, Sarah Hulsey, Valentine Hacquard & Danny Fox. 2008. The question-answer requirement for scope assignment. *Natural Language Semantics* 16. 205–237.
- Hanlon, Camille. 1986. Acquisition of set-relational quantifiers in early childhood. *Genetic, Social and General Psychology Monographs* 113(2). 213–264.
- Hayes, Andrew. 2018. *Introduction to mediation, moderation and conditional process analysis. A regression-based approach*. New York: Guilford Press.
- Houston, Derek M., Chi-hsin Chen, Claire Monroy & Irina Castellanos. 2020. How early auditory experience affects children's ability to learn spoken words. In M. Marschark & H. Knoors (eds.), *The Oxford handbook of deaf studies in learning and cognition*, 123–137. New York: Oxford University Press.
- Kennedy, Christopher & Kristen Syrett. 2022. Numerals denote degree quantifiers: Evidence from child language. In N. Gotzner & U. Sauerland (eds.), *Measurements, numerals and scales: Essays in honour of Stephanie Solt*, 135–62. Cham: Springer International Publishing.
- Kline, Rex. 1998. *Principles and practice of structural equation modeling*. New York: The Guilford Press.
- Kramer, Joel H., Dan Mungas, Katherine L. Possin, Katherine P. Rankin, Adam L. Boxer, Howard J. Rosen, Alan Bostrom, Lena Sinha, Ashley Berhel & Mary Widmeyer. 2014. NIH EXAMINER: Conceptualization and development of an executive function battery. *Journal of the International Neuropsychological Society* 20(1). 11–19.
- Labov, William. 1984. Field methods of the project of linguistic change and variation. In J. Baugh & J. Scherzer (eds.), *Language in use: Readings in sociolinguistics*, 28–53. Englewood Cliffs, NJ: Prentice Hall.
- LaBrant, Lou. 1933. A study of certain language developments of children in grades 4-12 inclusive. *Genetic Psychology Monographs* 14(5). 387–491.
- Landman, Fred. 1989. Groups, I. *Linguistics and Philosophy* 12. 559–605.
- Larson, Caroline, David Kaplan, Margarita Kaushanskaya & Susan Ellis Weismer. 2020. Language and inhibition: Predictive relationships in children with language impairment relative to typically developing peers. *Journal of Speech, Language, and Hearing Research* 63. 1115–1127.
- Larson, Richard K., Svitlana Antonyuk & Lei Liu. 2019. Superiority and scope freezing. *Linguistic Inquiry* 50(1). 233–252.
- Lefcheck, Jonathan S. 2016. piecewiseSEM: Piecewise structural equation modeling in R for ecology, evolution, and systematics. *Methods in Ecology and Evolution* 7(5). 573–579.
- Lefcheck, Jonathan S. 2019. piecewiseSEM: Piecewise Structural Equation Modeling in R. <https://cran.r-project.org/web/packages/piecewiseSEM/vignettes/piecewiseSEM.html> (10 Jul 2020).
- Link, Godehard. 1983. The logical analysis of plurals and mass terms: A lattice-theoretic approach. In Paul Portner & Barbara H. Partee (eds.), *Formal semantics - The essential readings*, 127–147. Cambridge, MA: Blackwell.
- Loban, Walter. 1963. *The language of elementary school children, Research Report No. 1*. Champaign, IL: National Council of Teachers of English.
- Luce, Paul A. & David B. Pisoni. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19(1). 1.

- May, Robert. 1985. *Logical form: Its structure and derivation*. Cambridge, MA: MIT Press.
- Mayer, Mercer. 1969. *Frog, where are you?* New York: Dial Press.
- Miller, Jon F. 1991. Quantifying productive language disorders. In Jon F. Miller (ed.), *Research on child language disorders: A decade of progress*, 211–220. Austin, TX: Pro-Ed.
- Miyake, Akira, Naomi Friedman, Michael Emerson, Alexander Witzki & Amy Howerter. 2000. The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology* 41. 49–100.
- Musolino, Julien. 2009. The logical syntax of number words: Theory, acquisition and processing. *Cognition* 111. 24–25.
- Negen, James & Barbara Sarnecka. 2012. Number-concept acquisition and general vocabulary development. *Child Development* 83(6). 2019–2027.
- Novick, Jared M., John C. Trueswell & Sharon L. Thompson-Schill. 2005. Cognitive control and parsing: Reexamining the role of Broca’s area in sentence comprehension. *Cognitive, Affective and Behavioral Neuroscience* 5(3). 263–281.
- Opfer, John, Dan Kim, Christopher J. Young & Francesca Marciani. 2019. Linear spatial–numeric associations aid memory for single numbers. *Frontiers in Psychology* 10. 1–12.
- Padilla-Reyes, Ramón. 2018. *Connections among scales, plurality and intensionality in Spanish*. The Ohio State University doctoral dissertation.
- Pagliarini, Elena, Gaetano Fiorin & Jakub Dotlačil. 2012. The acquisition of distributivity in pluralities. In A. K. Biller, E. Y. Chung & A. E. Kimball (eds.), *Proceedings of the Annual Boston University Conference on Language Development, Boston University*, vol. 3, 387–399. Boston: Cascadia Press.
- Pollard, Carl & Ivan Sag. 1994. *Head-driven phrase structure grammar*. Chicago: Chicago University Press.
- Preacher, Kristopher J., & Hayes, Andrew. 2008. Contemporary approaches to assessing mediation in communication research. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The Sage sourcebook of advanced data analysis methods for communication research*, 13–54. Thousand Oaks, CA: Sage.
- Rennie, B.C., & Dobson, J. 1969. On Stirling numbers of the second kind. *Journal of Combinatorial Theory* 7. 116–121.
- Rhemtulla, Mijke, Patricia É. Brosseau-Liard & Victoria Savalei. 2012. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods* 17(September). 354–373.
- Roberts, Craige. 1987. *Modal subordination, anaphora, and distributivity*. University of Massachusetts, Amherst dissertation.
- Russell, Bertrand. 1905. On denoting. *Mind* 66. 479–493.
- Schoolmaster, Donald R., Chad R. Zirbel & James Patrick Cronin. 2020. A graphical causal model for resolving species identity effects and biodiversity–ecosystem function correlations. *Ecology* 101(8). e03070.
- Schwarzschild, Roger. 1996. *Pluralities*. Dordrecht: Kluwer.
- Shah, Priti & Akira Miyake. 1996. The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General* 125(1). 4–27. <https://doi.org/10.1037/0096-3445.125.1.4>
- Shipley, Bill. 2000. A new inferential test for path models based on directed acyclic graphs. *Structural Equation Modeling* 7(2). 206–218.
- Shipley, Bill. 2002. *Cause and correlation in biology: A user’s guide to path analysis, structural equations and causal inference*. Cambridge: Cambridge University Press.
- Shipley, Bill. 2009. Confirmatory path analysis in a generalized multilevel context. *Ecology* 90(2). 363–368.
- Siegler, R.S. & John Opfer. 2003. The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science* 14. 237–243.
- Syrett, Kristen & Julien Musolino. 2013. Collectivity, distributivity and the interpretation of numerical expressions in child and adult language. *Language Acquisition: A Journal of Developmental Linguistics* 20(4), 259–291.
- Wolf, Erika J., Kelly M. Harrington, Shauna L. Clark & Mark W. Miller. 2013. Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement* 73(6). 913–934.
- Wynn, Karen. 1990. Children’s understanding of counting. *Cognition* 36(2). 155–193.
- Wynn, Karen. 1992. Children’s acquisition of the number words and the counting system. *Cognitive Psychology* 24(2). 220–251.
- Xu, Fei & Elizabeth S. Spelke. 2000. Large number discrimination in 6-month-old infants. *Cognition* 74 (1). B1–B11.

Appendix. PSEM with Poisson regressions, instead of OLS regressions, of the *cada* and *unos* outcome variables

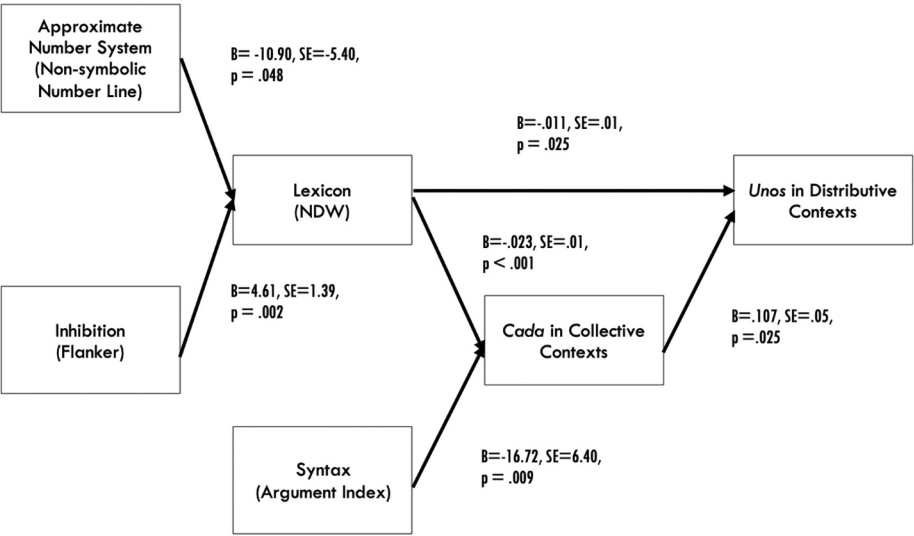


Figure A1. Piecewise Structural Equation Model – Fisher’s C Statistic = 12.123, $p = .436$, 12 degrees of freedom, No significant alternative endogenous variable–predicted paths, AIC = 33.123.

Table B1. Tests of Directed Separation — Tests Whether 3 Endogenous Variables (Lexicon, CadaC and UnosD) Can Be Predicted by Alternative Paths. All Are Non-significant ($>.05$).

Independence Claim	DF	Coefficient	p-value
Lexicon ~ Syntax + Inhibition + nonsymboliclambda	52	0.2897	0.7732
unosD ~ Syntax + Lexicon + cadaC	53	1.2171	0.2236
cadaC ~ Inhibition + Syntax + Lexicon	52	-0.8624	0.3885
unosD ~ Inhibition + Lexicon + cadaC	53	1.3918	0.1640
cadaC ~ nonsymboliclambda + Syntax + Lexicon	52	1.0550	0.2914
unosD ~ nonsymboliclambda + Lexicon + cadaC	52	0.3498	0.7265