

Distributive and collective interpretations: Data from children with developmental language disorder

Marta Ponciano, Department of Linguistics and Basque Studies, University of the Basque Country (UPV/EHU),
marta.ponciano@ehu.eus

Elena Castroviejo, Catalan Philology and General Linguistics, Universitat de Barcelona, elena.castroviejo@ub.edu

Urtzi Etxeberria, IKER (Center for Research on Basque Language and Texts): joint research unit 5478, French National Center for Scientific Research, urtzi.etxeberrria@iker.cnrs.fr

Agustín Vicente, Department of Linguistics and Basque Studies, University of the Basque Country (UPV/EHU);
IKERBASQUE: Basque Foundation for Science, agustin.vicente@ehu.eus

Sentences with plural expressions can be interpreted either *collectively* (where a group acts as a whole) or *distributively* (where each individual member acts independently). Distributive interpretations are often considered more complex (or harder), requiring additional syntactic and/or semantic computation. According to the *complexity hypothesis*, children with linguistic and cognitive processing limitations, such as those with Developmental Language Disorder (DLD), may find these interpretations particularly challenging. This study examines how Spanish-speaking children with DLD interpret ambiguous doubly quantified sentences that allow both collective and distributive readings, comparing their performance to typically developing (TD) peers. We analyzed access to distributive readings and interpretive preferences across development in children aged 5–12 years. Children with DLD were able to access both readings, and both groups showed similar overall preferences. The results do not reveal reliable differences between groups and, therefore, do not provide support for the complexity account.



1. Introduction

Sentences with plural subjects, such as (1), can give rise to two distinct interpretations: a *distributive* reading (2a), in which the predicate applies to each individual member of the group or plurality, and a *collective* reading (2b), in which the predicate applies to the group or plurality as a whole.¹

- (1) Tres niños sujetan dos globos.
 three boy.PL hold.3PL.PRS two balloon.PL
 ‘Three boys are holding two balloons.’

(Musolino, 2009)

- (2) a. Tres niños sujetan dos globos cada uno.² *distributive*
 three boy.PL hold.3PL.PRS two balloon.PL each one
 ‘Three boys are each holding two balloons.’
 b. Tres niños sujetan dos globos entre los tres. *collective*
 three boy.PL hold.3PL.PRS two balloon.PL among the three
 ‘Three boys are holding two balloons together.’

This type of ambiguity arises only with *mixed predicates* (e.g., *carry*, *paint*, *hold*), which are compatible with both readings (Link, 1983, 1987). In contrast, inherently collective (e.g., *gather*, *meet*) or inherently distributive predicates (e.g., *smile*, *have brown eyes*) allow only one interpretation (Champollion, 2016, 2019; Lasersohn, 1995).

It has been suggested that distributive readings are more complex than collective ones, partly because they are modeled as requiring a covert Distributive operator D in the semantic representation (e.g., Brisson, 1998; Champollion, 2019, 2020; Heim et al., 1991; Link, 1983, 1987), which applies the predicate to each individual member of the group (though see Landman, 1989, 2000, and Lasersohn, 1995, for the view that collective readings also require a dedicated operator). Some accounts assume that this operator is purely semantic (e.g., Brisson, 1998, 2003; Heim et al., 1991; Link, 1983, 1987), while others maintain that it is syntactically represented (e.g., Champollion 2019; Dotlačil, 2010, 2015; Dotlačil & Brasoveanu, 2021). Processing evidence seems to support this *complexity hypothesis*, although only one study (Dotlačil & Brasoveanu, 2021)

¹ See Landman (1995) and Lønning (1997) for overviews of the various possible interpretations of plural expressions. While our study focuses exclusively on collective and distributive readings, other readings, such as cumulative interpretations, have been observed (Champollion, 2020; Musolino, 2009; Winter, 2000). For example, the sentence in (1) may also receive a cumulative interpretation, where three boys jointly hold two balloons, but without specifying how the balloons are distributed. Some accounts treat collective readings as a subtype of cumulative ones (Ussery, 2007). Since these distinctions are not central to our research question, we group such readings under the label “collective” for simplicity.

² Note that in Spanish, the sentence is disambiguated by means of the complex quantifier *cada uno* (‘each one’).

offers clear support for it. A frequently cited study by Frazier et al. (1999), however, suggests that what is costly is not the distributive reading per se, but the process of disambiguating between readings (distributive and collective).

Another piece of evidence, though more indirect, in favor of the relative complexity of the distributive reading is the strong preference for collective readings observed in adult speakers who are faced with ambiguous stimuli (Brooks & Braine 1996; Gil, 1982; Kaup et al., 2002; Pagliarini et al., 2012; Syrett & Musolino 2013). However, this interpretation is challenged by developmental data: young typically developing (TD) children do not consistently favor collective readings (Brooks & Braine, 1996; Pagliarini et al., 2012; Syrett & Musolino, 2013).³ These data call into question the complexity view, suggesting that the observed adult bias may not stem from intrinsic complexity differences, but rather from distributional patterns in the input, in particular, from the frequent co-occurrence of distributive readings with overt markers like *each*. That is, adults would interpret an ambiguous sentence such as (1) collectively, because, typically, distributive readings are overtly marked by *each* (see Saponaro et al., 2025, for converging evidence from the domain of production).

Thus, the adult preference for collective readings in ambiguous contexts might not reflect the relative simplicity of collective interpretations, but rather a default strategy adopted in the absence of explicit cues. The role of *each*, on the other hand, remains unclear. According to the pragmatic view, listeners default to collective readings after pragmatically inferring that the speaker would have used *each* if intending to communicate distributivity (Dotlačil, 2010). According to the frequentist view, the default arises from the association of distributive readings with *each* and collective readings with the unmarked form (Gennari & MacDonald, 2005/2006). In both accounts, the presence (or absence) of the overt operator *each* allegedly influences interpretation preferences, independently of semantic or cognitive complexity.

In this study, we aim to explore the complexity hypothesis by focusing on children with Developmental Language Disorder (DLD), a neurodevelopmental condition affecting approximately 7–10% of the child population (Norbury et al., 2016; Tomblin et al., 1997). Children with DLD exhibit persistent difficulties with structural aspects of language, especially syntax and morphology, despite having typical sensory, intellectual, and neurological profiles (Bishop et al., 2016, 2017; Leonard, 2014). They often struggle with complex syntactic structures, such as passives, relative clauses, and embedded sentences (Balthazar & Scott, 2023; Georgiou & Theodorou, 2023; Leonard et al., 2013; Montgomery et al., 2018; Montgomery & Evans, 2009; Van Der Lely, 1998), and show cognitive limitations, including reduced working memory, slower

³ Multiple studies have shown that young typically developing children tend to either accept both collective and distributive interpretations of ambiguous sentences to the same extent or prefer the distributive interpretation (e.g., Brooks & Braine, 1996; Pagliarini, et al., 2012; Syrett & Musolino, 2013; among others).

processing speed, and weaker inhibitory control (Delage & Frauenfelder, 2020; Henry et al., 2012; Leonard et al., 2007; Montgomery, 2000, 2003).⁴ Given these linguistic and cognitive challenges, distributive interpretations, which, on the complexity account, require more syntactic or semantic computation and, thus, greater processing effort, may be especially difficult for this group.⁵

We propose that if children with DLD consistently opt for the collective reading in ambiguous contexts, such a preference would suggest that the distributive interpretation is comparatively more complex. In principle, it can be ruled out that the preference for collective readings is related to any role of *each*. TD children do not appear to fully grasp the distributive function of *each* until around age 7–8 (e.g., Brooks & Braine, 1996; Pagliarini et al., 2012), and it would be surprising if DLD children developed this understanding earlier than their TD peers. Recall that the role of *each* in making collective readings the default interpretation of ambiguous sentences is attributed either to frequency effects or to pragmatic reasoning. Now, children with DLD often show difficulties in statistical learning (e.g., Evans et al., 2009; Haebig et al., 2017; Lammertink et al., 2017), which suggests that they will not outperform their TD peers in this respect. On the other hand, the preference for collective readings can hardly be attributed to pragmatic reasoning, as children with DLD also tend to lag behind in pragmatic development (Arosio et al., 2017; Botting & Adams, 2005; Katsos et al., 2011). In sum, if DLD children show a preference for collective readings of ambiguous sentences, such a preference would suggest that they find distributive readings more difficult.

In short, studying children with DLD, by comparing them with typically developing peers, provides a valuable opportunity to test whether distributive readings are inherently more complex and demanding.

2. Aims of the present study

This study investigates differences between Spanish-speaking children with DLD and Spanish-speaking TD children in the interpretation of ambiguous sentences that allow both distributive and collective readings, such as those involving numerical quantifiers (e.g., *Three boys hold two balloons*). Using a sentence–picture matching task, we address three questions. First, we examine whether children with DLD differ from their TD peers in their ability to access distributive interpretations of ambiguous sentences. Second, we investigate

⁴ Whether their comprehension difficulties stem from these cognitive limitations remains unclear (Montgomery et al., 2016, 2021; among others) and is the subject of much debate.

⁵ While the present study does not aim to disentangle the possible underlying sources of difficulty (syntactic, semantic or cognitive), it provides new empirical insights into how this population processes semantic/syntactic ambiguity.

whether interpretive preferences for collective versus distributive readings differ between groups, and how these preferences change across development. Third, we explore whether deriving a distributive interpretation from an ambiguous sentence is more effortful than deriving a collective interpretation, and whether any such processing cost differs between children with DLD and TD children, using response times as an index of processing effort. Participants included children aged 5 to 12 years and a comparison group of adults aged 21 to 61 years.

2.1 Research questions

This study aims to address the following questions:

RQ1. Access

- a. Can children with DLD access distributive readings of ambiguous, doubly quantified sentences?
- b. How does their performance compare to that of their TD peers, and how does it change with age?

RQ2. Interpretive preference

- a. Do children with DLD show a stronger preference for collective readings than TD children?
- b. Do TD children show developmental changes toward adult-like collective preferences?

RQ3. Processing effort

- a. Are response times longer for distributive readings than for collective readings of ambiguous sentences (reflecting greater processing effort in deriving a distributive interpretation)?
- b. Is the eventual distributive–collective RT difference larger in children with DLD than in TD children?

2.2 Predictions

- Typically developing children

We expected TD children to behave in line with previous findings, showing a developmental trajectory from no clear preference in younger children to a growing preference for collective interpretations in ambiguous contexts as they get older (Brooks & Braine, 1996; Pagliarini et al., 2012).

Regarding response times, we expected that distributive readings would take longer to be selected than collective readings, reflecting their greater interpretive complexity. This prediction follows from previous work on adults, which suggests that processing distributive meanings requires additional cognitive effort (e.g., Dotlačil & Brasoveanu, 2021; Frazier et al., 1999).

- Children with DLD

If distributive readings are, indeed, more complex than collective ones, we expect children with DLD to show a stronger preference for collective interpretations, especially at younger ages. However, if both readings are similarly complex, we expect children with DLD to behave similarly to their TD peers (e.g., showing no clear preference).

Additionally, if distributive readings take longer to be selected than collective ones because of their complexity, we expect children with DLD to show even longer response times for these interpretations, compared to collective readings, on the one hand, and to TD children on the other hand.

3. Method

3.1 Participants

A total of 96 children (48 children with developmental language disorder and 48 children with typical language development) participated in the study, ranging in age from 5 to 12 years old. Initially, 71 children with language difficulties and 80 children with typical development were recruited. Children in the DLD group were screened using a battery of standardized tests administered by trained researchers. For the TD group, further selection was carried out, to ensure that participants met the criteria for typical development, specifically, children with no history of language delays or disorders, no reported concerns by teachers or parents, and age-appropriate academic performance were included. After applying our inclusion/exclusion criteria (see more information below) and accounting for experimenter error, data for several participants were lost. The final sample consisted of 48 children with DLD (14 girls, 34 boys, $M_{\text{age}} = 100.81$, $SD = 24.43$, $\text{Range} = 60\text{--}151$)⁶ and 48 age-matched children with TD (17 girls, 31 boys, $M_{\text{age}} = 100.62$, $SD = 24.05$, $\text{Range} = 60\text{--}147$, $t(94) = 0.037$, $p = .96$).

We did not conduct a power analysis to determine the minimum number of participants needed to detect true effects. Instead, we recruited as many children with DLD as possible, given the time and resources available.

⁶ Although more boys than girls were included in our DLD sample, the actual prevalence of DLD is similar across genders (Calder et al., 2022; Norbury et al., 2016; Tomblin et al., 1997). Disparities in diagnosis and referral may stem from behavioral differences, lower awareness, and gender-based diagnostic biases (McGregor, 2020; Whitehouse, 2010; Wittke & Spaulding, 2018; Zhang & Tomblin, 2000).

Children belonging to the experimental group were contacted and recruited through several local institutions, including early intervention services and an association supporting children with language difficulties (TEL EUSKADI). Children from the control group were recruited from two public schools (Antonio Trueba IPI in Portugalete, Basque Country, Spain, and Ibaibe Herri Ikastetxea in Barakaldo, Basque Country, Spain). All participants were native speakers of Spanish.⁷ All families completed a written informed consent form and a background (information) questionnaire.

All children in the DLD group met inclusion and exclusion criteria defined in accordance with diagnostic guidelines established by a Spanish expert committee (Aguado et al., 2015) and the Delphi Consensus Study (CATALISE; Bishop et al., 2017). Identification was based on a combination of standardized language assessment, clinical/formal diagnosis when available, and converging parental and educational reports, in line with current consensus recommendations. Specifically, inclusion criteria for the DLD group were: (a) an average non-verbal intellectual quotient (NVIQ) ≥ 75 , as measured by the Leiter International Performance Scale, Third Edition (LEITER-3 non-verbal intelligence scale; Roid et al., 2013); (b) performance ≥ 1.25 SD below age-level expectations (standard score < 81) on at least one of the three core composite scores of the Clinical Evaluation of Language Fundamentals – Fifth Edition, Spanish version (CELF-5; Wiig et al., 2018): Core Language, Receptive Language, or Expressive Language (**Table 2**); (c) normal hearing at 500, 1000, 2000 and 4000 Hz at 20 dB; (d) normal or corrected-to-normal vision; and (e) absence of other medical or neurological conditions (e.g., Autism Spectrum Disorder). Three children did not meet criterion (b) at the time of testing, but did have a formal external diagnosis from qualified clinical professionals.

Information regarding hearing, vision, and medical history was obtained via parental questionnaires and, in most cases, particularly for children without a prior diagnosis, confirmed through independent examinations. Children in the DLD group either had a formal diagnosis of developmental language disorder (58% of participants)⁸ or were in the process of obtaining one, but were, nevertheless, receiving speech-language services or clinical intervention at the time of the study. Parents of children in the TD group also completed the background questionnaire, reporting no history or diagnosis of language impairment, no medical or neurological conditions, no hearing loss, and normal or corrected-to-normal vision. These reports were supported by the children's school teachers. Descriptive data for all groups are presented in **Table 1**. Additional information on the standardized assessments used to characterize the DLD group (CELF-5

⁷ Some participants were simultaneous bilinguals who acquired Basque from birth or early infancy, alongside Spanish.

⁸ Some were additionally diagnosed with ADHD (but were medicated) or showed signs of dyslexia, which is quite common. DLD occurs quite frequently alongside ADHD or dyslexia (Young et al., 2002). In fact, comorbidity between DLD and other neurodevelopmental disorders is highly frequent, as they share difficulties across different cognitive areas (Hansen et al., 2018; Pennington, 2006).

Spanish and LEITER-3), including their administration, subtests, and psychometric properties, is provided in Appendix B.

Table 1: Descriptive characteristics of participants' chronological age.

DLD (n = 48)			TD (n = 48)		Comparison	
Age Group	Mean (SD)	Range	Mean (SD)	Range	t(df)	p-value
5-y.o. (n = 12)	64.83(3.92)	60–70	65.33(4.32)	60–71	–0.20(10)	.83
6-y.o. (n = 16)	79.37(3.88)	73–83	78.37(2.77)	74–83	0.59(14)	.56
7-y.o. (n = 12)	89.16(3.92)	84–94	90.33(3.01)	86–94	–0.57(10)	.57
8-y.o. (n = 24)	102.16(3.56)	96–107	102.33(2.42)	98–107	–0.13(22)	.89
9-y.o. (n = 12)	112(3.03)	109–116	111.5(2.81)	108–116	0.29(10)	.77
10–12-y.o. (n = 20)	138.2(8.29)	123–151	137.2(8.80)	121–147	0.26(18)	.79

Note. Mean, SD, and Range of chronological ages are reported in months. The numbers in parentheses indicate the total number of child participants in each age group. Age groups were defined for descriptive and illustrative purposes, to ensure a relatively balanced number of participants across groups, particularly at the older end of the age range. At age 10, there were two participants per group, and at age 12, there were three participants per group. All inferential statistical analyses were performed with age as a continuous variable.

Each child with DLD was individually matched to a TD child of the same age (+/– 4 months) at the time of the experimental task. We were unable to closely match one participant. We also attempted to match participants by gender; however, this was not possible for all cases. Children in the control group were not administered the full battery of assessments, primarily due to time constraints.

In addition to the child participants, we included a group of 35 adult native speakers of Spanish, to provide a baseline for interpretation preferences and also to (exploratorily at least) assess whether the oldest child groups approach adult-like performance.⁹ Adults were aged 21 to 61 years ($M_{\text{age}} = 32.2$, $SD = 8.78$), and reported no history of language, hearing, or

⁹ A reviewer rightly pointed out that, according to Pagliarini et al. (2012), even TD children aged 10–12 years may not display fully adult-like collective preferences. For that reason, we decided to include the adult group.

neurological disorders, and had normal or corrected-to-normal vision. All adults completed the same sentence–picture matching task as the children.

Table 2: Standard scores for language and cognitive assessment in children with DLD.

	DLD (n = 48)		
Variable	Mean	SD	Range
CELF-CLS ^a	72.2	9.06	53–91
CELF-RLS ^b	79	9.69	61–104
CELF-ELS ^c	70	9.06	45–83
LEITER-3 ^d	101	7.41	88–113

Notes:

^aCELF-5 CLS = Clinical Evaluation of Language Fundamentals Spanish version, 5th Edition: Core Language score. Scaled scores (M = 100, SD = 15).

^bCELF-5 RLS = Clinical Evaluation of Language Fundamentals Spanish version, 5th Edition: Receptive Language score. Scaled scores (M = 100, SD = 15).

^cCELF-5 ELS = Clinical Evaluation of Language Fundamentals Spanish version, 5th Edition: Expressive Language score. Scaled scores (M = 100, SD = 15).

^d LEITER-3 NON-VERBAL IQ (Leiter International Performance Scale – 3th Edition; Age-scaled scores (M = 100, SD = 15).

Neurotypical children were not tested on these standardized measures. Their scores were believed to be in the typical range.

Children in the DLD group showed impairments across expressive and/or receptive language domains of varying severity, as reflected in CELF-5 composite scores. Inclusion in the DLD group required scoring < 1.25 SD below the mean on at least one core composite score. While most children met this criterion, three participants did not fall below this cutoff on any single composite score, but had a formal external diagnosis of DLD and converging evidence from parental and educational reports. This approach seems to be consistent with current consensus recommendations for identifying DLD (e.g., Bishop et al., 2017).

3.2 Materials

We carried out the experiment in Spanish. Each trial consisted of a target sentence and four images appearing together on a computer screen. Every sentence was presented twice. Participants saw visual arrays of four pictures paired with test sentences (played auditorily twice), as shown in **Figure 1**. Auditory stimuli were recorded in a soundproof recording booth by a female native speaker of Spanish, using neutral (non-emphatic, non-stressed) speech prosody to avoid biases towards one interpretation (see Lohiniva & Panizza, 2016). They were later edited in PRAAT (Boersma & Weenik, 2024) to ensure approximately equal length. Sentences sounded quite natural, despite the unbiased prosody.

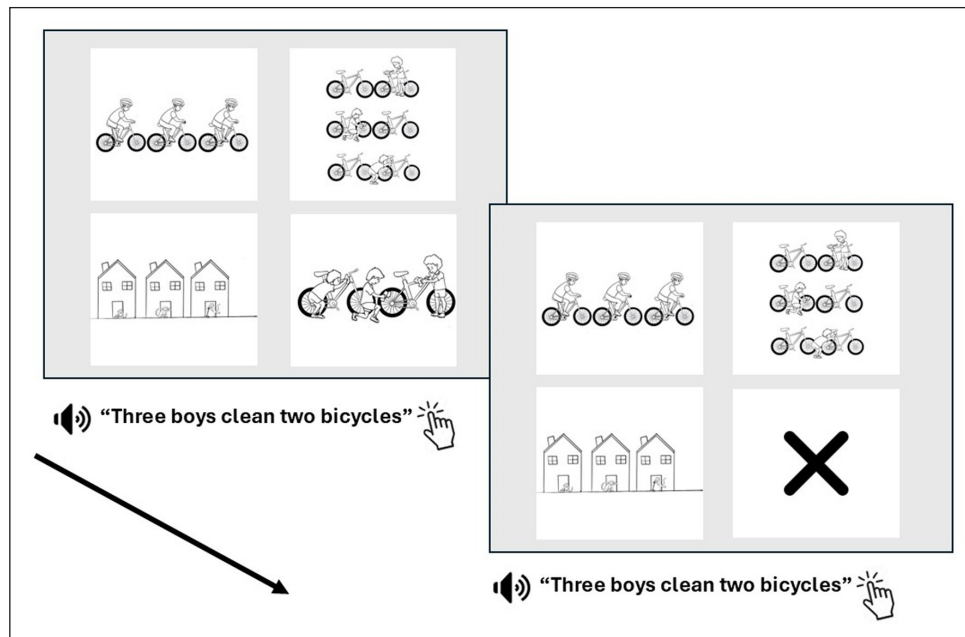


Figure 1: Schematic representation of a critical trial. Auditory stimuli adapted for illustration purposes.

We created pictures depicting either a collective or distributive context and some other inconsistent contexts.

Critical sentential stimuli were selected based on previous literature (Brooks & Braine, 1996; Drozd et al., 2017; Grinstead et al., 2021; Musolino, 2009; Syrett & Musolino, 2013). Our target sentences were doubly quantified sentences, mostly of the 3/2 or 2/3 type, like (3) and (4), which have multiple readings (as described above).

We included predicates that were ambiguous and equally easy and plausible to interpret either way (Champollion, 2020; Link, 1983, 1987). The predicate types varied and included activities (atelic), accomplishments (telic), achievements, and states (Vendler, 1967).

- (3) Dos niños rompen tres ventanas.
Two boy.PL break.3PL.PRS three window.PL
'Two boys break three windows.'

- (4) Tres niñas levantan dos cajas.
Three girl.PL lift.3PL.PRS two box.PL
'Three girls lift two boxes.'

We used 14 transitive verbs for the critical stimuli: *romper* ('break'), *limpiar* ('clean'), *comer* ('eat'), *sujetar* ('hold'), *mover* ('move/push'), *tener* ('have'), *cantar* ('sing'), *leer* ('read'), *pasear*

(‘walk’), *jugar* (‘play’), *ensuciar* (‘dirty’), *llevar* (‘carry’), *dibujar* (‘draw’), and *levantar* (‘lift’). The subjects had either human referents (girls, boys) or animal referents (elephants, monkeys, sharks, dogs, or pigs). No pretests were conducted to assess the availability of the target readings.

Numerals were restricted to small numbers, such as *two* and *three*, which are acquired quite early, around the age of 3, according to the developmental literature (Sarnecka & Gelman, 2004; Wynn, 1992). Also, TD children around that age (4- to 6-year-olds) use those numerals very often in the context of counting objects, as in *two dogs*, counting sequences, as in *One, two, three*, or to answer *how-many-type* questions (Gennari & MacDonald, 2005/2006). This was also done to avoid the possible arithmetical and mathematical difficulties that children with DLD might face (Cowan et al., 2005; Donlan et al., 2007; Katsos et al. 2011).

There were 16 critical trials containing ambiguous sentences (like (2) or (3)). In addition, there were 32 filler trials similar in structure, for a total of 48 test sentences (see **Table 3** for examples). The purpose of those was to keep participants attentive and to ensure that they were paying attention to the number of events and to the semantic content of the predicate. Filler trials were randomly interspersed between critical trials.

Table 3: Examples of sentence-picture sets in experimental and control conditions.

Condition	Sentences	Content of the images
Ambiguous	Dos niños sujetan tres globos. ‘Two boys hold three balloons.’	2 boys are holding 3 balloons.
		2 boys are each holding 3 balloons.
		2 boys are each holding 3 balls.
		2 girls are holding 2 bicycles.
Collective	Tres elefantes mueven dos coches. ‘Three elephants push two cars.’	3 elephants are pushing 2 cars.
		3 elephants are pushing 3 airplanes.
		3 boys are pushing 3 cars.
		3 elephants are standing together.
Distributive	Tres niñas levantan dos cajas. ‘Three girls lift two boxes.’	3 girls are each lifting 2 boxes.
		3 girls are each lifting 2 pianos.
		3 gorillas are each throwing 2 boxes.
		2 girls are walking.
Filler	Un niño come tres galletas. ‘A/One boy eats three cookies.’	One boy is eating 3 cookies.
		One boy is eating a whole chicken.
		An elephant is jumping rope.
		One girl is eating a cookie.

Of the 16 critical trials, four appeared in the collective condition, four in the distributive condition, and eight in the ambiguous condition. This design served two main purposes: (i) to compare participants' picture selections between collective versus distributive conditions in terms of speed and cognitive effort; and (ii) to examine participants' ability to identify both target responses (collective and distributive interpretations) compared to other, incorrect options, and to observe interpretive preferences.

Collective and distributive condition trials each presented an ambiguous sentence alongside three distractor pictures incompatible with either reading, and one picture consistent with the sentence under the respective reading (i.e., the "correct" picture). In collective conditions, the correct picture supported only the collective reading; in distributive conditions, it supported only the distributive reading.

Ambiguous condition trials showed an ambiguous sentence with two consistent images, one matching the collective reading and the other, the distributive reading, plus two distractors depicting different activities or participants. Distractors depicted other activities by the same number of participants or the same participants, or the same kind of event, but not the same participants. Picture positions were counterbalanced. Only in ambiguous trials could children (and adults) experience a conflict between collective and distributive interpretations.

Lastly, filler trials included non-ambiguous sentences (e.g., *One/A boy eats three cookies*) with three foil pictures and one consistent picture matching the sentence's unique meaning.

We also included 2 demonstration trials and 3 practice trials, designed to familiarize participants with the task.¹⁰ These trials were similar in format to the filler and critical trials but, in some cases, included nonliteral light verb constructions (e.g., *hacer la cama* 'to make the bed' or *poner la mesa* 'to set the table'). For these, two pictures were presented, each corresponding to a different interpretation of the sentence. Demonstration trials included one ambiguous sentence with a light verb construction (*hacer la cama*) and one non-ambiguous sentence. Practice trials consisted of one ambiguous sentence with a light verb construction (*poner la mesa*) and two non-ambiguous sentences. Practice trials were not analyzed.

Target trials, therefore, appeared in three conditions (collective, distributive and ambiguous). We generated four lists. The same items appeared in all four lists under different conditions. To make sure that participants did not see the same visual array multiple times and heard the same sentence three times, participants were randomly assigned to one of the four lists. Each list contained, as already mentioned, 48 audio-visual stimuli (4 collective, 4 distributive, 8 ambiguous, 32 fillers) presented in random order. Left-right and top-bottom position of the

¹⁰ In a way, following the pattern from CELF-5 (Wiig et al., 2018).

images was counterbalanced. The stimulus list order did not appear to have any effect on the results and was not considered for further analysis.

3.3 Procedure

Participants completed a forced-choice sentence-picture matching task. Children sat in front of a laptop and were instructed to help a female puppet character learn Spanish. The task was run using E-Prime 3.0 stimulus presentation software (Psychology Software Tools, 2016).

For each trial, participants first heard the prompt: “¿Cuál representa mejor la frase [test sentence]?” (‘Which one best captures the meaning of [test sentence]?’ referred to as Q1). After choosing one of the four pictures, they then heard: “¿Hay otro dibujo que también pueda representar la frase [test sentence]?” (‘Is there another picture that could also match [test sentence]?’ Q2).

Each sentence was, thus, heard twice: once in the first phase (Q1), followed by four pictures; and once again in the second phase (Q2), with the previously chosen picture replaced by an “X” sign and the remaining three pictures displayed. In the first phase, children selected the image that best matched the sentence. In the second, they were asked whether any other picture could also match. If only one match was possible, they were instructed to select the “X” sign; otherwise, they indicated a second possible match, based on their spontaneous preference. Children responded by tapping the screen with their fingers as quickly and accurately as possible.

After the demonstration and practice phase, all children completed the test trials successfully; no participants were excluded for failing to understand the task. Stimuli remained onscreen until a response was given. Picture selections and response times were recorded. Response times were measured in milliseconds from the onset of the question and picture presentation to the participant’s tap on the touchscreen. Auditory stimuli consisted of three components: the question prompt, a brief pause, and the target sentence. Because target sentence durations varied naturally, we individually adjusted the pause preceding each sentence, so that the total duration was approximately constant across trials. These adjustments were almost imperceptible and ensured that differences in RT primarily reflected the time participants took to select a picture in response to the sentence.

No feedback was provided during test trials. However, limited feedback was given during practice trials, to ensure participants understood that they could select two pictures when appropriate. At the end of the experiment, a short animated clip of the puppet thanked participants for their effort and participation. The task lasted approximately 12 to 15 minutes.

Each child was tested individually in a quiet room. Children with DLD were tested at a research or intervention center, and TD children were tested at their schools. Most participants

completed the procedure in a single session, although some children with DLD required a second session.

TD children completed only the experimental task. For children with DLD, the experimental task was followed by the CELF standardized language assessment, and then the LEITER-3 nonverbal cognitive abilities test, administered in fixed order. Breaks were provided as needed. All testing sessions were video-recorded.

All parents/caregivers were informed of the study's goals and provided written informed consent; the consent procedure was approved by the Ethics Committee for research with human beings (CEISH) of the University of the Basque Country (EHU), code M10_2019_205. Families of children with DLD received a report with the results of the standardized assessments. A summary of the study's findings was shared with participating schools as a newsletter-style poster for distribution to the families of TD children.

Adults completed the same sentence–picture matching task individually in a quiet location of their choosing. Instructions were identical to those given to the children, and all responses were recorded using the same equipment and procedure as for child participants.

3.4 Data analysis

Participants were required to achieve at least 80% accuracy on filler trials to be included in the analyses. All participants met this criterion; therefore, no exclusions were made on this basis.

All analyses were conducted in R (R Core Team, 2021) using RStudio (RStudio Team, 2020), with the packages *ggplot2* (Wickham, 2016), *lme4* (Bates et al., 2007), *dplyr* (Wickham et al., 2015), and *lmerTest* (Kuznetsova et al., 2015). We performed three analyses, each targeting a different aspect of participants' performance: (i) access to distributive readings, (ii) interpretive preferences for collective versus distributive readings, and (iii) processing effort, explored through response times. For all analyses, we fitted (generalized) mixed-effects regression models including random intercepts for participants and items. In line with recommendations by Barr et al. (2013), we initially evaluated more complex random-effects structures, including random slopes. In cases of convergence problems or singular fits, or when random effects contributed only minimally to model variance, we simplified the random structure by removing the effects with the lowest variance, step by step. The final models retained only random intercepts. The *bobyqa* optimizer was used for model fitting.

3.4.1 Access to distributive readings (RQ1)

To find out whether children with DLD can access distributive readings and how this ability compares to that of their TD peers (RQ1a–b), we fitted a mixed-effects logistic regression model

to the picture-selection data from the ambiguous condition trials. The dependent variable was whether a distributive reading was selected at any point during the trial (in either Q1 or Q2), coded binarily (1 = distributive picture selected, 0 = not selected). We chose this operationalization to capture any evidence of access to distributive interpretations, regardless of when this interpretation was selected within the trial. The optimal model included AGE as a continuous predictor (centered and scaled, measured in months) and GROUP (DLD vs. TD; sum-contrast coded), as well as their interaction, as fixed effects. AGE was treated as a continuous predictor here, to model developmental change as a gradual process, rather than as discrete age-group differences. Random intercepts for participants and items were included.

3.4.2 Interpretive preference (RQ2)

To assess children's interpretive preferences (RQ2a–b), we analyzed picture-selection responses from Q1 in ambiguous trials only. Trials in which participants selected an incorrect (non-target) picture were excluded. For each remaining trial, the selected picture was coded as 1 if it represented the collective reading, and 0 if it represented the distributive reading. The dependent variable was, therefore, the type of reading chosen (1 = collective, 0 = distributive). To maximize statistical power and retain the full variability in age, age was modeled as a centered continuous predictor. We fitted a mixed-effects logistic regression model with GROUP (DLD vs. TD), AGE (centered and scaled; measured in months), and their interaction as fixed effects, and random intercepts for participants and items. GROUP was treatment-contrast coded, with TD as the reference level. This model allows us to test whether preferences change with age in the TD group. The AGE x GROUP interaction tests whether the effect of age differs between TD and DLD children, that is, whether the developmental trajectory in DLD differs from that observed in TD. Age-group information is presented descriptively for visualization purposes. To explore adults' interpretive preferences, we fitted a separate intercept-only mixed-effects logistic regression model, with random intercepts for participants and items.

3.4.3 Processing effort (RQ3)

To examine potential differences in response times (in ms) between collective and distributive readings (RQ3a–b), we analyzed response times from Q1, comparing the two conditions (collective vs. distributive). RTs were measured from the onset of the question (including the pause and the target sentence) and picture presentation until the participant's tapping of the picture on the laptop touch screen. RTs were log-transformed to correct for non-normality. No

outlier removal was performed, as there was no objective criterion for defining extreme RTs.¹¹ RTs were analyzed for all correct responses; incorrect (non-target) responses were excluded, accounting for 14% of critical trials.

Based on model comparisons, using the Akaike Information Criterion (AIC), the final model (for the children's data) included the fixed factors GROUP (DLD vs TD) and CONDITION (distributive vs. collective), and their interaction (sum-contrast coded). We also included random intercepts for participants and items. We tried including CONDITION as a random slope by participant, but it led to a singular fit and was removed. The dependent variable was response time. The model included the $\text{CONDITION} \times \text{GROUP}$ interaction, to explore whether the effect of condition varied across groups. For adults, we fitted a separate model with CONDITION as the only fixed factor (sum-contrast coded), and random intercepts for participants and items.

Full model outputs and statistical results for each analysis are presented in the corresponding results sections below. All study materials, data, and analysis scripts have been made publicly available on the Open Science Framework (OSF) at:

<https://osf.io/knp7m/overview>.

4. Results

4.1 Fillers and picture-selection in collective versus distributive conditions

Participants performed accurately on filler trials (Q1), with mean accuracy of 98.6% in the DLD group, 99.9% in the TD group, and 99.4% in adults. Children's individual scores ranged from 87.5% to 99%. Filler trials were excluded from subsequent analyses. As expected, participants were also highly accurate in selecting collective and distributive interpretations in the collective and distributive conditions (DLD: 95.6%, TD: 95.8%), because the visual displays depicted only one interpretation (collective or distributive), and the alternatives were not genuine competitors. These condition trials were included to examine response-time differences between collective and distributive readings. We now turn to the main results.

4.2 Picture selection: Access to the distributive reading in ambiguous conditions

First, we inspected the overall proportion of distributive picture selections across groups and age groups. **Table 4** reports the mean proportion of distributive responses selected by each diagnostic

¹¹ It seemed difficult to detect true outliers, and the risk of eliminating extreme, but valid, RTs was great, so we decided not to remove extreme RTs, because it could introduce considerable bias.

group (DLD vs. TD) across age groups, as well as the overall means. Adults' distributive picture selections are also reported for comparison.

Table 4: Proportion of participants choosing the distributive picture in either Q1 or Q2.

	Proportion of distributive responses	
	DLD	TD
5-year-olds	0.50 (0.26)	0.79 (0.23)
6-year-olds	0.59 (0.28)	0.64 (0.3)
7-year-olds	0.56 (0.21)	0.65 (0.41)
8-year-olds	0.77 (0.24)	0.64 (0.34)
9-year-olds	0.67 (0.38)	0.77 (0.28)
10–12-year-olds	0.82 (0.21)	0.71 (0.44)
Overall	0.68 (0.27)	0.69 (0.34)
Adults	0.88 (0.23)	

The model revealed no significant group differences in distributive picture selection (see **Table 5**). Children with DLD and TD children showed comparable rates of distributive responses, indicating no evidence of reduced access or difficulties in accessing distributive readings in the DLD group. Inspection of individual data points in **Figure 2** indicates that several children in the DLD group and the TD group performed below chance level across the full age range (5 to 12 years). This suggests that access to distributive readings is not uniform across all individual children.¹²

Table 5: Model results for group comparisons in access to distributive readings.

Term	β	95% CI	z-value	p-value
Intercept	1.24	[0.75, 1.73]	4.95	<.001
AGE (centered)	1.71	[-0.02, 3.45]	1.93	.053
GROUP	-0.13	[-0.95, 0.69]	-0.30	.763
AGE $\times$ GROUP	2.64	[-0.80, 6.08]	1.50	.133

¹² In an additional model including only TD participants (children and adults), with 10–12-year-olds as the reference level, we found no statistically reliable differences between 10–12-year-olds and adults in access to distributive readings (see Table A1 in Appendix A).

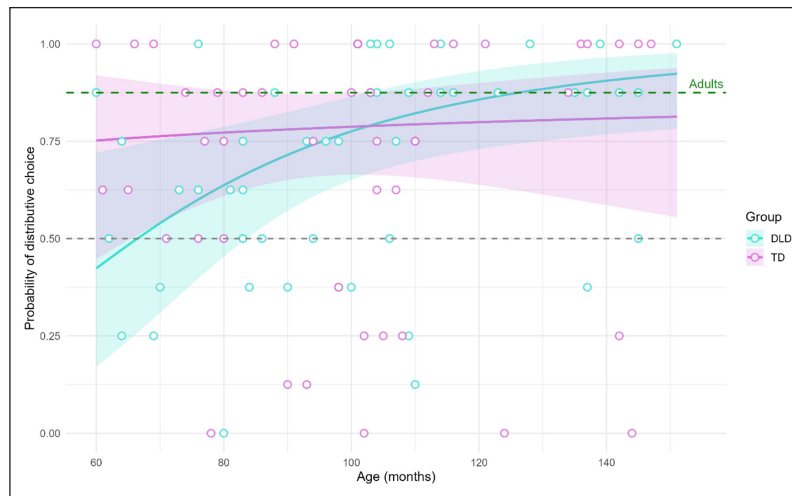


Figure 2: Developmental trajectory of distributive picture selections across age per diagnostic group.

Note: Lines and shaded bands represent model-predicted probabilities and 95% confidence intervals; points show participant-level mean proportions. The dashed green line indicates the adult mean.

4.3 Picture selection: Preferences

Table 6 shows the mean proportions (and standard deviations) of collective and distributive picture selections by age group and diagnostic group, with adults included for comparison.

Figure 3A visualizes these data as percentages of collective and distributive picture selections within each age group and diagnostic group, providing a descriptive overview of interpretive preferences. **Figure 3B** shows the model-predicted probabilities of choosing the collective picture.

Table 6: Descriptive statistics for collective and distributive picture selections (Q1, ambiguous trials) by age group and diagnostic group. Adults are included for comparison.

	Proportions of collective and distributive picture selections			
	DLD		TD	
Age group	Collective	Distributive	Collective	Distributive
5-year-olds	0.69 (0.13)	0.31 (0.13)	0.52 (0.09)	0.48 (0.09)
6-year-olds	0.61 (0.20)	0.39 (0.20)	0.70 (0.20)	0.30 (0.20)
7-year-olds	0.62 (0.13)	0.38 (0.13)	0.57 (0.33)	0.43 (0.33)
8-year-olds	0.66 (0.15)	0.34(0.15)	0.69 (0.24)	0.31 (0.24)
9-year-olds	0.79 (0.23)	0.21 (0.23)	0.75 (0.14)	0.25 (0.14)
10–12-year-olds	0.80 (0.15)	0.20 (0.15)	0.79 (0.26)	0.21 (0.26)
	Collective		Distributive	
Adults	0.83 (0.21)		0.17 (0.21)	

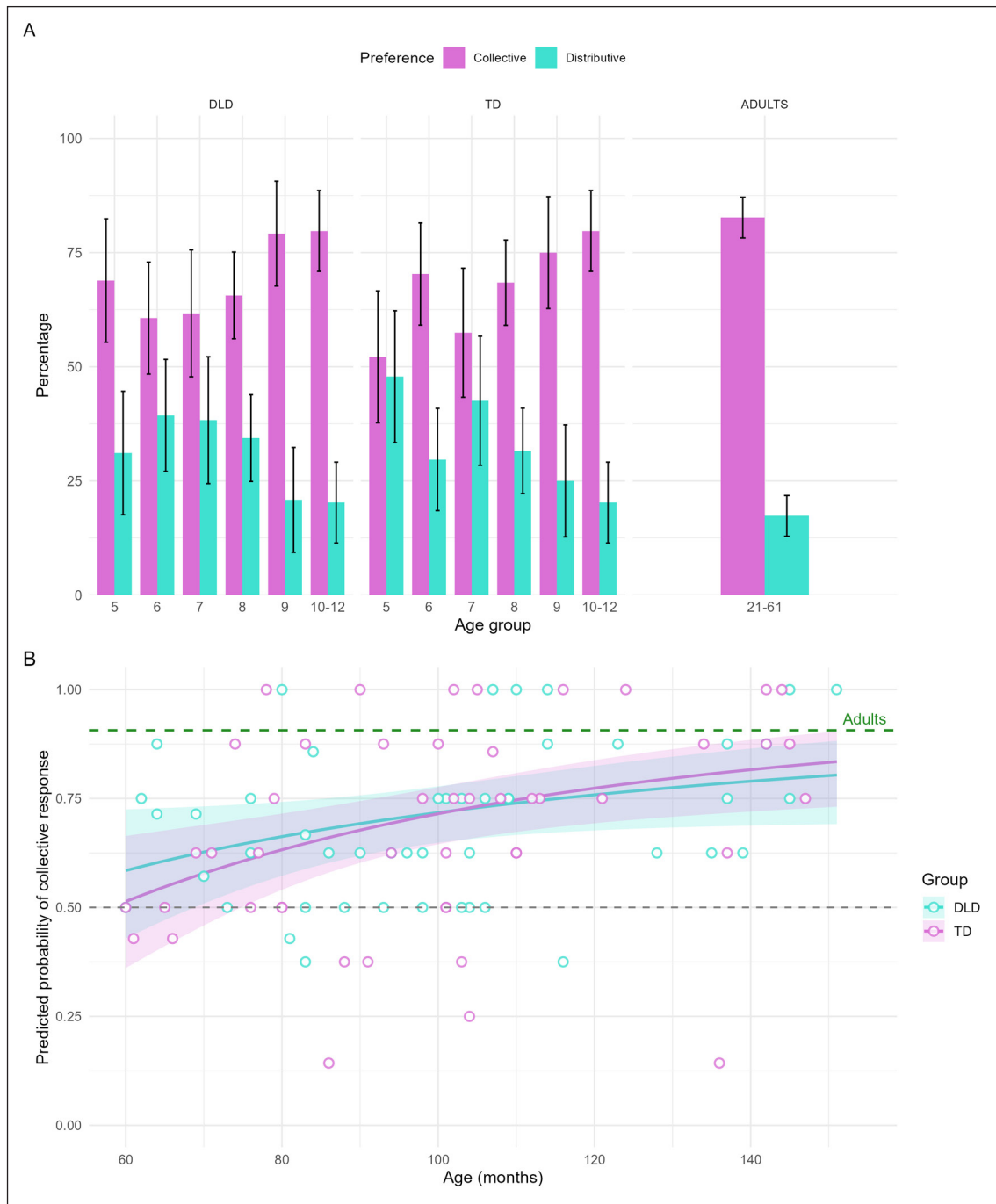


Figure 3: (A) Proportion of collective and distributive picture selections by age group and group (TD vs DLD). Error bars represent 95% confidence intervals. (B) Model-predicted probabilities of collective picture selections across the age range for DLD and TD children. Shaded bands represent 95% confidence intervals. Points represent participant-level mean responses. The dashed green line indicates the model-predicted probability for adults as a reference.

The model showed that the intercept (representing TD children, mean age) was significantly different from zero ($\beta = 1.50$, $z = 5.23$, $p = < .001$), indicating that at the reference point (TD, mean age), TD children are more likely than chance to choose collective responses. The main effect of GROUP was not significant ($\beta = -0.17$, $z = -0.45$, $p = .652$), suggesting no difference between TD and DLD children. The main effect of AGE was significant ($\beta = 1.69$, $z = 2.82$, $p = .005$), suggesting that preference for collective responses increases with age in TD children. Crucially, the GROUP and AGE interaction was not significant, indicating that the developmental trajectory of collective preferences did not differ between DLD and TD children. The summary of the output of this model can be seen in **Table 7**.

Table 7: Model results for group comparisons in interpretive preferences (reference level: TD children).

Term	β	95% CI	z-value	p-value
Intercept	1.50	[0.94, 2.06]	5.23	<.001
GROUP (DLD)	-0.17	[-0.91, 0.57]	-0.45	.652
AGE (centered)	1.69	[0.52, 2.87]	2.82	.005
GROUP (DLD) x AGE	-0.54	[-2.18, 1.11]	-0.64	.522

The intercept from the adult's model was positive and significant ($\beta = 2.27$, $z = 5.24$, $p < .001$), showing an overall significant preference for collective interpretations.¹³

These results suggest that both groups' preferences change with age in a similar way.

4.4 Response times (picture-selection times)

The model for RTs revealed no significant differences between groups ($\beta = 0.02$, $t = 0.51$, $p = .610$). However, there was a robust main effect of CONDITION ($\beta = -0.090$, $t = -5.04$, $p < .001$), reflecting longer response times for distributive interpretations than for collective interpretations. There was no interaction between GROUP and CONDITION. The full model is reported in **Table 8**, and the overall response-time pattern is illustrated in **Figure 4**.

¹³ Although not a primary focus of the study, we conducted an exploratory analysis of preferences in the TD group, to assess whether the oldest age group (10–12-year-olds) differed from adults; no significant differences were found (see Table A2 in Appendix A).

Regarding adults, CONDITION was also a significant factor in their model ($\beta = -0.09$, $t = -4.03$, $p < .001$), with longer response times for distributive interpretations than for collective interpretations. Taken together, these results indicate that all groups (TD children, children with DLD, and adults) showed the same direction of response-time differences across conditions.

Table 8: Model results for group comparisons in response times.

Term	β	95% CI	t	df	p -value
Intercept	8.75	[8.71, 8.79]	420.78	38.39	<.001
GROUP	0.02	[-0.05, 0.08]	0.51	93.74	.610
CONDITION	-0.09	[-0.13, -0.06]	-5.04	625.44	<.001
GROUP $\times$ CONDITION	0.06	[-0.02, 0.13]	1.53	624.44	.127

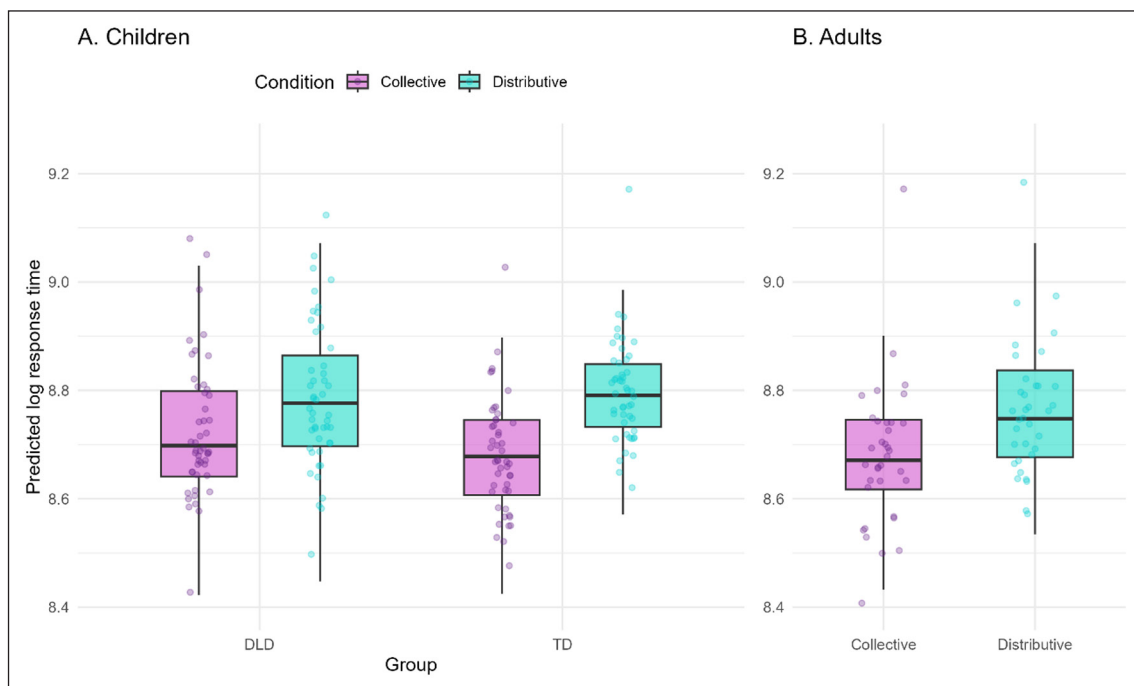


Figure 4: Predicted response times for collective and distributive readings. (A) Children's predicted log response times per group and condition. (B) Adults' predicted log response times per condition. Boxplots represent model-predicted distributions; points show participant-level means.

We discuss these results in the context of our predictions and RQs.

5. Discussion

This study investigated whether distributive interpretations of ambiguous doubly quantified sentences are more complex than collective ones, using children with DLD as a test case. Distributive readings have been said to be syntactically, semantically, and cognitively more demanding than collective readings. Since children with DLD are known to have difficulties with complex syntactic structures and often show cognitive limitations (e.g., reduced working memory, weaker inhibitory control, slower processing speed), we hypothesized that they might struggle to access distributive readings if these readings are indeed more complex, and, therefore, show a stronger preference for collective interpretations. We used a sentence–picture matching task in which children, as well as adults, were presented with a sentence containing two numerical quantifiers (e.g., *Three girls lift two boxes*) and four pictures. There were three experimental conditions: (i) ambiguous, where both collective and distributive readings were visually represented; (ii) collective-only, where only the collective reading was depicted; and (iii) distributive-only, where only the distributive reading was represented. Children and adults were first asked to select the picture that best matched the sentence (Q1), and then to indicate whether another picture could also match the sentence (Q2). We now revisit our research questions in light of our results.

Regarding our first research question, we analyzed picture-selection data from ambiguous trials to determine whether children with DLD can access distributive readings and how their performance compares to that of their TD peers across development (5–12 years). We looked at whether participants selected a distributive interpretation in either Q1 or Q2. The results show that children with DLD, including the youngest participants (5-year-olds), were able, overall, to access distributive readings of ambiguous doubly quantified sentences. Their performance closely matched that of TD children across the full age range. These findings suggest that children with DLD show no difficulties in accessing multiple readings of ambiguous numerically quantified sentences, at least with respect to collective and distributive interpretations, and that distributive interpretations are within their grammatical reach and processing capacity. Performance in the TD group is consistent with previous research showing that typically developing children are able to access distributive readings from an early age (Musolino, 2009; Syrett & Musolino, 2013). In fact, we know from those studies that TD children as young as three can access multiple readings (collective, distributive, cumulative) of ambiguous sentences (a.o. Musolino, 2009; Syrett & Musolino, 2013).

Descriptively, however, performance was not uniform across participants. Inspection of individual data points revealed that some children in both groups performed below chance levels across the full age range (5 to 10–12 years). This descriptive pattern shows that access to

distributive interpretations was generally present, but not consistent across all children. Similar variability has been reported in previous studies of typically developing children (e.g., Pagliarini et al., 2012, for Italian; Syrett & Musolino, 2013, for English), and the factors contributing to these differences are unclear. It is important to note that task-related factors may have facilitated access to distributive readings. The inclusion of both ambiguous and unambiguous conditions could have increased participants' overall sensitivity to distributivity.¹⁴

Turning now to our second research question, we asked whether children with DLD show a stronger preference for collective readings than TD children, and whether TD children show developmental changes toward adult-like collective preferences. To address these questions, we analyzed picture-selection responses from Q1 on ambiguous trials only. Contrary to our predictions, we found no evidence that children with DLD differed from their TD peers in their interpretive preferences. Children with DLD did not show stronger collective preferences than their TD peers. Instead, we found a significant effect of age: collective preferences increased with age across the sample, and this developmental pattern appeared similar in both groups. Adults, as expected, showed a robust preference for collective readings, consistent with many previous findings (e.g., Brooks & Braine, 1996, Experiment 2; Syrett & Musolino, 2013, Experiment 2).

At the same time, these findings should be interpreted cautiously. The absence of significant group differences should not be taken as evidence of equivalence between TD and DLD children. Given the modest sample sizes and the limited number of ambiguous trials, the present study may have lacked sufficient power to detect subtle group differences or developmental divergences. Consequently, the present findings are best interpreted as showing no detectable evidence of differences between groups in the current dataset, rather than definitive evidence that such differences do not exist. Replication with larger samples and a greater number of ambiguous items would, therefore, be desirable.

It is also possible that potential difficulties with distributive interpretations emerge earlier in development than in the age range examined here. Thus, children with DLD may have experienced greater difficulty with distributive readings at younger ages, but largely overcome these difficulties by the ages tested in the present study (by age 5, the typical age of earliest diagnosis).

The developmental increase in collective preferences observed in both groups replicates a substantial body of cross-linguistic work showing that, in the absence of explicit distributive markers such as *each*, children gradually move toward adult-like collective interpretations of

¹⁴ Anecdotally, some adult participants reported that they would not have considered a distributive interpretation unless it had been explicitly presented as an option. While these observations were not systematically collected or analyzed, they are consistent with claims in the theoretical literature that distributive interpretations of sentences without explicit distributive markers are often judged (as) marginal or even unavailable (Champollion, 2020).

ambiguous sentences (e.g., Brooks & Braine, 1996; de Koster et al., 2017, 2018; de Koster et al., 2021; Padilla-Reyes, 2018; Pagliarini et al., 2012). This trend has often been attributed to pragmatic and/or frequency-based learning from the input, rather than to the intrinsic complexity of distributive interpretations. Given the similar trajectories observed in the TD and DLD groups, this suggests that children with DLD are able to engage in the kind of pragmatic inferencing required to prefer collective readings or, alternatively, that this is an area in which statistical learning is preserved (Kapa & Mettler, 2024).

Taken together, the present findings do not support the prediction that distributive readings are more difficult for children with DLD than for their TD peers. However, given the limitations discussed above, this conclusion should be regarded as tentative. Given that no differences between groups were observed, the present findings do not seem to support the prediction that distributive readings are inherently more complex than collective readings, on the assumption that DLD children exhibit difficulties with complex structures.

Finally, our third research question explored whether response times differed between collective and distributive condition trials and whether this effect varied across groups. We analyzed RTs from Q1 on non-ambiguous trials, including only correct responses. Both TD and DLD children, as well as adults, took longer to select distributive interpretations than collective interpretations. In principle, this pattern is consistent with the idea that distributive interpretations may require additional processing (Dotlačil & Brasoveanu, 2021; Frazier et al., 1999). Importantly, however, there were no differences in RTs between groups, indicating that the magnitude of this slowdown was comparable across TD children and children with DLD.

At the same time, we must acknowledge, as suggested by a reviewer, that these RT results must be interpreted with caution, because they are subject to a potential confound between interpretive complexity and visual complexity. Distributive pictures necessarily contained more objects and were, therefore, visually more complex than collective pictures (see **Figure 1** above). This design choice followed from our primary goal of examining how different interpretations are derived from the same ambiguous sentence, rather than simply verifying a single interpretation. As a result, longer RTs in the distributive condition trials may reflect either increased interpretive complexity or longer visual inspection times.

Notably, however, the primary contribution of the present study was to assess whether differences between children with DLD and their TD peers could provide evidence for a complexity-based account. Although the RT pattern is compatible with such an account, the absence of reliable group differences in RTs means that the study of children with DLD still does not provide evidence in favor of complexity.¹⁵

¹⁵ One of our reviewers suggested an alternative design that could more directly test the processing complexity of distributive readings. In this design, the pictures would be matched in the number of objects. For example, one could

5.1 Limitations

Before concluding, several limitations of the present study should be acknowledged. First, as we have already pointed out, due to recruitment constraints, the sample size was relatively small for the age-range covered. As a result, we cannot conclude that children with DLD do not experience greater difficulties with distributive readings than typically developing children. What we can say, however, is that we currently lack evidence that they do have such difficulties. Second, standardized language and cognitive assessments were not administered to the TD group. Instead, typical development was confirmed through parental and teacher reports. While this approach is common in developmental research, future studies would benefit from administering the same standardized assessment battery to both groups to strengthen group comparisons.

Summing up, comparable designs with larger samples will be necessary to confirm the robustness of the observed patterns.

6. Conclusion

For the first time, this study examined how children with DLD and TD children interpret ambiguous sentences with numerical quantifiers that can be interpreted collectively or distributively. By comparing these groups, we aimed to test the *complexity hypothesis*, which posits that distributive readings are inherently more complex than collective ones, and, therefore, potentially more challenging for children with DLD, who are known to struggle with language structure and processing capacity.

Our results show that children in both groups were able to access distributive interpretations (at least, in this task), and that they both seemed to have similar interpretive preferences across the age range tested. These findings do not reveal reliable group differences and, therefore, do not support the complexity account. However, they provide independent evidence that children with DLD can, in fact, access distributive readings at levels comparable to their TD peers. The current results suggest that part of becoming an adult language user involves gradually learning or realizing that distributive readings are typically overtly marked with *each*, and sentences without such markers are usually interpreted collectively.

present a sentence like “Two boys cleaned two bikes” with a collective picture, and separately “Two boys cleaned one bike” with a distributive picture. This approach would allow any differences in response times to be attributed specifically to interpretive complexity, rather than to visual differences. While we could have included such trials, using different sentences to equalize visual complexity would address a different question: namely, whether verifying a distributive event is slower than verifying a collective one, rather than whether deriving a distributive interpretation from an ambiguous sentence is more costly.

Appendix A

Table A.1: Model results for access to distributive readings in TD participants (reference level: 10–12-year-olds).

Term	β	95% CI	z-value	p-value
Intercept	2.28	[0.17, 4.40]	2.11	.035
AGE 5	0.36	[-2.85, 3.57]	0.22	.826
AGE 6	-1.57	[-4.44, 1.31]	-1.07	.285
AGE 7	-0.93	[-4.08, 2.21]	-0.58	.561
AGE 8	-1.20	[-3.85, 1.44]	-0.89	.373
AGE 9	-0.08	[-3.23, 3.07]	-0.05	.960
Adults	1.56	[-0.77, 3.88]	1.31	.190

Table A.2: Model results for interpretive preferences in TD participants (reference level: 10–12-year-olds).

Term	β	95% CI	z-value	p-value
Intercept	1.69	[0.74, 2.64]	3.48	<.001
AGE 5	-1.57	[-2.95, -0.20]	-2.24	.025
AGE 6	-0.62	[-1.92, 0.67]	-0.94	.347
AGE 7	-1.31	[-2.71, 0.09]	-1.84	.066
AGE 8	-0.66	[-1.84, 0.53]	-1.08	.279
AGE 9	-0.37	[-1.79, 1.05]	-0.51	.611
Adults	0.30	[-0.73, 1.33]	0.57	.568

Appendix B Background measures: Clinical assessment

Some children with DLD were diagnosed by independent and experienced speech and language therapists from clinical/public institutions, according to the standard procedures (ICD-11; WHO, 2022 or DSM-5; APA, 2013), but not all of them were. Therefore, we used a battery of assessments. The Clinical Evaluation of Language Fundamental CELF-5 (Spanish version) is a validated and reliable measure of language competence for 5-to-15-year-olds. Its composite scores have demonstrated high reliability ($\alpha = .91-.94$), and the instrument has shown strong sensitivity and specificity for identifying language difficulties. The CELF-5 consists of subtests that assess sentence comprehension, sentence repetition, sentence formulation, following directions, defining words, semantic relationships, and other aspects of language. Non-verbal cognitive ability was assessed using the LEITER-3. It is a non-verbal intelligence test appropriate for individuals from two and

a half to 75+ years. It is quick and easy to administer, and the intelligence index (IQ) is derived from four subtests: Sequential Order, Form Completion, Classification and Analogies, and Figure Ground.

As mentioned, children in the TD group were not assessed on those standardized measures, due to time restrictions. The session would have taken too long to be included during school hours. However, parents' and teachers' reports confirmed that these children had no known linguistic or cognitive impairments (or any diagnosis of such impairments). Future studies should make sure to test both control and experimental groups on the same battery of tests.

Abbreviations

PL = plural

PRS = present

3PL = third person plural

Data accessibility statement

All data, code, and materials are available at <https://osf.io/knp7m/overview>.

Ethics and consent

This study was conducted under the approval of the Ethics Committee for Research with Human Beings (CEISH) of the University of the Basque Country (EHU), approval code M10_2019_205.

Acknowledgements

We are very grateful to the Developmental Language Disorder Association (TEL Euskadi) in the Basque Country, Spain, specifically to Alicia Solavera, to the speech and language therapy center Hitz Zentroa in Llodio (Basque Country, Spain), to the Early Attention Service in Vitoria-Gasteiz, and to the public schools Antonio Trueba IPI in Portugalete (Basque Country, Spain) and Ibaibe Herri Ikastetxea in Barakaldo (Basque Country, Spain) for helping us with participant recruitment and logistics. We are enormously indebted to all child participants and their families, who generously dedicated their time and support to our research. We would also like to thank all adult participants who generously volunteered to take part in this study. We are also grateful to Olatz Etxeberria and Estíbaliz Vera Ortíz de Zarate for designing the pictures used in the study. Finally, we would like to thank the anonymous reviewers and editors of *Glossa Psycholinguistics* for their insightful comments and constructive suggestions, which greatly improved this article.

This research has been partially supported by project FUNLAT (PID2021-122233OBI00 and PID2025-169832NB-I00), funded by MICIU/AEI/10.13039/501100011033 and by “ERDF/EU,” and by the IT1811-26 Research Group (Basque Government). This work has also been supported

by a predoctoral grant from the Predoctoral Training Programme for Non-Doctoral Research Staff of the Department of Science, Universities and Innovation of the Basque Government (PRE_2019_1_0291) (M.P).

Competing interests

The authors have no competing interests to declare.

Authors' contributions

M. Ponciano: conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing – original draft, writing – review & editing, visualization, and project administration. E. Castroviejo: conceptualization, methodology, resources, writing – review & editing, funding acquisition, supervision, and project administration. U. Etxeberria: conceptualization, methodology, and resources. A. Vicente: conceptualization, methodology, resources, writing – review & editing, funding acquisition, supervision, and project administration.

ORCID IDs

Marta Ponciano: <https://orcid.org/0000-0002-7062-0518>

Elena Castroviejo: <https://orcid.org/0000-0003-0371-1214>

Urtzi Etxeberria: <https://orcid.org/0000-0001-6560-1614>

Agustín Vicente: <https://orcid.org/0000-0002-5331-5546>

References

- Aguado, G., Coloma, C. J., Martínez, A. B., Mendoza, E., Montes, A., Navarro, R., & Serra, M. (2015). Documento de consenso elaborado por el comité de expertos en TEL sobre el diagnóstico del trastorno. *Revista de Logopedia, Foniatría y Audiología*, 35(4), 147–149. <https://doi.org/10.1016/j.rlfa.2015.06.004>
- American Psychiatric Association (APA). (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing. <https://doi.org/10.1176/appi.books.9780890425596>
- Arosio, F., Foppolo, F., Pagliarini, E., Perugini, M., & Guasti, M. T. (2017). Semantic and pragmatic abilities can be spared in Italian children with SLI. *Language Acquisition*, 13(4), 418–429. <https://doi.org/10.1080/15475441.2017.1308254>
- Balthazar, C. H., & Scott, C. M. (2023). Sentences are key: Helping school-age children and adolescents build sentence skills needed for real language. *American Journal of Speech-Language Pathology*. Advance online publication. https://doi.org/10.1044/2023_AJSLP-23-00038

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Sarkar, D., Bates, M. D., & Matrix, L. (2007). The lme4 package. *R Package Version*, 2(1), 74.
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Adams, C., Archibald, L., Baird, G., Bauer, A., Bellair, J., Boyle, C., Brownlie, E., Carter, G., Clark, B., Clegg, J., Cohen, N., Conti-Ramsden, G., Dockrell, J., Dunn, J., Ebbels, S., ... Whitehouse, A. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, 58(10), 1068–1080. <https://doi.org/10.1111/JCPP.12721>
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & Consortium, C. (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLOS ONE*, 11(7), Article e0158753. <https://doi.org/10.1371/journal.pone.0158753>
- Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer* (Version 6.4.15) [Computer software]. <http://www.praat.org/>
- Botting, N., & Adams, C. (2005). Semantic and inferencing abilities in children with communication disorders. *International Journal of Language & Communication Disorders*, 40(1), 49–66. <https://doi.org/10.1080/13682820410001723390>
- Brisson, C. (1998). *Distributivity, maximality, and floating quantifiers* [Doctoral dissertation, Rutgers University].
- Brisson, C. (2003). Plurals, *all*, and the nonuniformity of collective predication. *Linguistics and Philosophy*, 26(2), 129–184. <https://doi.org/10.1023/A:1022771705575>
- Brooks, P. J., & Braine, M. D. S. (1996). What do children know about the universal quantifiers *all* and *each*? *Cognition*, 60, 235–268. [https://doi.org/10.1016/0010-0277\(96\)00712-3](https://doi.org/10.1016/0010-0277(96)00712-3)
- Calder, S. D., Brennan-Jones, C. G., Robinson, M., Whitehouse, A., & Hill, E. (2022). The prevalence of and potential risk factors for developmental language disorder at 10 years in the Raine Study. *Journal of Paediatrics and Child Health*, 58(11), 2044–2050. <https://doi.org/10.1111/jpc.16149>
- Champollion, L. (2016). Covert distributivity in algebraic event semantics. *Semantics and Pragmatics*, 9, 15. <https://doi.org/10.3765/sp.9.15>
- Champollion, L. (2019). Distributivity in formal semantics. *Annual Review of Linguistics*, 5(1), 289–308. <https://doi.org/10.1146/annurev-linguistics-011718-012528>
- Champollion, L. (2020). Distributivity, collectivity and cumulativity. In L. Mathewson, C. Meier, H. Rullmann, & T. E. Zimmermann (Eds.), *The Wiley Blackwell companion to semantics*. Wiley. <https://doi.org/10.1002/9781118788516.sem021>
- Cowan, R., Donlan, C., Newton, E. J., & Lloyd, D. (2005). Number skills and knowledge in children with specific language impairment. *Journal of Educational Psychology*, 97, 732–744. <https://doi.org/10.1037/0022-0663.97.4.732>

- de Koster, A. M. B., Dotlačil, J., & Spenader, J. K. (2017). Children's understanding of distributivity and adjectives of comparison. In M. LaMendola & J. Scott (Eds.), *Proceedings of the 41st annual Boston University Conference on Language Development* (pp. 373–386). Cascadilla Press.
- de Koster, A. M. B., Hendriks, P., & Spenader, J. K. (2021). Child-like adults: Dual-task effects on collective vs. distributive sentence interpretations. *Frontiers in Psychology*, 12, 556120. <https://doi.org/10.3389/fpsyg.2021.556120>
- de Koster, A. M. B., Spenader, J. K., & Hendriks, P. (2018). Are children's overly distributive interpretations and spreading errors related? In A. B. Bertolini & M. J. Kaplan (Eds.), *Proceedings of the 42nd annual Boston University Conference on Language Development* (pp. 413–426). Cascadilla Press.
- Delage, H., & Frauenfelder, U. H. (2020). Relationship between working memory and complex syntax in children with developmental language disorder. *Journal of Child Language*, 47(3), 600–632. <https://doi.org/10.1017/S0305000919000722>
- Donlan, C., Cowan, R., Newton, E. J., & Lloyd, D. (2007). The role of language in mathematical development: Evidence from children with specific language impairments. *Cognition*, 103, 23–33. <https://doi.org/10.1016/j.cognition.2006.02.007>
- Dotlačil, J. (2010). *Anaphora and distributivity: A study of same, different, reciprocals and others*. LOT International Series. LOT.
- Dotlačil, J. (2015). Why is distributivity so hard? New evidence from distributive markers and licensors in Czech. In *Slavic Languages in the perspective of formal grammar: Proceedings of FDSL 10.5, Brno 2014* (pp. 139–154). Peter Lang.
- Dotlačil, J., & Brasoveanu, A. (2021). The representation and processing of distributivity and collectivity: Ambiguity vs. underspecification. *Glossa: A Journal of General Linguistics*, 6, 6. <https://doi.org/10.5334/gjgl.1131>
- Drozd, K., van der Lely, H. K. J., & Montalto, R. (2017). Children's comprehension of distributive universal quantification. *Lingua*, 198, 89–109. <https://doi.org/10.1016/j.lingua.2017.07.005>
- Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 52, 321–335. [https://doi.org/10.1044/1092-4388\(2009/07-0189\)](https://doi.org/10.1044/1092-4388(2009/07-0189))
- Frazier, L., Pacht, J. M., & Rayner, K. (1999). Taking on semantic commitments II: Collective versus distributive readings. *Cognition*, 70, 87–104. [https://doi.org/10.1016/S0010-0277\(99\)00002-5](https://doi.org/10.1016/S0010-0277(99)00002-5)
- Gennari, S., & MacDonald, M. (2005/2006). Acquisition of negation and quantification: Insights from adult production and comprehension. *Language Acquisition*, 13(2), 125–168. https://doi.org/10.1207/s15327817la1302_5
- Georgiou, G. P., & Theodorou, E. (2023). Comprehension of complex syntax by non-English-speaking children with developmental language disorder: A scoping review. *Clinical Linguistics & Phonetics*, 37(11), 1050–1068. <https://doi.org/10.1080/02699206.2022.2135024>
- Gil, D. (1982). Quantifier scope, linguistic variation, and natural language semantics. *Linguistics and Philosophy*, 5, 421–472. <https://doi.org/10.1007/BF00355582>

- Grinstead, J., Padilla-Reyes, R., & Nieves-Rivera, M. (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>
- Haebig, E., Saffran, J. R., & Ellis Weismer, S. (2017). Statistical word learning in children with autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry*, 58, 1251–1263. <https://doi.org/10.1111/jcpp.12734>
- Hansen, B. H., Oerbeck, B., Skirbekk, B., Petrovski, B. É., & Kristensen, H. (2018). Neurodevelopmental disorders: Prevalence and comorbidity in children referred to mental health services. *Nordic Journal of Psychiatry*, 72(4), 285–291. <https://doi.org/10.1080/08039488.2018.1444087>
- Heim, I., Lasnik, H., & May, R. (1991). Reciprocity and plurality. *Linguistic Inquiry*, 22, 63–102.
- Henry, L. A., Messer, D. J., & Nash, G. (2012). Executive functioning in children with specific language impairment. *Journal of Child Psychology and Psychiatry*, 53(1), 37–45. <https://doi.org/10.1111/j.1469-7610.2011.02430.x>
- Kapa, L. L., & Mettler, H. M. (2024). Statistical learning among preschoolers with and without developmental language disorder: Examining effects of language status, age, and prior learning. *Journal of Speech, Language, and Hearing Research*, 67(9), 3081–3093. https://doi.org/10.1044/2024_JSLHR-23-00602
- Katsos, N., Roqueta, C. A., Clemente Estevan, R. A., & Cummins, C. (2011). Are children with specific language impairment competent with the pragmatics and logic of quantification? *Cognition*, 119(1), 43–57. <https://doi.org/10.1016/j.cognition.2010.12.004>
- Kaup, B., Kelter, S., & Habel, C. (2002). Representing referents of plural expressions and resolving plural anaphors. *Language and Cognitive Processes*, 17(4), 405–450. <https://doi.org/10.1080/01690960143000272>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package ‘lmerTest’. *R Package Version*, 2, 734.
- Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2017). Statistical learning in specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 60, 3474–3486. https://doi.org/10.1044/2017_JSLHR-L-16-0439
- Landman, F. (1989). Groups, I. *Linguistics and Philosophy*, 12(5), 559–605. <https://doi.org/10.1007/BF00627774>
- Landman, F. (1995). Plurality. In S. Lappin (Ed.), *The handbook of contemporary semantic theory*. Blackwell.
- Landman, F. (2000). *Events and plurality: The Jerusalem lectures*. Kluwer. <https://doi.org/10.1007/BF00627774>
- Lasersohn, P. (1995). *Plurality, conjunction and events*. Springer. <https://doi.org/10.1007/978-94-015-8581-1>
- Leonard, L. B. (2014). *Children with specific language impairment*. MIT Press. <https://doi.org/10.7551/mitpress/9152.001.0001>

- Leonard, L. B., Deevy, P., Fey, M. E., & Bredin-Oja, S. L. (2013). Sentence comprehension in specific language impairment: A task designed to distinguish between cognitive capacity and syntactic complexity. *Journal of Speech, Language, and Hearing Research*, 56(2), 577–589. [https://doi.org/10.1044/1092-4388\(2012/11-0254\)](https://doi.org/10.1044/1092-4388(2012/11-0254))
- Leonard, L. B., Weismer, S. E., Miller, C. A., Francis, D. J., Tomblin, J. B., & Kail, R. V. (2007). Speed of processing, working memory, and language impairment in children. *Journal of Speech, Language, and Hearing Research*, 50(2), 408–428. [https://doi.org/10.1044/1092-4388\(2007/029\)](https://doi.org/10.1044/1092-4388(2007/029))
- Link, G. (1983). The logical analysis of plurals and mass terms: A lattice-theoretic approach. In P. Portner & B. H. Partee (Eds.), *Formal semantics: The essential readings* (pp. 127–147). Blackwell. <https://doi.org/10.1002/9780470758335.ch4>
- Link, G. (1987). Generalized quantifiers and plurals. In P. Gärdenfors (Ed.), *Generalized quantifiers* (pp. 151–180). Springer. https://doi.org/10.1007/978-94-009-3381-1_6
- Lohiniva, K., & Panizza, D. (2016). Children’s interpretation of ambiguous plural sentences in context. In J. Scott & D. Waughtal (Eds.), *Proceedings of the 40th annual Boston University Conference on Language Development* (pp. 216–228). Cascadia Press.
- Lønning, J. T. (1997). Plurals and collectivity. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language*. Elsevier. <https://doi.org/10.1016/B978-044481714-3/50023-0>
- McGregor, K. K. (2020). How we fail children with developmental language disorder. *Language, Speech, and Hearing Services in Schools*, 51(4), 981–992. https://doi.org/10.1044/2020_LSHSS-20-00003
- Montgomery, J. W. (2000). Verbal working memory and sentence comprehension in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 43(2), 293–308. <https://doi.org/10.1044/jslhr.4302.293>
- Montgomery, J. W. (2003). Working memory and comprehension in children with specific language impairment: What we know so far. *Journal of Communication Disorders*, 36(3), 221–231. [https://doi.org/10.1016/S0021-9924\(03\)00021-2](https://doi.org/10.1016/S0021-9924(03)00021-2)
- Montgomery, J. W., & Evans, J. L. (2009). Complex sentence comprehension and working memory in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 52(2), 269–288. [https://doi.org/10.1044/1092-4388\(2008/07-0116\)](https://doi.org/10.1044/1092-4388(2008/07-0116))
- Montgomery, J. W., Evans, J. L., Fargo, J. D., Schwartz, S., & Gillam, R. B. (2018). Structural relationship between cognitive processing and syntactic sentence comprehension in children with and without developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 61(12), 2950–2976. https://doi.org/10.1044/2018_JSLHR-L-17-0421
- Montgomery, J. W., Gillam, R. B., & Evans, J. L. (2016). Syntactic versus memory accounts of the sentence comprehension deficits of specific language impairment: Looking back, looking ahead. *Journal of Speech, Language, and Hearing Research*, 59(6), 1491–1503. https://doi.org/10.1044/2016_JSLHR-L-15-0302
- Montgomery, J. W., Gillam, R. B., & Evans, J. L. (2021). A new memory perspective on the sentence comprehension deficits of school-age children with developmental language disorder: Implications for theory, assessment, and intervention. *Language, Speech, and Hearing Services in Schools*, 52(2), 449–466. https://doi.org/10.1044/2021_LSHSS-20-00128

- Musolino, J. (2009). The logical syntax of number words: Theory, acquisition and processing. *Cognition*, 111, 24–45. <https://doi.org/10.1016/j.cognition.2008.12.008>
- Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, 57(11), 1247–1257. <https://doi.org/10.1111/jcpp.12573>
- Padilla-Reyes, R. E. (2018). *Connections among scales, plurality, and intensionality in Spanish* [Doctoral dissertation, The Ohio State University].
- Pagliarini, E., Fiorin, G., & Dotlačil, J. (2012). The acquisition of distributivity in pluralities. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of the 36th annual Boston University Conference on Language Development* (pp. 387–399). Cascadia Press.
- Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, 101(2), 385–413. <https://doi.org/10.1016/j.cognition.2006.04.008>
- Psychology Software Tools, Inc. (2016). *E-Prime 3.0* [Computer software]. <https://support.pstnet.com/>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Roid, G. H., Miller, L. J., Pomplun, M., & Koch, C. (2013). *Leiter International Performance Scale – Third Edition (LEITER-3)*. Western Psychological Services.
- RStudio Team. (2020). *RStudio: Integrated development environment for R* [Manual]. RStudio, PBC.
- Saponaro C., Carioti, D., Riva, M., & Guasti, M. T. (2025). Reaching meaning through language: What can children tell us about distributivity? *Journal of Child Language*. Published online, 1–21. <https://doi.org/10.1017/S0305000925000224>
- Sarnecka, B. W., & Gelman, R. (2004). Six does not just mean a lot: Preschoolers see number words as specific. *Cognition*, 92(3), 329–352. <https://doi.org/10.1016/j.cognition.2003.10.001>
- Syrett, K., & Musolino, J. (2013). Collectivity, distributivity, and the interpretation of plural numerical expressions in child and adult language. *Language Acquisition*, 20(4), 259–291. <https://doi.org/10.1080/10489223.2013.828060>
- Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, 40(6), 1245–1260. <https://doi.org/10.1044/jslhr.4006.1245>
- Ussery, C. (2007). Processing plural DPs: Collective, cumulative, and distributive interpretations. *University of Massachusetts Occasional Papers in Linguistics*, 33.
- van der Lely, H. K. J. (1998). SLI in children: Movement, economy, and deficits in the computational-syntactic system. *Language Acquisition*, 7(2–4), 161–192. https://doi.org/10.1207/s15327817la0702-4_4
- Vendler, Z. (1967). *Linguistics in philosophy*. Cornell University Press. <https://doi.org/10.7591/9781501743726>

- Whitehouse, A. J. (2010). Is there a sex ratio difference in the familial aggregation of specific language impairment? A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 53(4), 1015–1025. [https://doi.org/10.1044/1092-4388\(2009/09-0078\)](https://doi.org/10.1044/1092-4388(2009/09-0078))
- Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer. https://doi.org/10.1007/978-3-319-24277-4_9
- Wickham, H., François, R., Henry, L., & Müller, K. (2015). *Dplyr: A grammar of data manipulation* (R package version 0.4.3). <https://cran.r-project.org/package=dplyr>
- Wiig, E. H., Semel, E., & Secord, W. A. (2018). *Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5)*. NCS Pearson Assessments. (Adaptación Española/ Spanish Adaptation). ISBN:978-84-9035-623-4.
- Winter, Y. (2000). Distributivity and dependency. *Natural Language Semantics*, 8, 27–69. <https://doi.org/10.1023/A:1008313715103>
- Wittke, K., & Spaulding, T. J. (2018). Which preschool children with specific language impairment receive language intervention? *Language, Speech, and Hearing Services in Schools*, 49(1), 59–71. https://doi.org/10.1044/2017_LSHSS-17-0024
- World Health Organization. (2022). *ICD-11: International classification of diseases* (11th revision). <https://icd.who.int/>
- Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, 24(2), 220–251. [https://doi.org/10.1016/0010-0285\(92\)90008-P](https://doi.org/10.1016/0010-0285(92)90008-P)
- Young, A. R., Beitchman, J. H., Johnson, C., Douglas, L., Atkinson, L., Escobar, M., & Wilson, B. (2002). Young adult academic outcomes in a longitudinal sample of early identified language impaired and control children. *Journal of Child Psychology and Psychiatry*, 43(5), 635–645. <https://doi.org/10.1111/1469-7610.00052>
- Zhang, X., & Tomblin, J. B. (2000). The association of intervention receipt with speech-language profiles and social-demographic variables. *American Journal of Speech-Language Pathology*, 9(4), 345–357. <https://doi.org/10.1044/1058-0360.0904.345>



