Making beds and dying of boredom literally: A developmental study on the comprehension of nonliteral uses of language in autism

Marta Ponciano
University of the Basque Country (UPV/EHU), Spain

Agustín Vicente
Ikerbasque, Basque Foundation for Science
University of the Basque Country (UPV/EHU), Spain

José V. Hernández-Conde
University of Valladolid, Spain

Elena Castroviejo
University of the Basque Country (UPV/EHU), Spain

Abstract: Comprehension of simple nonliteral uses of language was investigated in three- to nine-year-old autistic and linguistically matched typically developing (TD) children, by assessing their understanding of nonliteral uses of language with potential literal senses. Children were tested on conventional metaphors, idioms, hyperboles, and light verb constructions. The aim of the study was to determine whether autistic children showed a genuinely strong tendency to interpret nonliteral uses of language literally across development. A total of 166 children (N = 42 Autistic children; N= 124 TD children) were tested using a paradigm with online (response times) and offline (picture selection) measures. Overall, there were no significant group differences on the picture selection task, but autistic children were slower in spite of increasing verbal age. Both groups showed continuous improvement of their understanding of literal and nonliteral senses with increasing verbal mental age. The results, nevertheless, call for a reflection on the (possible) literalist behavior in autism, indicating that it is important to take into account individual variation, as we observed different kinds of performance within the autistic group.

Keywords: nonliteral uses of language; autism; literalism; development; heterogeneity

Corresponding author(s): Marta Ponciano, Department of Linguistics and Basque Studies, University of the Basque Country (UPV/EHU), Centro de Investigación Micaela Portilla Ikergunea (UPV/EHU), Justo Vélez de Elorriaga, 1, 01006 Vitoria-Gasteiz, Spain. Email: marta.ponciano@ehu.eus.

ORCID ID(s): Marta Ponciano: https://orcid.org/0000-0002-7062-0518; Agustín Vicente: https://orcid.org/0000-0002-5331-5546; José V. Hernández-Conde: https://orcid.org/0000-0002-8502-6570; Elena Castroviejo: https://orcid.org/0000-0003-0371-1214

Citation: Ponciano, M., Vicente, A., Hernández-Conde, J.V. & Castroviejo, E. (2025). Making beds and dying of boredom literally: A developmental study on the comprehension of nonliteral uses of language in autism. *Language Development Research, 5*(2), 93–129. https://doi.org/10.34842/ldr2025-826

Introduction

Nonliteral Language in Autism

Difficulties understanding nonliteral uses of language by interpreting them literally are common across the autistic spectrum in everyday situations (Morra, 2016). Literalism has been taken to be characteristic of autism since the pioneering investigations of Kanner (1943), who observed that autistic individuals exhibited an unusual tendency to interpret non-literal language literally, and Asperger (1991), who noted that the children he studied exhibited characteristic difficulties in understanding jokes (Geurts et al., 2019).

However, previous studies in laboratory settings have typically yielded contradictory results (Gernsbacher & Pripas-Kapit, 2012; Kalandadze et al., 2018; Lampri et al., 2023; Vicente et al., 2024). Some studies report systematic difficulties, often showing tendencies to interpret nonliteral language in a literal manner, while others indicate no significant differences from typically developing controls. The discrepancies between studies' findings may depend on the type of nonliteral use that is being tested or the methodology used (Kalaldadze et al., 2018). Concerning types of nonliteral uses tested, many studies on autistic children, adolescents or young adults report special difficulties understanding irony and sarcasm (Deliens et al., 2018; Happé, 1993; MacKay & Shaw, 2004; Saban-Bezalel et al., 2019), while several studies targeting scalar implicatures or indirect speech acts do not find similar differences (on scalar implicatures: Chevallier et al., 2010; Su & Su, 2015; Hochstein et al., 2018; Pijnacker et al., 2009; Van Tiel & Kissine, 2018; and Schaeken et al., 2018, and Mazzaggio et al., 2021 for discrepant results; on indirect speech acts: Kissine et al., 2015, Marocchini et al., 2022; and Paul & Cohen, 1985, and Ozonoff & Miller, 1996 for opposite results).

Focusing on figurative language, results on autistic individuals' comprehension of metaphors and idioms are mixed (Kalandadze et al., 2018, 2019; Lampri et al., 2024). Since Norbury's (2005) seminal work, many researchers hold that differences in metaphor comprehension are related to differences in general linguistic abilities, such that if autistic and non-autistic participants are matched on linguistic abilities, no significant differences should emerge in metaphor comprehension (Andrés-Roqueta & Katsos, 2017). However, this "structural language hypothesis" does not account for all the variability observed: some studies that match autistic and non-autistic individuals on linguistic abilities have found that autistic people exhibit more difficulties than neurotypicals with metaphor interpretation. In particular, several of these studies find that autistic participants exhibit a stronger tendency than neurotypicals to understand metaphors, as well as idiomatic expressions, literally (Chahboun et al., 2016, 2017; Vulchanova et al., 2012; Walenski & Love, 2017). Even if autistic participants in these studies end up selecting the nonliteral interpretation in a forced-choice task, online measures such as eye-tracking reveal a more significant interference of literal options in the autistic population than in neurotypicals (O Shea et al., 2024, Martin-González et al., 2025).

There is not much work on the development of figurative language comprehension (or on nonliteral language comprehension in general) in autism research. Rundblad & Annaz (2010) tested *conventional* metaphor comprehension in autistic children (n=11), with ages ranging from 4 to 11, using a picture selection task. Results showed difficulties in all ages, as well as a lack of relation between verbal mental age (VMA) and accuracy, which in principle is surprising since other studies with older children (10-12 yr.olds) have found such a relation (Kasirer & Mashal, 2014, 2016; see also Olofson et al., 2014, and Pastor-Cerezuela et al., 2020, who, in contrast with Kasirer & Mashal, 2016, found novel metaphors to be more difficult than conventional metaphors for autistic children). In a cross-sectional similar study on *novel* metaphors with autistic children and teens, Van Herwegen & Rundblad (2018) found more difficulties in the autistic group than in a control group matched on chronological age. Van Herwegen & Rundblad selected some of the youngest participants of their initial sample (children of around 9 years) for a longitudinal study, which showed improvements on metaphor comprehension related to age. In a recent study, also on novel metaphors, Martin-González et al. (2025) tested two groups of children, autistic and non-autistic, matched on VMA (chronological ages ranging from 3 to 12), on a picture selection task. The gaze patterns of the children were recorded. The results did not find differences between the autistic and the non-autistic groups in the picture selection task, but there were significantly more looks at literal competitors in the autistic group in the metaphorical condition than in the TD group.

In sum, several decades of research on nonliteral uses of language in autism has thrown mixed results, with some evidence suggesting that autistic individuals may interpret nonliteral language literally more often than neurotypicals. In this paper we add to the developmental research on literalism in autism: we explored whether autistic children have easier access than their TD peers to the literal interpretations of ambiguous expressions, and whether TD children have easier access than their autistic peers to the nonliteral interpretations of such expressions.

Our main research question is about a particular kind of literalism. There are several different notions of literal meaning at the level of utterances (Allott & Textor, 2022). Depending on the notion at stake, literalism can be understood in one way or another. For the purposes of this paper, we propose to understand a literal interpretation of any given utterance as a thoroughly compositional interpretation, i.e., one generated on the basis of fixed rules of grammar that take as input stable and concrete meaning assignments. Correspondingly, by nonliteral uses of language, or nonliteral language, we mean any use of language that is intended not to be processed compositionally. According to this notion of *the literal*, the metaphor *Juliet is the Sun* counts as a nonliteral use of language, but also do hyperboles such as *to die of boredom*, conventional metaphors such as *to be a pig* (meaning: to be greedy and selfish), and even light verb constructions such as *to take the bus* or *to make the bed*. Our main research question, thus, is whether autistic individuals at some point in development exhibit this kind of literalism, i.e., whether they exhibit a more significant tendency than neurotypicals to understand utterances in a word-by-word, compositional manner.

As we are aware that this notion of the literal may be controversial, the following paragraphs argue why we consider it a legitimate notion. We take it that the most controversial point concerns light verb constructions (LVCs). We follow Fleischer (1997), Fellbaum et al. (2006) and Wittenberg (2016) in considering that light verbs are "light" because they have lost most of their semantic weight, such a loss representing a shift away from the literal. According to this view, when a light verb appears in a LVC, the verb is no longer being used in its original, literal sense, and instead, takes on a quasi-auxiliary role within the construction. However, LVCs do also retain a literal interpretation in the sense here characterized (Wittenberg, 2016). Such literal interpretation is given by applying rules of grammar to meanings that represent specific rather than unspecific notions (Taylor, 2006). Light verb constructions are not dissimilar from other idiomatic expressions, where the meaning of the whole cannot be derived from the literal meanings of its parts. For this reason, we think that LVCs can be considered part of the spectrum of nonliteral language (Fleischer, 1997, Fellbaum et al., 2006). Yet, we do not need to endorse such a view; what we are after is the contrast between a literal interpretation of a LVC, as we have defined it, and an interpretation that is not literal in that same sense.

The most habitual interpretation of a LVC is its nonliteral interpretation. *To make the bed* is almost never understood as "to create or generate a bed". Actually, the nonliteral meaning of the phrase is arguably its *conventional* meaning. On most occasions, if someone is told to make a bed, they are expected to lay down sheets, etc.; otherwise, they would be taken as not complying with the directive. This suggests that what we call the *nonliteral meaning* of a LVC may actually be taken to be its literal meaning. While this is correct, the issue is terminological: we use a notion of literal meaning such that the conventional and the literal do not have to coincide. Actually, we are interested in finding out whether autistic children at some point of development may go for a compositional meaning of a phrase instead of going for its conventional and habitual meaning.

The metaphors, hyperboles and idioms that we test in our study, together with items displaying LVCs, are instances of nonliteral conventional language. However, to say that they are instances of *conventional* nonliteral language does not mean that the nonliteral meaning is, or may be, the conventional meaning. In the case of metaphors and hyperboles, *conventional* is opposed to *novel*. That is, to say that a metaphor is conventional is, roughly, to say that it has been lexicalized, and has become another meaning of the word (see Bowdle & Gentner, 2005). However, this does not entail that such a meaning has become conventional in the sense of being dominant.

These expressions are also part of utterances whose more frequent meaning is their nonliteral meaning. For example, *to die of boredom* rarely, if ever, means to be killed by boredom. In cases like this hyperbolic use of *die*, however, it is arguably less controversial to hold that the most frequent meaning is not the literal meaning of the expression. This is because at the word level speakers would have it clear that *die* has the literal meaning it has. That is, even though at the phrase level the nonliteral meaning of *to die of boredom* is clearly more frequent than its compositional meaning, the verb *die* is still dominantly used in its literal sense. However, despite there being such a difference between e.g., *to take the bus* and *to die of boredom* with respect to verb meanings, we take it that it is a difference in degree: *take* is used in its nonliteral, modified, sense more often than *die*, such that *take*'s literal, concrete, sense is more difficult to access than the literal meaning of *die* in nonliteral constructions.

In sum, the purpose of the study was to explore whether autistic children exhibit more facility than TDs in interpreting expressions whose more frequent meaning is their nonliteral meaning in a literal, compositional way. We also wanted to investigate how difficult it is for autistic children to acquire nonliteral meanings, even when such meanings are very familiar and salient. Understanding these differences is important, as it can provide insights into the cognitive and linguistic processing styles of autistic individuals. The study thus seeks to contribute to broader discussions in autism research and pragmatic theory, ultimately fostering more effective communication strategies and educational approaches tailored to the needs of autistic children.

Current Study

In this study, we tested autistic children ranging in age from four to nine, with verbal mental ages (VMA) ranging from three to nine, on their understanding of familiar nonliteral expressions in their L1, Spanish. Autistic children and TD children matched on VMA were compared. The (main) interest of the study was to compare the two groups' development of nonliteral language comprehension based on linguistic development, focusing particularly on response times. This focus allowed us to evaluate how easily each group processed literal versus nonliteral interpretations of familiar expressions.

In Castroviejo et al. (2024), we conducted the typical developmental study using offline (picture selection) and online (response time) data collection from the picture selection task used in this paper. Unlike the present study, which treats VMA as a continuous variable, the aforementioned study grouped participants into discrete age categories. We observed a linear progression across ages in *nonliteral* readings. The progression started with around 68% of correct responses at age three and reached *ceiling* performance in the oldest group (age nine). The development of *literal* interpretations was less straightforward: three-year-olds exhibited an accuracy of around 75%, and performance remained at that level until they were five. From five to eight they became more accurate. Concerning reaction times (RT), it was observed that access to literal interpretations was easier than access to nonliteral interpretations for three- and four-year-olds, that both kinds of interpretations began to be equally accessible at around five years, and that the pattern may reverse again at nine.

The results indicated distinct patterns in children's performance between two groups of experimental items: those categorized as Light Verb Constructions (LVC) and those classified under (what we called) the Metaphor-Hyperbole-and-Idiom continuum (MHI). For example, LVC items involved expressions like *hacer la cama* 'to make the bed', where the verb itself contributes with little semantic content. In these LVC items, children were more accurate in the nonliteral than the literal condition. In contrast, items in the MHI category, which included metaphorical expressions like *ser una tortuga* (*lit*. 'to be a turtle'; meaning: 'to be slow'), idioms like *partirse de risa* (*lit*. 'to split oneself with laughter'; meaning: 'to burst out laughing'), and hyperboles like *morirse de aburrimiento* ('to die of boredom') showed that younger children (ages three to five) were more accurate in the literal condition than in the nonliteral one.

The task was a forced-choice task including no competitors (i.e., in the literal condition, a representation of the nonliteral meaning; and in the nonliteral condition, a representation of the literal meaning). Children had to decide between a representation of either the literal or the nonliteral meaning, and two distractors. This paradigm was chosen primarily because the study aimed to compare the accuracy and ease of access to literal meanings *versus* the accuracy and ease of access to nonliteral meanings in both groups, rather than to examine whether children preferred one interpretation over the other. We also thought that the inclusion of competitors might confound the results, particularly when the research question focuses on how accurately and rapidly children access one meaning versus the other. Children may be able to understand a nonliteral expression and do so quickly, but this might not be the case if a literal competitor is explicitly presented. For instance, Köder and Falkum (2020) reported that four- and five-year old TD children exhibited a stronger tendency than younger and older children to choose a representation of the literal meaning of a metonymy, even though their gaze patterns were similar to those of other ages. What Köder and Falkum observed may mean that four-and five-year-old TDs understand metonymies, but that, when prompted to choose between competing representations in a picture selection task, they chose the literal one. Finally, we also considered that including competitors could affect the autistic group more than the non-autistic group, since autistic children may experience more uncertainty in general than TD children (Vicente et al., 2024).

Research Questions and Hypotheses

As said, the main goal of this study is to test the literalism trend in autistic and non-autistic children matched on verbal mental age across development. By assessing performance on familiar expressions with a more frequent or dominant nonliteral use, we can observe (a) whether autistic children experience more difficulties accessing nonliteral meanings across development, and (b) whether autistic children have a comparably easier access to literal meanings, even when such literal meanings are unusual. Our research questions are spelled out below.

RQ1: Are there any differences in the performance of TD and autistic children when interpreting nonliteral expressions across development (from 3 to 9 years of verbal mental age)?

RQ2: Do autistic children exhibit an easier access to literal, compositional, meanings at different developmental stages?

We hypothesized that autistic children would have more difficulties understanding nonliteral uses of language across the board than TD peers. At the same time, in keeping with the idea that autistic individuals are more literalist than non-autistic individuals (Vicente & Falkum, 2023), we expected autistic children to experience less difficulties than TD children understanding the literal but uncommon use of the expressions in our sample. We thus hypothesized that autistic children would perform better than TDs in the literal condition. Concerning latencies, we hypothesized that the said differences between autistic and TD children should have an effect on response times (RT): in general, we thought we could observe autistic children accessing literal meanings faster than nonliteral meanings. However, we were more confident about observing the reverse pattern in the TD group: TD children should access nonliteral interpretations faster than literal, word-by-word interpretations. We considered the prediction about autistic children's faster access to literal interpretations to be just one possibility, given the nature of the items with highly frequent nonliteral meanings and the age range of the participants. In any event, we did expect to find this kind of difference in access in the youngest autistic children.

We did not have specific predictions regarding the classification of items (LVC category *vs* MHI category) for autistic participants. The investigation of this dimension was more exploratory in nature. We did not predict anything in development either regarding those two item categories. Previous literature has focused extensively on metaphors, metonymies, and irony, but not so much on conventionalized expressions such as LVCs. For this reason and the observed bimodal distribution in typically developing children's performance, we included those two categories into the analyses.

According to the literalist hypothesis, in atypical development we expected to observe a more prolonged literalist stage, where children would perform better in the literal than in the nonliteral condition. With increasing verbal mental age, autistic children should improve on the nonliteral interpretation, since there should be an effect of frequency corresponding to an improvement on linguistic skills generally. We also expected them to be more sensitive to frequency effects as their VMA increased, resulting in more difficulties in accessing literal interpretations.

Method

Participants

We tested a total of 166 Spanish speaking children aged three to nine–years. Children were divided into two groups based on whether they were typically developing (TD) (N=124) or had a diagnosis of autism (N=42) (see Table 1 for the participants' descriptive data). The control group was recruited from participation in the previous study (Castroviejo et al., 2024).

For this experiment, 42 Spanish-speaking children (9 girls and 33 boys) diagnosed with autism were recruited (a) through the Early Intervention program (Atención Temprana, Álava), Araba, Spain; (b) from the APNABI Autism association (Asociación de familias de personas con un Trastorno del Espectro del Autismo), Bizkaia, Spain; (c) from the Ilargia Intervention Center, in Logroño, Spain, and (d) from mainstream public schools, Bizkaia, Spain. Most of the children had received a formal diagnosis of autism in conformity to the criteria of DSM-5 (and/or ICD11) by a team of professional neuropsychologists, pediatricians and speech therapists specialized in ASD diagnosis. For those who had not received a diagnosis prior to the study, diagnosis of autism was confirmed by a licensed clinician on the basis of DSM-criteria, using The Autism Diagnostic Observation Schedule (ADOS-2) (Lord et al., 2012) and sometimes additionally Autism Diagnostic Interview – Revised (ADI-R) (Rutter et al. 2003). All autistic participants were attending mainstream classrooms following inclusive education criteria, and the majority of them were receiving intervention (either group intervention, individual intervention, or both, via public or private services).

We did not conduct a formal power analysis before starting the study, so the sample size was not determined based on statistical criteria. Instead, we recruited as many eligible children as we could within the age range of interest, guided by previous literature and the design of the broader study. In the end, practical factors like time constraints and participant availability largely shaped the final sample size.

The chronological age (CA) of autistic children ranged from four to nine years (M_{age}= 80.69 months, *SD*= 18.29 months). Autistic children's nonverbal intellectual abilities, as measured by the Leiter-3 scale (Roid, Miller et al., 2013), ranged from 73 to 125 (*M*=98.88; *SD*=11.87), with average or above-average non-verbal intelligence. Two children's cognitive measures were absent, and mean imputation was used to replace those missing values. There was no evidence to believe that they were below the threshold on non-verbal abilities. Also, there was no mention of any intellectual disability in their diagnoses.

A total of 143 Spanish-speaking TD children were initially recruited from various mainstream public schools in Bizkaia, North of Spain; 19 children were eventually excluded from analyses (due to one of the following reasons: i.a. being in the process of receiving a diagnosis of Generalized/Pervasive developmental disorder (PDD-NOS)/Autism Spectrum Disorder (ASD), poor performance (below chance levels) in the filler items or being absent from the testing sessions). The final group consisted of 124 Spanish-speaking children (55 boys and 69 girls) between the ages of 2;11 and 9;11 (year, month) (M_{age}=72.05 months; *SD*=22.88 months, among the TD children whose results were included within the analyses, one was slightly below the age of three, 2;11;29 (years; months; days)).

Groups were matched on receptive vocabulary as measured by the Peabody Picture Vocabulary Scale/Test (PPVT-3, Dunn et al., 2006; ASD: $M_{age\ equivalent}$ = 67.45 months, *SD* = 20.65 (34-119); TD: $M_{age\ equivalent}$ = 72.05 months, *SD* =22.88 (35-119)), t(77) = -1.21 , p = .22, although TD children were not tested with the PPVT. Their receptive vocabulary measures were assumed to be in line with their chronological age (CA), meaning their age equivalents were considered to be similar or equal to their chronological age. The assumption was necessary because the TD group was originally part of a developmental study where VMA was not a variable of interest. Autistic children were significantly older (t(87)= 2.47, p = .01) than TD children, and had a significant difference between their CA and their VMA (t(41) =4.71, p = 2.753 e-05).

Table 1. ***Participants' descriptive data.***

	ASD		**TD**	
N (*M*:*F*)	42 (33:9)		124 (55:69)	
Variable	Mean (*SD*)	Range	Mean (SD)	Range
Age in months	80.69 (18.29)	49-118	72.05 (22.88)	35-119
VMA in months	67.45 (20.65)	34-119	72.05 (22.88)	35-119
Non-verbal IQ	98.88 (11.87)	73-125	-	-

Notes: Matching measures are in boldface. VMA (Verbal Mental Age) age equivalence (standardized score) is the verbal mental age measure obtained from PPVT-III. It is an age-scaled score from the direct score. Non-verbal IQ test (standard scores) from LEITER-3 test, composite scores have a mean of 100 and an SD of 15. In the case of TD children CA and VMA are assumed to be the same. TD, typically developing; ASD, autism spectrum disorders.

Materials

Stimuli consisted of Spanish expressions whose nonliteral interpretation is highly frequent, but whose literal reading is plausible though awkward. Frequency was tested by means of a norming study (Castroviejo et al., 2024). Participants were tested on the interpretation of expressions including hyperboles (*morir de aburrimiento* 'to die of boredom'), metaphors (*ser una tortuga*; *lit*. 'to be a turtle'; *nonlit*. 'to be slow'), idioms (*partirse de risa*; *lit*. 'to split oneself with laughter'; *nonlit*. 'to burst out laughing'), and light verb constructions (*hacer la cama* 'to make the bed'), as in (1).

(1) a. Sergio se muere de aburrimiento.
Sergio SE dies of boredom
'Sergio is dying of boredom.'

b. Unax es una tortuga.
Unax is a turtle
'Unax is a turtle.'

c. Tania se parte de risa.
Tania SE breaks of laughter
'Tania bursts out laughing.'

d. Juan hace la cama.
Juan makes the bed
'Juan makes the bed.'

All stimuli were a combination of a visual array of three pictures and an orally presented sentence. Only one picture reflected the target meaning of the sentence, and the other two were distractors. We created 16 critical sentences that contained the aforementioned linguistic expressions (see Appendix, Table A1.). There were also filler items. Each sentence had a literal and nonliteral version/ visual representation, in which the same linguistic expression is used to refer to the literal or nonliteral interpretation (see Figure 1). Participants saw 8 nonliteral and 8 literal versions of the sentences. The purpose of this design was to assess participants' ability to access each meaning (frequent non-compositional or infrequent but compositional) independently. Each participant saw a total of 32 experimental trials, 16 critical sentences (8 nonliteral and 8 literal) and 16 filler items.

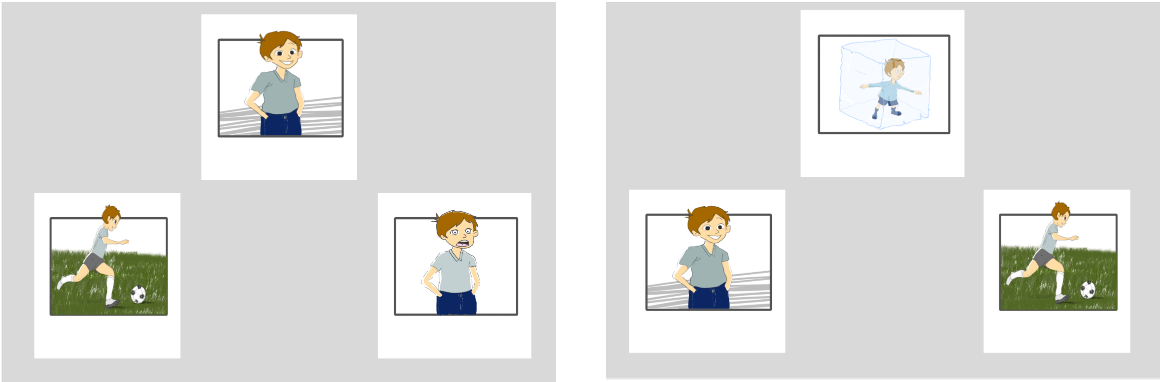

Figure 1. ***Example of the visual display that participants saw on a computer screen showing the literal and nonliteral conditions of the critical item quedarse helado ('to be shocked, lit. to get frozen'), respectively. Participants saw one of the two visual grids.***

Filler trials consisted of simple, common sentences in which the sentence referred to only one picture, such as *María anda en bici* ('Maria is riding her bike') or *Tomás está dormido* ('Tomás is asleep') or *Uxue come un helado* ('Uxue is eating an icecream'), similar in structure to experimental items. They were meant to be understood compositionally (i.e., literally). Filler sentences were included to keep children's attention and ensure that they were actively participating in the task. There were four pre-test trials very similar to filler sentences (e.g. *Andrés bebe agua* 'Andrés is drinking water'; *Hace sol* 'It is sunny').

Going back to the stimuli, the alternative images for two experimental items (the correct response in each condition is underlined) are illustrated in (2) and (3).

(2) Juan hace la cama. 'Juan makes the bed.'	(3) Pedro está hecho polvo Pedro is done dust 'Pedro is exhausted.'
a. Juan is lying on his bed. b. Juan is looking at his bed. c. Juan is building a bed (literal). d. Juan is arranging the linen (nonliteral).	a. Pedro is watching TV. b. Pedro is smiling. c. Pedro has turned into dust (literal). d. Pedro is exhausted (nonliteral).

Two counterbalanced lists were created, to make sure that participants did not see the same trial twice – once in the literal, once in the non-literal interpretation – and we assigned them randomly to one of two lists. Thus, participants did not see both target meanings in the same visual array. For each participant, order of presentation of the stimuli was randomized. E-prime 3.0 stimulus presentation software (Psychology Software Tools) was used to build and run the experiment.

The total list of critical items was split into two categories, following the distribution of performances found in the analyses of TD children (Castroviejo et al., 2024): light verb constructions on the one hand (i.e., *hacer la cama* 'to make the bed') and a continuum of metaphors, hyperboles and idioms on the other (i.e., *partirse de risa* 'to burst into laughter') (remember, LVC and MHI, respectively). The continuum included several phenomena since it seemed to be difficult to disentangle hyperboles from metaphors (see Wilson & Carston, 2007) and idioms based on metaphors from metaphors (Vega-Moreno, 2005). Each category contained an unequal number of items (five in the LVC category and nine in the MHI category: three hyperboles, three metaphors and three idioms). Accordingly, the effect of item category was taken into account in the following analyses.

Comprehension of these expressions undoubtedly depends on frequency factors and degree of exposure to these linguistic inputs. Since there is no easy access to corpora made up from transcripts of children's linguistic input and child-ambient speech in Spanish, specifically what the child incidentally learns from adult speech, we could not include those estimates for the critical items. We opted instead for a subjective measure. We ran a small norming study by means of an online questionnaire with adult native speakers of Peninsular Spanish (N= 63, ages ranging from 18 to 66 years old) to confirm that the stimuli were highly frequent in their nonliteral, dominant sense. As explained in Castroviejo et al., (2024), the results of the norming study confirmed that the expressions were highly frequent in their nonliteral sense (see a detailed presentation and discussion of the study in Castroviejo et al., 2024). On the other hand, the results reported in Castroviejo et al. show that the expressions used were familiar enough to TD children: even the youngest 3-year-olds were able to select the picture representing the nonliteral interpretation with above 60% of accuracy).

Procedure

The study was designed as a sentence-picture matching task. Participants were asked to help a friend, a cartoon character, to learn some things from the target language. First participants saw the cartoon character appearing on the computer screen together with our target sentences. The stimuli sentences were presented orthographically and auditorily twice (via a pre-recorded female voice). The reason for hearing the stimuli twice was motivated by the need to ensure that the youngest children had enough opportunities to engage with the task. We also wanted to make sure that they did not have trouble with their inattention and/or impulsivity. After listening to the stimulus, participants saw a visual display with three pictures (see

Figure 1). The position for presenting the images was counterbalanced between participants and items and order of the expressions. The participants' task was to select the visual target which matched the sentence including the expression they had previously heard, by clicking on the screen. The computer-mediated interface was especially appealing for children.

TD children were tested individually at their school in a quiet room away from their class. Autistic children either came to the laboratory to participate in the experiment, accompanied by their caregivers; or those who belonged to APNABI Autism Association/Ilargia Intervention center were individually tested at their center during one of the intervention sessions. PPVT-III and LEITER-3 tests were administered either at the laboratory or at their respective centers in a different videotaped session within a few weeks of the first session, to avoid participants becoming tired, distracted or restless. Breaks were given as often as were required. This second testing session lasted about one hour. Before testing took place, children verbally assented to participating in the experiment.

Each child was shown first four pretest trials and, if they could answer those appropriately, they were then given the complete experimental package which consisted of 32 items. The task took participants around 10 minutes to complete. Measures of accuracy and response times (RT) were collected to determine ease of processing. Reaction times (RTs, in ms) were calculated from the time the visual display appeared, just after the sentence was played for the second time, until the moment the participant tapped for a quick selection on the touch-screen display.

Analysis

Prior to the analysis, we intended to remove participants who scored less than 50% accuracy on filler items, but none was below this threshold. All statistical analyses were conducted in R (R Core Team, 2022), using the lme4 and lmerTest packages (Bates et al. 2007; Kuznetsova et al., 2017). We fitted generalized linear mixed models on accuracy (correct versus incorrect responses / rates of successful target responses) and linear mixed-effects models on response times, including the four factors, verbal mental age (VMA) as a centered continuous variable, group (ASD, TD), condition (literal, nonliteral) and category of item (LVC, MHI) together with their interactions. We coded all the variables using sum-contrast coding. The random effects structure included random intercepts by items and by participants. We intended to include a maximal random effects structure (Barr et al., 2013) but simplified it due to convergence issues. Non-convergence issues were handled by using the 'bobyqa' optimizer as well. We report relevant t-values and p-values for reaction time and z-values and p-values for accuracy.

We took two different approaches. We ran two analyses. The first one, the correct response analysis, provided us with a general overview of the abilities to identify the correct target image or response (either literal or nonliteral interpretation) compared to the non-target images. Second, the response time analysis explored the ease with which participants chose the target response and provided us with an overview of the different experimental manipulations of the study. Finally, an additional analysis involved a descriptive examination of autistic children's performance given some variability observed in the data. These analyses are presented in detail in the following sections, along with the data that emerged from them.

Results

Of the 16 initial critical items one was an exploratory item, and another one was removed due to high rates of incorrect responses across the board. The latter was disproportionately difficult for the majority of children. The final number of trials that were analyzed for each participant was 14.

For both groups, we collected participants' chronological age (in months). For the autistic group, we also gathered standardized measures, including VMA (from PPVT-III), which serve as matching variables and potential explanatory factors.

Correct Response Analyses

In terms of task performance, both groups performed above chance levels, with overall high picture selection accuracy. For critical items, the ASD group had a mean accuracy of *M* = 77.21 (*SD* = 41.98) and the TD group *M* = 83.58 (*SD* = 37.05). For filler items, the ASD group's mean accuracy was *M* = 88.83 (*SD* = 31.51), and the TD group's was *M* = 93.69 (*SD* = 24.30).

To examine the effects of group, condition, category, and verbal mental age, we fitted a generalized linear mixed model (estimated using *bobyqa* optimizer, binomial distribution, Logit link) on correct responses (accurate performance or correct choice of the target picture, i.e., choice of literal picture in the literal interpretation trials, choice of nonliteral picture in the nonliteral interpretation trials). As anticipated in the Analysis section, the model included verbal mental age as a continuous variable (centered), group, condition and type of item as fixed effects (sum-contrast coded) together with a three-way interaction between VMA, group and condition and another three-way interaction between group, condition and category. Random intercepts for participants and items were included in the model as random effects. The results of the model are summarized in Table 2. From visual inspection of the data (Figure 2), nearly all children appeared to perform above chance (chance = .50), consistently selecting the target picture with substantial accuracy.

The model revealed differences in accuracy. There is a significant interaction between VMA and CONDITION ($z = -3.02$, $p = .002$). Specifically, in the nonliteral condition, performance tends to increase as VMA increases. There is also a three-way interaction between GROUP, CONDITION and CATEGORY ($z = 2.56$, $p = .011$). We failed to find a significant difference between groups (ASD *versus* TD) across VMA for each of our conditions (The results are shown in Table 2, while the pattern of results is shown in Figure 2). The first interaction was not analyzed further because we were mainly interested in the differences between groups.

Table 2. *Logistic regression model output for correct responses.*

Term	β	95% CI	z-value	p-value
Intercept	1.93	[1.42, 2.45]	7.34	< .001
VMA (centered)	0.04	[0.03, 0.04]	7.03	< .001
GROUP	−0.40	[−0.80, 0.00]	−1.94	.052
CONDITION	−0.25	[−0.55, 0.05]	−1.61	.108
CATEGORY	−0.84	[−1.82, 0.14]	−1.69	.091
VMA × GROUP	0.01	[−0.01, 0.03]	0.57	.566
VMA × CONDITION	−0.02	[−0.04, −0.01]	−3.02	.002
GROUP × CONDITION	−0.08	[−0.69, 0.52]	−0.27	.790
GROUP × CATEGORY	0.53	[−0.02, 1.08]	1.90	.057
CONDITION × CATEGORY	−2.09	[−2.65, −1.53]	−7.34	< .001
VMA × GROUP × CONDITION	0.00	[−0.03, 0.03]	−0.02	.986
GROUP × CONDITION × CATEGORY	1.44	[0.34, 2.54]	2.56	.011

In order to establish the origin of the three-way interaction, additional multiple comparisons with Tukey contrasts were run. For the autistic group in the MHI category, no significant differences were observed between conditions. However, for the LVC category, there was a significant difference between conditions ($p = .01$) with more accurate performance in the nonliteral condition. For the control group, significant differences were observed between conditions in both item categories (i.e., MHI and LVC) ($p < .001$), with better performance in the nonliteral condition for LVC items and slightly better performance in the literal condition for MHI category of items. We also explored differences between groups for each item category and condition. No significant differences were observed except for the literal condition in the MHI category of items ($p = .001$) in which autistic participants performed below their TD verbal mental age matched-peers. Now, while for the control group there is differential performance depending on the category of the items in different conditions or interpretation trials, this pattern is partially absent in autistic participants.

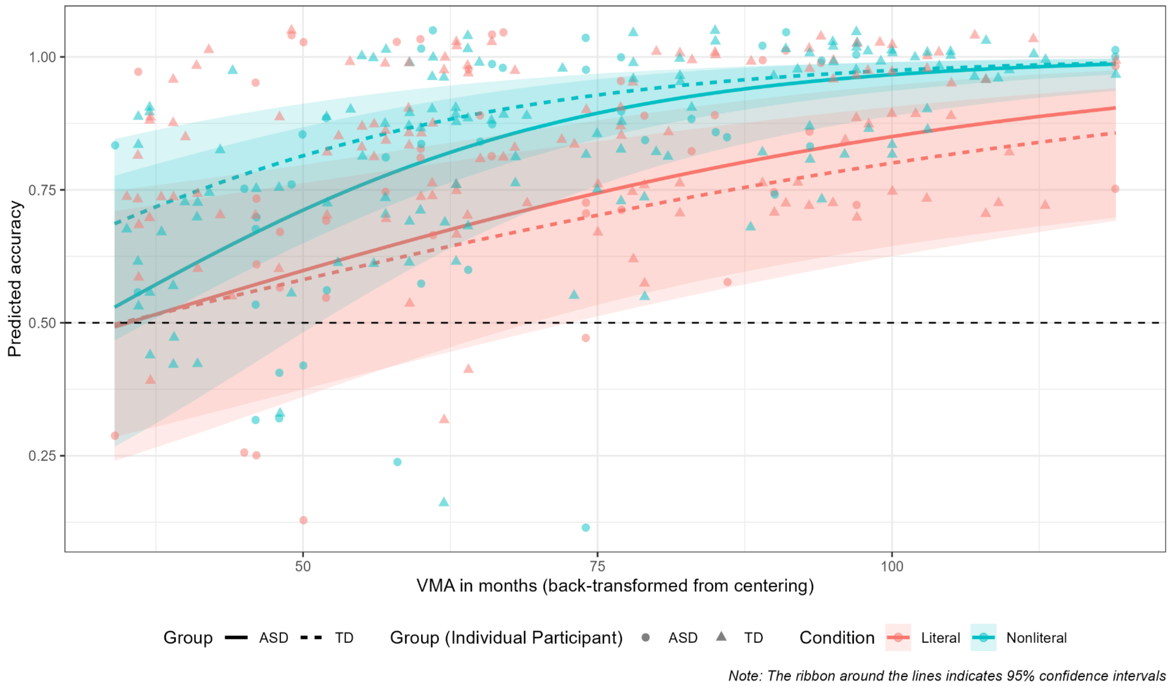

Figure 2. ***Results of correct responses with verbal mental age as a continuous predictor and a superimposed fitted accuracy slope by age. Individual dots and triangles show participant averages. Error ribbons show 95-percent confidence intervals.***

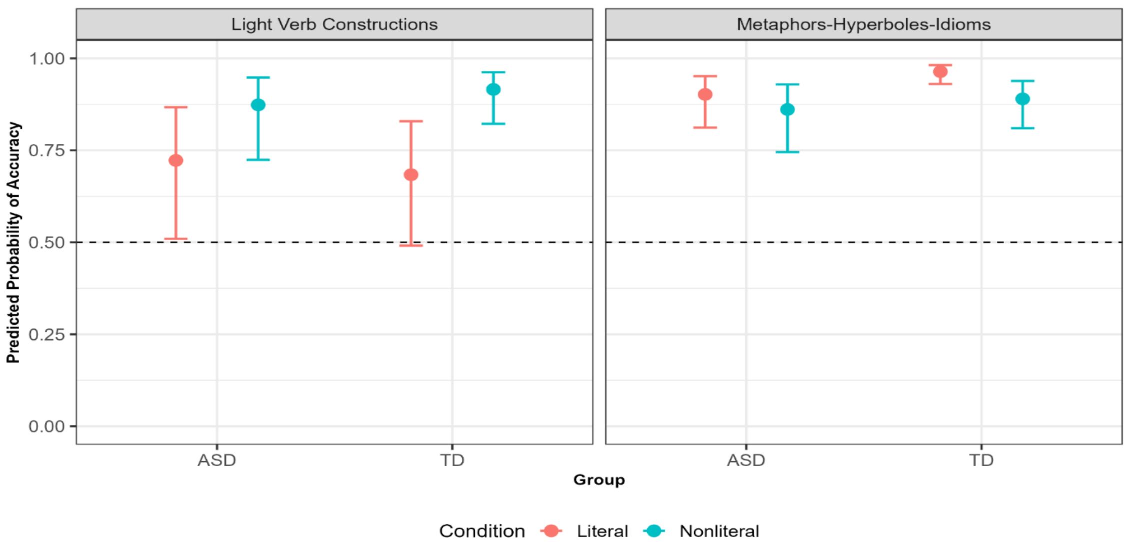

Figure 3. ***Results of correct responses per group on type of item x condition. Error bars denote one standard error of the mean.***

For autistic children, additional analyses confirmed that performance was related to verbal mental age, and it was not predicted by chronological age. Two models for accuracy were run for autistic participants only, with either chronological age or verbal mental age (centered) as continuous predictor variables (see Appendix Tables A2–A3). The model showed a main effect of VMA (z-value = 3.74, $p < .001$) as compared to no effect in the model with CA as an independent variable (z-value = 1.54, $p = .123$). CA was not a good predictor of performance in the autistic group.

Reaction Time Analyses

Excluding incorrect responses, both groups showed relatively fast response times in filler trials. The ASD group had a mean RT of $M = 2929.73$ ms ($SD = 1384.63$), while the TD group responded more quickly, with a mean RT of $M = 2413.80$ ms ($SD = 1120.50$). This intergroup difference was statistically significant, with autistic participants responding more slowly than their typically developing peers ($t(836) = 8.17$, $p < .001$). These values reflect the relative ease of target selection in filler items and establish a baseline for comparing response times in critical trials. For critical trials, the mean RT for the ASD group was $M = 3913.83$ ms ($SD = 1899.36$), compared to $M = 3127.03$ ms ($SD = 1637.42$) for the TD group. Comparing critical to filler trials provides context for understanding how each group responded to the task. Within-group comparisons showed significant differences in RTs between critical and filler trials for both groups (ASD: $t(762) = –9.15$, $p < .001$; TD: $t(2353) = –13.97$, $p < .001$), indicating that critical trials were more demanding.

To formally test effects on RTs, we fitted a linear mixed-effects model using RT (in ms) as the dependent measure. Extremely fast or slow responses were excluded. We calculated the mean RT for each item and for each participant per group and we excluded trials that exceeded 3SD above those means. 194 trials (out of 5146 observations) were excluded (3%) according to these criteria. This choice of outlier removal did not affect the overall interpretation of results. RTs were analyzed only for correct responses. This led to removing 17% of critical trials.

The model included a three-way interaction between verbal mental age (centered), group, and condition, together with a second three-way interaction between group, condition, and category, and all lower-order effects. Random intercepts for participants and items were included. The results are summarized in Table 3. The overall model revealed that there is an interaction for response times between VMA and GROUP ($t = 2.10$, $p = .037$), in which we get shorter response times as VMA increases particularly in the TD group. The control group shows consistently quicker responses as compared to the ASD group which shows slightly slower responses. Another significant interaction was found between CONDITION and CATEGORY ($t = 5.88$, $p < .001$). Multiple comparisons of means using Tukey contrasts revealed the source of the interaction. While in the LVC category of items, literal conditions are significantly slower than nonliteral conditions ($p < .001$), in the MHI category the reverse holds, nonliteral conditions take more time to get responded than literal conditions ($p < .001$). We did not find any significant differences between autistic and typically developing children regarding the conditions and categories of items, nor within each of these groups. Autistic children showed slower responses in general. Figure 5 displays overall RTs for each category of item, LVC and MHI.

Table 3. ***Regression model output for reaction latencies.***

Term	β	95% CI	t-value	df	p-value
Intercept	3,586.50	[3,346.20, 3,826.81]	29.25	37.80	< .001
VMA (centered)	−28.61	[−36.29, −20.92]	−7.30	160.21	< .001
GROUP	604.84	[275.51, 934.17]	3.60	168.56	< .001
CONDITION	90.38	[−56.73, 237.50]	1.20	1683.85	.229
CATEGORY	−200.84	[−579.50, 177.83]	−1.04	15.35	.315
VMA × GROUP	16.48	[1.12, 31.84]	2.10	160.10	.037
VMA × CONDITION	2.84	[−3.63, 9.31]	0.86	1676.11	.390
GROUP × CONDITION	135.80	[−157.62, 429.23]	0.91	1680.88	.364
GROUP × CATEGORY	−190.58	[−479.77, 98.60]	−1.29	1663.64	.197
CONDITION × CATEGORY	889.22	[592.73, 1,185.71]	5.88	1700.66	< .001
VMA × GROUP × CONDITION	−7.37	[−20.29, 5.54]	−1.12	1674.92	.263
GROUP × CONDITION × CATEGORY	−474.13	[−1,064.75, 116.49]	−1.57	1694.54	.116

Although autistic participants took slightly longer to respond, no differences were found either between groups or within each group between conditions or interpretation trials. The response pattern was similar for both groups. From the visual inspection of the data, one can conclude that the tendency is towards literal responses to take more time than nonliteral responses overall. However, once the item category is taken into account the pattern continues to be that for LVC but reverses for MHI items. The lack of condition effect in both autistic and typically developing children may very well stem from large within group variation on RTs.

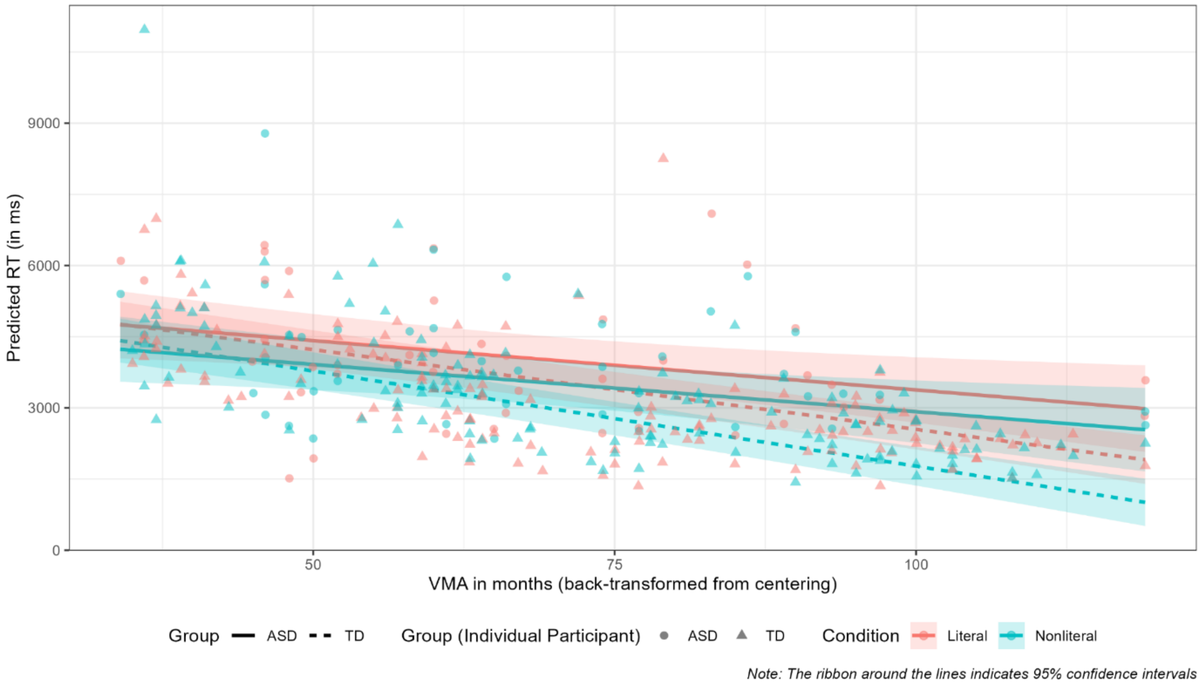

Figure 4. ***Results of response times with verbal mental age as a continuous predictor and superimposed fitted slopes. Individual dots and triangles show participant averages. Error ribbons show 95-percent confidence intervals.***

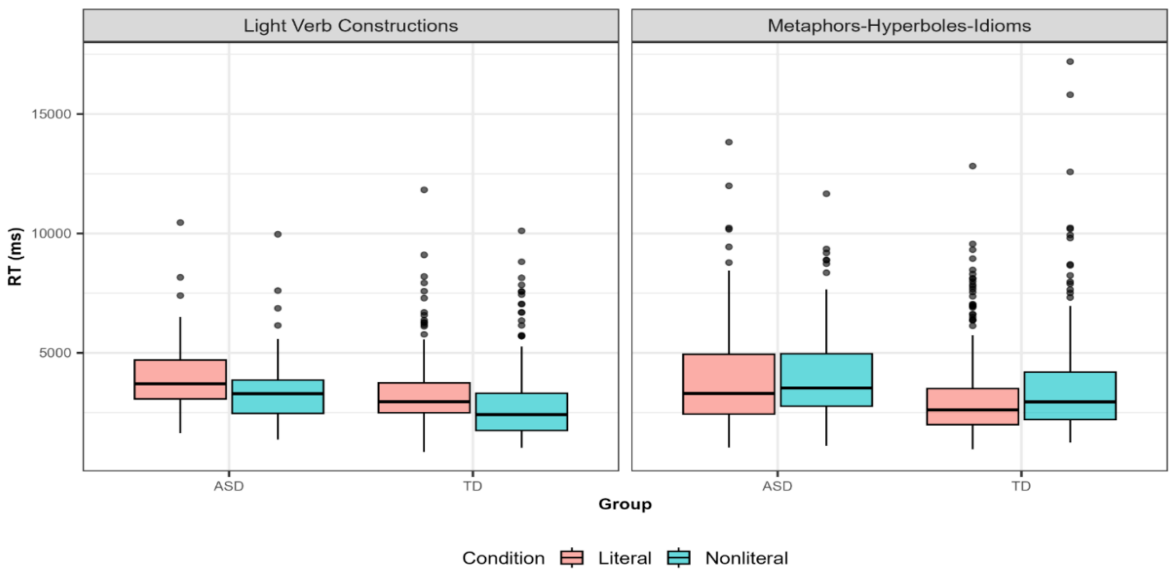

Figure 5. Turkey box plots to represent RT by group, condition and category.

We were interested in mean performance across literal and nonliteral trials broadly and within that we explored the effect of LVC and MHI specifically. To sum up these results, with increasing verbal mental age, performance changed for the best for both groups in both conditions (i.e., nonliteral and literal interpretation trials). There was an effect of the group, critically suggesting that performance on correct responses from autistic children differed from TD peers, but that only seemed to be the case in the MHI category of items, in particular, in the literal condition. No other differences were found between both groups. While the performance in different categories of items (LVC *versus* MHI) differed regarding both literal and nonliteral interpretation trials (i.e., conditions) for controls, the trend was much more subtle for autistic children for which there was a difference between interpretations for LVC, but not for MHI. The only difference we found was in the literal condition in the MHI category of items. Regarding response times, the picture that emerges is quite different. There were no differences between groups for each condition, nor within each group regarding each condition x category of item pairing. What seemed to be the case is that overall, autistic children tended to respond more slowly than their TD peers across verbal mental ages. The general trend where nonliteral responses are faster than literal responses in LVC and slower than literal responses in MHI does not hold consistently within each group. This inconsistency is likely due to significant individual variation among participants, but it could also be influenced by how participants understood or engaged with the task, as this might affect their response patterns or make certain items more challenging. Additionally, variability in familiarity with specific items or expressions could influence processing speed, potentially disrupting the trend. Factors like reduced attention or fatigue during parts of the task might also contribute to the inconsistencies in response times, although we attempted to visualize this, and it did not seem to be the case.

All the above analyses considered the 42 autistic participants as a group. Given the variable performance specifically at the youngest verbal mental ages (and the variable profiles that we included with minimally verbal children to highly verbal), we decided to explore individual differences in the autistic group descriptively, focusing on qualitative observations (see Figure 6), following Panzeri et al. (2022), Saban-Bezalel et al. (2019) and many others on individual variation of performance. Group mean performance might have been masking a wide range of individual variation on performance, limiting interpretability, so a closer inspection of the data seemed warranted.

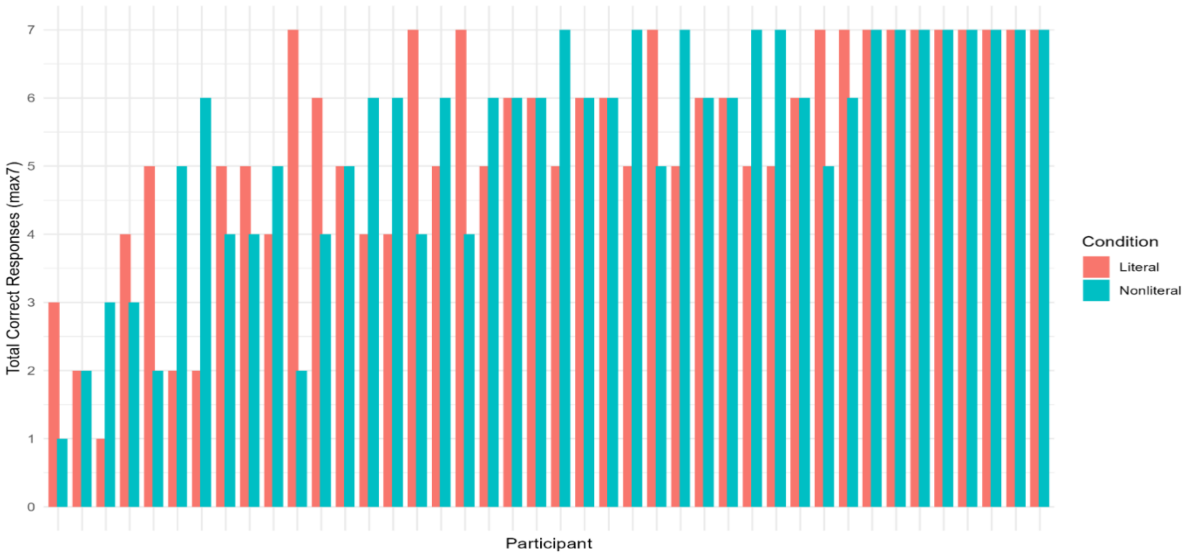

Figure 6. Individual autistic participants' scores (sum of correct responses) in literal and nonliteral interpretation trials.

Descriptive and Exploratory Post-hoc Analyses of the Autistic Group

Leaving reaction latencies aside, in this final analysis, we focused on performance on individual participants, descriptively. The idea was to simply describe profiles of autistic children regarding their response performance. We looked at their performance across conditions and types of items and described it in terms of z-scores (i.e., above or below the group's mean). We also wanted to explore whether their performance was related to child individual characteristics (i.e., CA, VMA, the difference between these two variables, and non-verbal IQ). In some cases, we could also add autism severity scores to these variables. By *autism severity,* we refer to the ADOS calibrated severity score, a metric derived from the transformation of ADOS-2 raw scores, as proposed by Gotham, Pickles and Lord (2009), with a range from 1 to 10. However, as we did not have ADOS scores for all children, we could not use autism severity as a variable for every participant. We mention autism severity scores only in the case of subgroups for which we had more than half of such scores. Performance was far more variable than that of TDs, and some fairly exploratory patterns were found.

There were four children who overall performed low across conditions and item categories, showing a very small proportion of correct responses (LVC *lit.* -0.9*z and nonlit. -0.7z; MHI lit. -1z and nonlit. -1z*). Their profile was heterogeneous: NVIQ ranging from 73 to 113; VMA from 3;10 years to 7;6 years; difference between CA and their VMA from 6 months to 2;9 years; and autism severity from 5 to 7.

By contrast, there were nine autistic children who had slightly higher scores, above the group's mean (LVC *lit.* +0.7*z* and *nonlit.* +0.5*z;* MHI *lit.* +0.4*z* and *nonlit.* +0.5*z*). They performed *at ceiling* irrespective of condition or item category. The high scorers group formed a heterogeneous group with regard to NVIQ (ranging from 82 to 117), VMA (4;2 years to 9;11 years), difference between CA and their VMA (for those who got higher VMA: 1 month to 1;4 years, for those who got lower VMA: 1 month to 3;11 years).

Interestingly, there were four children who performed better in the literal interpretation trials than on the nonliteral ones across item categories (*lit.* LVC M = 100, SD = 0 *vs nonlit*. LVC M= 42.85, SD= 53.45; and *lit.* MHI M=100, SD = 0 *vs nonlit*. MHI M=57.14, SD=51.35; lit. +0.6*z* and *nonlit.* -0.6*z*). They seemed to be the only ones exhibiting a strong literalism (including LVC expressions). These four children formed again a heterogeneous group with respect to NVIQ (ranging from 83 to 106), VMA (3 years to 5;4 years), difference between CA and their VMA (from 7 months to 2;6 years), and autism severity (5 to 7, though one severity score was missing).

Discussion

The aim of this paper is to add to the literature on nonliteral language development in autistic children. In particular, we aimed to test whether autistic children go through a phase of strong literalism, such that they would understand highly frequent nonliteral uses of language in a word-by-word, compositional, way. For that purpose, we used a picture selection task where children were required to choose between representations of either the literal or nonliteral meaning of an expression, along with two distractors. This paradigm was chosen primarily because it allowed us to compare the accuracy and ease of access to literal *versus* the accuracy and ease of access to nonliteral meanings in both groups, rather than to assess children's preference for one interpretation over the other. The expressions we used as stimuli were not just pieces of conventional nonliteral language, but their nonliteral uses were conventional *and* more frequent than their word-by-word literal counterparts. The expressions included frequent idioms, hyperboles and metaphors, as well as light verb constructions.

The primary aim of this study was to compare literal and nonliteral interpretation trials to identify potential differences between groups and also within each group. To assess the potential impact of VMA, VMA was included as a linear predictor for both autistic and typically developing participants. Additionally, item categories (LVC *versus* MHI) were included for exploratory analysis, based on the results obtained in Castroviejo et al. (2024). The study aimed to answer the following two research questions. The first one was:

RQ1: Are there any differences in the performance of TD and autistic children when interpreting nonliteral expressions across verbal mental age?

Our hypothesis was that autistic children's performance on nonliteral uses in general would be lower compared to their TD peers, even when matched on VMA. We were aware that some empirical evidence tends to support the opposite hypothesis (see Kalandadze, 2019), as well as some theories (Andrés-Roqueta & Katsos, 2017) that hold that comprehension of the nonliteral uses of the expressions in the current experiment would relate to VMA. However, based on some contrary evidence supporting literalism, and the fact that literalism is an interpretive tendency for many autistic people, especially in their infancy and youth, we hypothesized that we would find group differences, expressed either in difficulties in accessing nonliteral meanings, or in a relative ease in accessing the literal ones (or both at the same time). We did not have any specific predictions for autistic participant's performance in item categories.

With regard to RQ1, and contrary to what was predicted, results suggest that autistic children did not differ from VMA-matched TD children's performance regarding literal and nonliteral interpretation trials across development. The above-chance performance in the task suggests that the task itself was accessible and appropriate for assessing comprehension in both groups, indicating that autistic children are capable of processing both types of interpretations comparably to their TD peers when matched by VMA. In spite of those high scores, children from both groups exhibited an upwards linear performance as their vocabulary size developed.

In the same line, we observed that VMA (receptive vocabulary) was a predictor of success on both literal and nonliteral conditions in the autistic group. As said, since Norbury (2005), several authors have argued that language abilities (particularly lexical knowledge in Norbury's own case), play a key role in figurative language comprehension (e.g., Andrés-Roqueta & Katsos, 2020). Some authors have found that the hypothesis that figurative language comprehension relates to language level applies particularly well to *conventional* figurative language (Kasirer & Mashal, 2014; Olofson et al., 2014), which is the kind of nonliteral language that we were concerned with in our study: all the expressions in our list of stimuli were highly conventional in their nonliteral use. In this regard, as we did not compare conventional with novel nonliteral uses of language, we cannot add to the discussion concerning the role of linguistic development (or, rather, of receptive vocabulary levels) to these different kinds of nonliteral language. However, in line with Norbury (2005), Kasirer & Marshal (2014), and Olofson et al. (2014), we found a clear effect of VMA on overall task performance, suggesting that vocabulary breadth may facilitate the comprehension of both literal and nonliteral language in autistic children. As said, we had not predicted that nonliteral language would be so clearly predicted by VMA, because we thought that autistic children's difficulties with nonliteral language could be independent from their linguistic development. As we explain below, we found some evidence that this could be the case only for a small subgroup of children whose degree of literalism did not correspond to their vocabulary level.

We also explored differences between and within each group regarding item categories. We found that autistic participants only differed significantly from TD children in the MHI item category during literal interpretation trials, showing slightly lower performance. This was a surprising finding, which could suggest more difficulties in accessing literal interpretations when there is competition with nonliteral interpretations. On the other hand, we observed no significant differences between literal and nonliteral interpretation trials within the MHI item category for autistic participants, which would mean that it only makes sense to speak about some difficulties in autistic children when we compare autistic and non-autistic performance. However, we suspect that the observed difference is due to the heterogeneity of the autistic group: as we mention below, a majority of autistic children exhibited an interpretive flexibility comparable to their TD peers.

As already shown in the previous study on typical developmental trajectories (Castroviejo et al., 2024), TD controls exhibited higher accuracy scores in the nonliteral condition (as compared to the literal condition) for LVCs across developmental stages. The opposite pattern emerged for MHI items, in which case the literal condition got higher correct responses in comparison to the nonliteral condition across ages. Crucially, the difference between both types of items for TD controls lay in the literal interpretation trials. Autistic participants in our sample also followed that performance pattern except for the literal interpretation trials in the MHI item category. Underlyingly, their overall performance pattern seemed to be similar to the bimodal performance of TD children.

Concerning response times, autistic participants were overall slower to respond than their TD peers in both conditions as well as in the filler trials. There were apparent differences between groups across verbal mental ages. With increasing age, TD children's responses became faster, possibly meaning that the task became easier for them. That seems not to be the case for autistic children. The task might have still been somewhat demanding or effortful, and attention and/or motor skills may have also played a role. No differences were found between literal and nonliteral interpretation trials between groups and within each group, nor for each item category. What we found is that once we collapsed groups, the emerging pattern converges with the results from the picture selection task.

Our second research question was the following:

RQ2: Do autistic children exhibit an easier access to literal, compositional meanings at different developmental stages?

Despite being implicitly addressed in the preceding research question, autistic children at group level were not more accurate in literal interpretation trials than in nonliteral interpretation trials. Nor did they perform better in literal interpretation trials at any point across VMA development. Most importantly, we did not find differences across interpretation trials in response times for autistic children. Here we found differences in picture selection between interpretation trials in the LVC item category, with more accurate performance in nonliteral interpretations and no differences between conditions in the MHI item category. However, these empirical findings might underlyingly suggest bimodal performance as in typical development. This would most likely be the case if there were not so much within-group variation. In other words, there might be a lingering tendency toward better performance in literal interpretation trials for MHI and in nonliteral ones for LVC.

One reason why we may not have found support for a general pattern of literalism may be the lack of ecological validity. This forced-choice paradigm, although widely used, falls short of accurately representing interpretation in naturalistic settings. Still, even with a relatively modest sample size and a minimally demanding task with visuals as supportive contexts, we found some subgrouping, in particular a few children who did better in literal interpretation trials, therefore exhibiting literalism.

Although 25 autistic children exhibited a behavior similar to that of TDs of similar vocabulary size, we observed that a small number of children (four) displayed a lower performance in all items and conditions, that another group of nine children performed *at ceiling* in all items and conditions (LVC 100% in both conditions; MHI *lit.* 100% and MHI *nonlit.* 98%), and that a group formed by another four participants adjusted to a strong literalist profile. Concerning the first group of four children, whose performance was lower than the rest in the literal and nonliteral conditions, our hypothesis is that their structural language abilities did not correspond to their vocabulary size: that is, despite being VMA-matched to TDs on receptive vocabulary, their sentence comprehension abilities were probably not matched to that of their TD peers.

The group that was more interesting for the purposes of this paper was the group of four children who were more accurate than their VMA-matched TD peers only in the literal condition in both categories (LVC ASD 100% vs TD 60% and MHI ASD 100% vs TD 92%), and who performed significantly less accurately in the nonliteral condition (ASD; LVC 43% and MHI 57%; TD; LVC 79% and MHI 77%). Children in this group were more accurate on literal interpretation trials than on nonliteral interpretation trials, that is, they seemed to have obvious difficulties understanding expressions in a non-compositional way. It is particularly noteworthy that they were 100% accurate on the literal interpretation of expressions like *hacer la cama* ('to make the bed') and below 50% on their nonliteral interpretation. As mentioned above, these four children, on the other hand, formed a heterogeneous group with respect to NVIQ (ranging from 83 to 106), VMA (3 years to 5;4 years), the difference between CA and their VMA (from 7 months to 2;6 years), and autism severity (5 to 7, though one severity score was missing). This suggests that literalism may be more acute in some autistic individuals than in others, and that it may relate to factors we did not explore. Perhaps literalism is characteristic of autism in the sense that only autistic individuals exhibit such a way of understanding nonliteral language, while at the same time not being widespread in the autistic population. In other words, it is not observed in all autistic individuals, and as such cannot be considered a defining feature, but it can be considered a cue validity feature (Rosch, 1978).

The paradigm that we employed had some limitations, so the current findings should be interpreted with caution. First, the task was simple and easy, as we already mentioned in several places and as accuracy results showed, and children were given the right interpretation as an option, which is not what occurs in real life contexts. In general, multiple-choice paradigms significantly reduce the uncertainty surrounding the comprehension of any utterance of an expression used nonliterally. Second, an obvious limitation of the current study is the relatively small size of our autistic sample which naturally impacts on the exploratory analysis of possible subgroups. The relatively low number of autistic participants and consequently low statistical power for detecting true effects was in large part due to the challenges of recruiting participants. Accordingly, findings for subgroups should be seen as an exploration of possible interpretive performance patterns coming from individual variation. Increasing the autistic sample size would allow for stabler developmental trajectories and possibly more homogeneous clusters. A further constraint of child-friendly paradigms is the reduced number of critical items used with the addition of an unbalanced number of item types in each item category.

Two limitations of our study relate to the notion of VMA. The first limitation is that our VMA measures are of receptive vocabulary, while the task consisted in understanding sentences. Especially in the case of autistic children, receptive vocabulary may not provide an adequate measure of linguistic development, and, in fact, we suspect that the poor performance in both conditions of a subgroup of autistic children may relate to the fact that their linguistic abilities did not correspond to the VMA measure. The second limitation is assuming that TD children's receptive vocabulary is aligned with their chronological age, without verifying this through direct testing. We did not test receptive vocabulary in the TD participants because they constituted the experimental group of a TD developmental study in which VMA was not a variable. This implied that we had to assume that chronological age would match verbal mental age in the TD group.

Lastly, in order to try to explain why certain subgroups performed the way they did, we realized that we had limited information about the participants in the study. As mentioned above, we saw nothing in common between the children who exhibited a stronger literalist behavior, for instance. In that respect, we missed having data about ToM abilities, executive function, local processing or rigid behaviors, all of which may relate to literalism. However, since we had ADOS-2 (Lord et al., 2012) scores for more than half of the participants, we were able to observe that subgroups were also heterogeneous with respect to the social-affective scores of the ADOS-2. These scores can be considered an indirect measure of ToM abilities, further highlighting the complexity of the subgroup differences.

Conclusion

The current study contributes to research studying nonliteral language in providing empirical evidence on how different expressions in Spanish are interpreted and processed by both autistic and typically developing pre-school and early elementary school-aged children across a wide verbal mental age range (three to nine). In particular, this study compared autistic children to their TD peers on their comprehension of expressions that had a dominant nonliteral meaning. We wanted to see whether autistic children could exhibit a strong form of literalism (word-by-word processing) with that kind of expressions.

Overall, our results revealed that autistic children performed similarly to typically developing peers. Their verbal mental age trajectories did not differ between the groups either, regardless of the literal or nonliteral interpretation trials. Finally, their correct response performance in both groups increased with verbal mental age. Contrary to expectations, difficulties with nonliteral uses of language were not spotted across the autistic spectrum. However, typically developing children performed differently in two types of nonliteral expressions: they were more accurate with respect to literal conditions for Metaphors, Hyperboles and Idioms and with respect to the nonliteral conditions for Light Verb Constructions. This performance pattern was not as clearly observed in the autistic group, as no differences were found between literal and nonliteral interpretation trials for Metaphors, Hyperboles and Idioms, although we hypothesize that the lack of a similar dichotomous pattern is masked by the variability in individuals' performances.

In trying to account for such variable performance, we observed that a small group of children were strongly literalists, performing better in literal interpretation trials even in examples involving light verb constructions. At the same time, there were some very good performers who could flexibly entertain both interpretations. This suggests that there is a lot of variability concerning literalism across the spectrum. However, given the characteristics of our design, we cannot exclude that the literalism observed in many autistic individuals may appear in more demanding communicative contexts. Future work should explore further this possible heterogeneity in pragmatic profiles within the spectrum and relate eventual subtypings to other variables or traits.

References

Allott, N., & Textor, M. (2022). Literal and metaphorical meaning: in search of a lost distinction. *Inquiry*, 1–28. https://doi.org/10.1080/0020174X.2022.2128867

Andrés-Roqueta, C., & Katsos, N. (2017). The Contribution of Grammar, Vocabulary and Theory of Mind in Pragmatic Language Competence in Children with Autistic Spectrum Disorders. *Frontiers in Psychology, 8*, 996. https://doi.org/10.3389/fpsyg.2017.00996

Asperger, H. (1991). 'Autistic psychopathy' in childhood. In U. Frith (Ed.), *Autism and Asperger Syndrome* (pp. 37–92). Cambridge University Press. https://doi.org/10.1017/CBO9780511526770.002

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* (Version 2) [R package]. https://cran.r-project.org/web/packages/lme4/

Bowdle, B. F., & Gentner, D. (2005). The Career of Metaphor. *Psychological Review, 112*(1), 193–216. https://doi.org/10.1037/0033-295X.112.1.193

Castroviejo, E., Ponciano, M., Hernández-Conde, J.V. and Vicente, A. (2024), Development of nonliteral interpretations in typically developing Spanish speaking children: light verb constructions and figurative expressions. *Studia Linguistica, 78*, 8-34. https://doi.org/10.1111/stul.12222

Chahboun, S., Vulchanov, V., Saldaña, D., Eshuis, H., & Vulchanova, M. (2016). Can You Play with Fire and Not Hurt Yourself? A Comparative Study in Figurative Language Comprehension between Individuals with and without Autism Spectrum Disorder. *PLoS One, 11*(12), e0168571. https://doi.org/10.1371/journal.pone.0168571

Chahboun, S., Vulchanov, V., Saldaña, D., Eshuis, H., & Vulchanova, M. (2017). Can you tell it by the prime? A study of metaphorical priming in high-functioning autism in comparison with matched controls. *International Journal of Language & Communication Disorders, 52*(6), 766-785. https://doi.org/10.1111/1460-6984.12314

Chevallier, C., Wilson, D., Happé, F., & Noveck, I. (2010). Scalar Inferences in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 40*(9), 1104–1117. https://doi.org/10.1007/s10803-010-0960-8

Deliens, G., Papastamou, F., Ruytenbeek, N., Geelhand, P., & Kissine, M. (2018). Selective Pragmatic Impairment in Autism Spectrum Disorder: Indirect Requests Versus Irony. *Journal of Autism and Developmental Disorders, 48*(9), 2938–2952. https://doi.org/10.1007/s10803-018-3561-6

Dunn, L. M., Dunn, L. M., & Arribas, D. (2006). *PPVT-III Peabody: Test de vocabulario en imágenes.* Tea Ediciones.

Fellbaum, C., Geyken, A., Herold, A., Koerner, F., & Neumann, G. (2006). Corpus-based Studies of German Idioms and Light Verbs. *International Journal of Lexicography, 19*(4), 349–360. https://doi.org/10.1093/ijl/ecl031

Fleischer, W. (1997). *Phraseologie der deutschen Gegenwartssprache*: De Gruyter. https://doi.org/10.1515/9783110947625

Gernsbacher, M. A., & Pripas-Kapit, S. R. (2012). Who's missing the point? A commentary on claims that autistic persons have a specific deficit in figurative language comprehension. *Metaphor and Symbol, 27*(1), 93–105.

Geurts, B., Kissine, M., & Tiel, B. van. (2019). Pragmatic reasoning in autism. In K. Morsanyi & R. M. J. Byrne (Eds.), *Thinking, Reasoning, and Decision Making in Autism* (1st ed., pp. 113–134). Routledge. https://doi.org/10.4324/9781351060912-6

Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition, 48*(2), 101–119. https://doi.org/10.1016/0010-0277(93)90026-R

He, A. X., & Wittenberg, E. (2020). The acquisition of event nominals and light verb constructions. *Language and Linguistics Compass, 14*(2), e12363. https://doi.org/10.1111/lnc3.12363

Hochstein, L., Bale, A., & Barner, D. (2018). Scalar Implicature in Absence of Epistemic Reasoning? The Case of Autism Spectrum Disorder. *Language Learning and Development, 14*(3), 224–240. https://doi.org/10.1080/15475441.2017.1343670

Kasirer, A., & Mashal, N. (2014). Verbal creativity in autism: Comprehension and generation of metaphoric language in high-functioning autism spectrum disorder and typical development. *Frontiers in Human Neuroscience, 8,* 615.

Kasirer, A., & Mashal, N. (2016). Comprehension and generation of metaphors by children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 32,* 53–63.

Kalandadze, T., Bambini, V., & Næss, K.-A. B. (2019). A systematic review and meta-analysis of studies on metaphor comprehension in individuals with autism spectrum disorder: Do task properties matter? *Applied Psycholinguistics, 40*(6), 1421–1454. https://doi.org/10.1017/S0142716419000328

Kalandadze, T., Norbury, C., Nærland, T., & Næss, K.-A. B. (2018). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism: The International Journal of Research and Practice, 22*(2), 99–117. https://doi.org/10.1177/1362361316668652

Kanner, L. (1944). Early infantile autism. *The Journal of Pediatrics, 25*, 211–217. https://doi.org/10.1016/S0022-3476(44)80156-1

Kasirer, A., & Mashal, N. (2016). Comprehension and generation of metaphors by children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 32,* 53–63. https://doi.org/10.1016/j.rasd.2016.08.003

Kissine, M., Cano Chervel, J., Carlier, S., De Brabanter, P., Ducenne, L., Pairon, M.-C., Deconinck, N., Delvenne, V., & Leybaert, J. (2015). Children with autism understand indirect speech acts: Evidence from a semi-structured act-out task. *PloS One, 10,* e0142191. https://doi.org/10.1371/journal.pone.0142191

Köder, F., & Falkum, I. L. (2020). Children's metonymy comprehension: Evidence from eye-tracking and picture selection. *Journal of Pragmatics, 156,* 191–205. https://doi.org/10.1016/j.pragma.2019.07.007

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *lmerTest: Tests in linear mixed effects models* (Version 2.0-7) [R package]. https://cran.r-project.org/package=lmerTest

Lampri, S., Peristeri, E., Marinis, T., & Andreou, M. (2023). Figurative language processing in autism spectrum disorders: A review. *Autism Research,* 1–16. https://doi.org/10.1002/aur.3069

Lord C., Luyster R., Gotham K., & Guthrie W. (2012). *Autism Diagnostic Observation Schedule.* 2nd. Torrence, CA: Western Psychological Services. (ADOS-2) *Manual (Part II): Toddler Module.*

MacKay, G., & Shaw, A. (2004). A comparative study of figurative language in children with autistic spectrum disorders. *Child Language Teaching and Therapy, 20*(1), 13–32. https://doi.org/10.1191/0265659004ct261oa

Marocchini, E., Di Paola, S., Mazzaggio, G., & Domaneschi, F. (2022). Understanding indirect requests for information in high-functioning autism. *Cognitive Processing, 23*(1), 129–153. https://doi.org/10.1007/s10339-021-01056-z

Martín-González, I., Schroeder,K., Castroviejo, E., Lossius Falkum, I., & Vicente, A. (2025). Novel metaphor processing in young autistic children. *Language Development Research, 5*(2), 28-66. https://doi.org/10.34842/im-g

Mazzaggio, G., Foppolo, F., Job, R., & Surian, L. (2021). Ad-hoc and scalar implicatures in children with autism spectrum disorder. *Journal of Communication Disorders, 90*, 106089. https://doi.org/10.1016/j.jcomdis.2021.106089

McMahon C.M., Lerner M.D., & Britton N. (2013) Group-based social skills interventions for adolescents with higher-functioning autism spectrum disorder: a review and looking to the future, *Adolescent Health, Medicine and Therapeutics, 4,* 23-38, https://doi.org/10.2147/AHMT.S25402

Melogno, S., Pinto, M., & Di Filippo, G. (2017). Sensory and Physicopsychological metaphor comprehension in children with ASD: A preliminary study on the outcomes of a treatment. *Brain Sciences, 7*(12), 85. https://doi.org/10.3390/brainsci7070085

Morra, L. (2016). Raising awareness of how Asperger persons perceive their capacity to use Metaphors. *Medicina e Storia, 11*(2), 129–146.

Norbury, C. F. (2005). The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology,* I(3), 383–399. https://doi.org/10.1348/026151005X26732

Nunberg, G., Sag, I. A., & Wasow, T. (1994). Idioms. *Language, 70*(3), 491–538. https://doi.org/10.2307/416483

O'Shea, A., Cersosimo, R. & Engelhardt, P.E.(2024). Online Metaphor Comprehension in Adults with Autism Spectrum Disorders: An Eye Tracking Study. J *Journal of Autism and Developmental Disorders,* 1-15. https://doi.org/10.1007/s10803-024-06562-5

Olofson E. L., Casey D., Oluyedun O., van Herwegen J., Beccera A., Rundblad G. (2014). Youth with autism spectrum disorder comprehend lexicalized and novel primary conceptual metaphors. *Journal of Autism and Developmental Disorders, 44*(10), 2568–2583. http://doi.org/10.1007/s10803-014-2129-3

Ozonoff, S., & Miller, J. N. (1996). An exploration of right-hemisphere contributions to the pragmatic impairments of autism. *Brain and Language, 52*(3), 411–434. https://doi.org/10.1006/brln.1996.0022

Panzeri, F., Mazzaggio, G., Giustolisi B, Silleresi, S., & Surian, L. (2022). The atypical pattern of irony comprehension in autistic children. *Applied Psycholinguistics, 43(*4), 757-784. https://doi.org/10.1017/S0142716422000091

Pastor-Cerezuela, G., Fernández-Andrés, M.-I., Tordera-Yllescas, J. C., & González-Sala, F. (2020). Metaphor comprehension in children with and without autism spectrum disorder. *Research in Autism Spectrum Disorders*, *76*, 101588. https://doi.org/10.1016/j.rasd.2020.101588

Paul, R., & Cohen, D. J. (1985). Comprehension of Indirect Requests in Adults with Autistic Disorders and Mental Retardation. *Journal of Speech, Language, and Hearing Research*, *28*(4), 475–479. https://doi.org/10.1044/jshr.2804.475

Pijnacker, J., Hagoort, P., Buitelaar, J., Teunisse, J.-P., & Geurts, B. (2009). Pragmatic Inferences in High-Functioning Adults with Autism and Asperger Syndrome. *Journal of Autism and Developmental Disorders, 39*(4), 607–618. https://doi.org/10.1007/s10803-008-0661-8

Psychology Software Tools. (2016). *E-Prime 3.0* (Version 3.0) [Computer software]. https://support.pstnet.com/

R Core Team (2022). *R: A language and environment for statistical computing.* Vienna, 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.* Stoelting Company.

Rosch, Eleanor (1978), Principles of categorization, in Eleanor Rosch & Barbara B. Lloyd (Eds.), *Cognition and Categorization*, Hillsdale, New Jersey: Lawrence Erlbaum, pp. 27–48.

Rundblad G., & Annaz D. (2010). The atypical development of metaphor and metonymy comprehension in children with autism. *Autism, 1,* 29-46. https://doi.org/10.1177/1362361309340667

Rutter, M., LeCouteur, A. & Lord, C. (2003). *Autism Diagnostic Interview - Revised*. Los Angeles, CA: Western Psychological Services. Adaptación española 2006.

Saban-Bezalel, R., Dolfin, D., Laor, N., & Mashal, N. (2019). Irony comprehension and mentalizing ability in children with and without Autism Spectrum Disorder. *Research in Autism Spectrum Disorders, 58,* 30–38. https://doi.org/10.1016/j.rasd.2018.11.006

Schaeken, W., Van Haeren, M., & Bambini, V. (2018). The understanding of scalar implicatures in children with Autism Spectrum Disorder: Dichotomized responses to violations of informativeness. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01266

Su, Y., & Su, L.-Y. (2015). Interpretation of logical words in mandarin-speaking children with autism spectrum disorders: Uncovering knowledge of semantics and pragmatics. *Journal of Autism and Developmental Disorders, 45.* https://doi.org/10.1007/s10803-014-2350-0

Taylor, J. R. (2006). Polysemy and the lexicon. In K. G., M. Achard, R. Dirven, & F. J. Ruiz de Mendoza Ibánez (Eds.), *Cognitive Linguistics: Current Applications and Future Perspectives* (pp. 51-80). Berlin: Mouton de Gruyter.

Van Herwegen, J., & Rundblad, G. (2018). A cross-sectional and longitudinal study of novel metaphor and metonymy comprehension in children, adolescents, and adults with autism spectrum disorder. *Frontiers in Psychology*, 9, 945. https://doi.org/10.3389/fpsyg.2018.00945

Van Tiel, B., & Kissine, M. (2018). Quantity-based reasoning in the broader autism phenotype: A web-based study. *Applied Psycholinguistics, 39*(6), 1373–1403. https://doi.org/10.1017/S014271641800036X

Vega-Moreno, R. E. (2005). Idioms, transparency and pragmatic inference. *UCL Working Papers in Linguistics, 17,* 389–425.

Vicente, A., Ábalos, Z., Martín-González, I., Ramos-Cabo, S., Castroviejo, E. (2024) Autism. In Vandenbroucke, M., Declercq, J., Brisard, F. & D'hond, S. (Eds.), *Handbook of Pragmatics* (pp. 1-40). John Benjamins. https://doi.org/10.1075/hop.27.aut4

Vicente, A., & Falkum, I. L. (2023). Accounting for the preference for literal meanings in autism spectrum conditions. *Mind & Language, 38,* 119–40. https://doi.org/10.1111/mila.12371

Vulchanova, M., Talcott, J. B., Vulchanov, V., Stankova, M., & Eshuis, H. (2012). Morphology in autism spectrum disorders: Local processing bias and language. *Cognitive Neuropsychology, 29*(7–8), 584–600. https://doi.org/10.1080/02643294.2012.762350

Vulchanova, M., & Vulchanov, V. (2022). Rethinking figurative language in autism: What evidence can we use for interventions? *Frontiers in Communication, 7.* https://doi.org/10.3389/fcomm. 2022.910850

Walenski, M., & Love, T. (2017). The real-time comprehension of idioms by typical children, children with specific language impairment and children with autism. *Journal of Speech Pathology & Therapy*, 3(1), 130. https://doi.org/10.4172/2472-5005.1000130

Wilson, D., & Carston, R. (2007). A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In N. Burton-Roberts (Ed.), *Pragmatics* (pp. 230-259). Palgrave.

Wittenberg, E., (2016). With light verb constructions from syntax to concepts. PhD Dissertation. Universität Potsdam. *Potsdam Cognitive Science Series.*

Data, Code and Materials Availability Statement

List of materials, data and analysis script are available on the open science framework online data OSF repository at: https://osf.io/v3kq9/

Ethics Statement

Ethical approval was issued by University of the Basque Country's (UPV/EHU) Ethics Committee for research with human beings (CEISH), code M10_2019_205. Written informed consent was obtained from parents/caregivers of all participants.

Authorship and Contributorship Statement

MP: Conceptualization, Software, Investigation, Data curation, Formal analysis, Writing - Original Draft Preparation, Writing - Review & Editing. AV: Conceptualization, Methodology, Resources, Writing - Original Draft Preparation, Writing - Review & Editing, Supervision. JVHC: Formal analysis. EC:

Conceptualization, Methodology, Resources, Writing - Original Draft Preparation, Writing - Review & Editing, Supervision. All authors approved the final version of the manuscript and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

The authors wish to thank all the children and their families for their participation in this study. Special thanks are also extended to: Early Attention Service and Autism Association from Araba; APNABI Association from Bizkaia; CEIP Gabriel Celaya HLHI Eskola Publikoa and CPI Antonio Trueba IPI Eskola Publikoa in Portugalete; C.E.I.P. Itsasoko Ama Ikastetxea H.L.H.I. in Santurtzi (Spain). We are also grateful to our colleagues at the Lindy Lab (UPV/EHU), as well as to the editors of LDR and the three anonymous reviewers who commented on the original version of this paper. They have significantly helped us improve our work.

Earlier versions of this research were presented at the 31st Colloquium on Generative Grammar (CGG31, 2022); the 3rd Joint Meeting of the Society for Philosophy and Psychology (SPP) and the European Society for Philosophy and Psychology (ESPP) (SPP/ESPP, 2022); and the 9th Experimental Pragmatics Conference (XPRAG, 2022). We are grateful to the audiences at these events for their insightful feedback.

This research has been partially supported by grant PID2021-122233OB-I00, funded by MICIU/AEI/10.13039/501100011033 and by "ERDF/EU", and by the IT1537-22 Research Group (Basque Government). This work has also been supported by the Basque Government's Department of Education through its Predoctoral Training Programme for Non-Doctoral Research Staff (Programa Predoctoral de Formación de Personal Investigador No Doctor, 2020–2024) (M.P.).

Appendix

Table A1. ***Sentences used as experimental items. The nonliteral translations of the items are also shown.***

Item	Literal interpretation	Nonliteral translation / interpretation
Benito hace pesas	Benito makes (builds) weights.	Benito lifts weights.
El técnico corta el agua	The technician cuts the water.	The technician turns off the water.
Ibai se queda helado al oír la noticia	Ibai gets frozen when hearing the news.	Ibai is shocked by the news.
Idoia le da la mano a Martín	Idoia gives Martín her hand.	Idoia shakes Martín's hand.
Ione coge el autobús	Ione catches (lifts) the bus.	Ione takes the bus.
Juan hace la cama	Juan makes (builds) the bed.	Juan makes the bed.
Juan le pone mala a Elena	Juan makes Elena sick.	Juan drives Elena mad.
Pedro está hecho polvo	Pedro is made of dust.	Pedro is exhausted.
Sandra duerme con los angelitos	Sandra sleeps with the angels (literally).	Sandra sleeps with the angels / peacefully.
Sergio se muere de aburrimiento	Sergio is (literally) dying of boredom.	Sergio is dying of boredom / is extremely bored.
Silvia está en las nubes	Silvia is (physically) in the clouds.	Silvia has her head in the clouds / is distracted.
Tania se parte de risa	Tania splits herself with laughter.	Tania bursts out laughing / laughs uncontrollably.
Unax es una tortuga	Unax is a turtle.	Unax is a snail / is very slow.
Xabi pone la mesa	Xabi places the table (somewhere).	Xabi lays the table.

Table A2. Model results with verbal mental age as a predictor in autistic children

Term	β	95% CI	z-value	p-value
Intercept	1.73	[1.23, 2.24]	6.71	< .001
VMA (centered)	0.04	[0.02, 0.06]	3.74	< .001
CONDITION	−0.18	[−0.68, 0.32]	−0.69	.489
VMA × CONDITION	−0.02	[−0.05, 0.00]	−1.81	.071

Table A3. Model results with chronological age as a predictor in autistic children

Term	β	95% CI	z-value	p-value
Intercept	1.59	[1.07, 2.10]	6.05	< .001
AGE (centered)	0.02	[0.00, 0.04]	1.54	.123
CONDITION	−0.01	[−0.44, 0.42]	−0.03	.975
AGE × CONDITION	−0.02	[−0.04, 0.01]	−1.52	.129