

Feasibility and Relevance of an Eye-Tracking-Based Assessment of Morphosyntactic Comprehension in Young Autistic Children

Autism & Developmental Language Impairments
Volume 11: 1–18
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/23969415261451402
journals.sagepub.com/home/dli



Lisa Michel¹, Camille Bataillon¹, Frédérique Bonnet-Brilhault^{1,2}, Joëlle Malvy^{1,2}, Aude Rambault^{1,2}, Laurice Tuller¹ and Marianne Latinus^{1,3} 

Abstract

Background and Aims: Autism spectrum disorder is an early-onset neurodevelopmental condition, characterized by wide heterogeneity, including in language skills. Existing morphosyntactic comprehension assessment tasks are inadequate for many autistic children, leading to low task completion rates. Challenges in the assessment of morphosyntactic comprehension in these children may result in under-informed clinical profiles and hamper in-depth exploration of language profiles in research. Our aim was to develop a passive task (requiring no overt behavior) to assess syntactic comprehension in young French-speaking autistic children. We present the creation of MOSYCAT, a linguistically meaningful intermodal preferential looking task with eye-tracking, and then focus on its feasibility, acceptability, and potential clinical relevance.

Methods: MOSYCAT tests four structures of varying syntactic complexity, simple intransitive sentences, simple subject–verb–object sentences, composite past tense sentences, and sentences with direct object pronominal clitics. Children viewed two videos while they listened to a sentence that matched only one of the videos. Four eye-tracking parameters were measured during their exploration of the videos: Exploration proportion (the proportion of time spent exploring the matching video), exploration accuracy (the proportion of trials in which children explored significantly more of the matching video), orientation latency (the latency of the first saccade toward the matching video), and orientation accuracy (the accuracy of the first saccade). A first version was administered to 56 non-autistic children, and then a reduced version to 23 additional non-autistic children and 21 autistic children of varying clinical profiles. All children were aged 2;6 to 5;11, a period when syntactic performance skills typically mature, but when many autistic children are not (yet) producing sentences.

Results: Feasibility and acceptability were high, including in the autistic group, and more than 85% of autistic children provided usable data. A principal component analysis on MOSYCAT parameters demonstrated that two dimensions, mainly related to accuracy (exploration proportion, exploration accuracy) and speed of response (orientation latency), captured 86% of performance variability in non-autistic children. These dimensions correlated with standardized receptive morphosyntactic measures. The eye-tracking parameters yielded normal distributions, suggesting that MOSYCAT can be used to assess morphosyntactic abilities in all children, demonstrating its clinical relevance. These parameters improved with age in the non-autistic children, suggesting that MOSYCAT is sensitive to the maturation of morphosyntactic comprehension skills. As expected, autistic children performed below the non-autistic children, yet nearly all completed the task, contrary to other tasks, suggesting that MOSYCAT can assess language comprehension in autistic children from across the autism spectrum.

Conclusions and Implications: MOSYCAT, a linguistically informed tool designed to assess morphosyntactic comprehension in young autistic children using a passive, eye-tracking task, proved to be clinically relevant and feasible for young

¹Université de Tours, INSERM, Imaging Brain & Neuropsychiatry iBrain U1253, Tours, France

²Fédération Hospitalo-Universitaire Autisme et Troubles du Neurodéveloppement (FHU Exac-t), Département universitaire de psychiatrie de l'enfant et de l'adolescent, Service Autisme et TND, CHRU de Tours, Tours, France

³Centro de Estudios en Neurociencia Humana y Neuropsicología, Facultad de Psicología, Universidad Diego Portales, Santiago, Chile

Corresponding Author:

Marianne Latinus, Université de Tours, INSERM, Imaging Brain & Neuropsychiatry iBrain U1253, Département universitaire de psychiatrie de l'enfant et de l'adolescent, Service Autisme et TND, 2 Boulevard Tonnellé, 37044 TOURS Cedex 9, France.
Email: Marianne.latinus@univ-tours.fr



autistic children. As such, it has both clinical and research potential for use in the identification and exploration of language profiles in autism. It should be noted that the assessment was not entirely standardized. Variable lighting conditions and break durations, as well as post-hoc calibration correction, were applied. These methodological choices, some of which mirror those common in clinical practice, were made to allow for wide assessment of the autistic children, many of whom required considerable support. Although promising, MOSYCAT is not ready for direct clinical transfer at this time.

Keywords

Language comprehension, morphosyntax, eye-tracking, intermodal preferential looking, autism

Autism spectrum disorder (ASD) is an early-onset neurodevelopmental condition, characterized by wide clinical heterogeneity, including in language skills. These vary in several ways: the aspects of language that may manifest difficulties (receptive versus expressive modality, linguistic domain), the intensity of language challenges, as well as how both of these may change throughout the life of a single individual (Schaeffer et al., 2025). Research on autistic children, from early developmental periods through school age, has highlighted heterogeneous trajectories, both in terms of the timing of speech onset and the linguistic skills developed thereafter (Anderson et al., 2007). Some autistic children will develop language normally, or even precociously, while others will develop language late in life, and around 30% (depending on definition criteria and evaluation methods) will remain with minimal or no expressive language skills (Grandgeorge et al., 2009; Schaeffer et al., 2023; Tager-Flusberg & Kasari, 2013).

The heterogeneity of language profiles in autism has been widely demonstrated in production (Schaeffer et al., 2025), but comprehension has been comparatively less studied, in part due to the difficulty in assessing comprehension (Muller & Brady, 2016). Moreover, study of receptive language abilities has so far over-whelmingly focused on single-word vocabulary, despite frequent calls for studies that go beyond the lexicon (e.g., Plesa Skwerer et al., 2016). While expressive morphosyntactic abilities in autistic children have received considerable attention, notably regarding similarities, for many autistic children, with morphosyntactic difficulties observed in children with Developmental Language Disorder (Schaeffer et al., 2025), comparatively little work has been devoted to sentence comprehension, particularly in young children. Here, we propose a new tool to evaluate morphosyntax in the receptive modality based on a passive evaluation of morphosyntactic comprehension.

To evaluate morphosyntax in its receptive modality, researchers and clinicians mainly rely on standardized tasks and/or parental report. Parent reports are informative and have a clear advantage for assessing language skills in children who may not be able to sustain many forms of direct testing. However, they do not provide a complete picture of comprehension skills, in particular because estimating comprehension skills is difficult for parents, who may

over- or under-estimate their children's skills (Tomasello & Mervis, 1994; Venker et al., 2016). Direct assessment of morphosyntactic comprehension skills in young children typically employs act-out tasks, which have the advantage of using material that is easily obtainable and attractive to children. One of the most used standardized tools for assessing comprehension in very young children (ages 2 to 6) acquiring French is from the *EVALO 2-6* battery (Coquet et al., 2009). The *EVALO* morphosyntactic comprehension task relies on the manipulation of figurines to act-out scenes expressed orally by the evaluator, with items of increasing complexity, ranging from sentences such as *Un personnage marche* "A person is walking" to sentences such as *Le chat que porte la dame tombe* "The cat the lady is carrying falls down." This task requires understanding rather complex instructions (cognitive skills) and handling of figurines (fine motor skills). To this, we can add that the task also requires the young child to inhibit his or her own desire to manipulate or play with the material, to comply with the evaluator's instructions (behavioral regulation and executive function). Hence, this task, while informative for many children, is inadequate for testing many autistic children due to its reliance on several non-linguistic skills that may interfere with task participation or completion, a difficulty which has been widely reported for standardized language tasks (Kasari et al., 2013; Plesa Skwerer et al., 2016; Tager-Flusberg, 1999).

Inadequate assessment of receptive morphosyntax makes it impossible to pursue the goal of identifying receptive language profiles in young autistic children. Yet, establishing such profiles is crucial, clinically, so that clinicians and caregivers can understand the level of comprehension they may expect from a child and adapt their daily communication/treatment plans. Adequate assessment of receptive morphosyntax is also fundamental for research on language in autism to allow for advancement of our understanding of issues such as the relation between language and nonverbal intellectual quotient (NVIQ) and how comprehension and production skills may differ in autism. In the literature on typical and autistic children, there is evidence for language comprehension in the absence of language production, though, in autism, this evidence has so far concentrated mainly on lexical skills (e.g., Chen et al., 2024; Swensen et al., 2007). This literature could be enriched through

more studies using less demanding, passive tasks, in other words, tasks that do not require an active, voluntary response on the part of the child, such as expressive language production, or active behavior (e.g., pointing). The need for such studies has been underlined by several researchers, who have noted the relevance of using techniques such as electroencephalography or eye-tracking (Houston-Price et al., 2007; Plesa Skwerer et al., 2016; Tager-Flusberg & Kasari, 2013).

One of the commonly used paradigms for assessing language development with eye-gaze measures is intermodal preferential looking (IPL; Golinkoff et al., 1987). This paradigm consists of presenting an auditory linguistic stimulus along with side-by-side visual stimuli, one of which matches the auditory stimulus and one of which does not. This method assesses the child's comprehension since the child who has understood the auditory linguistic stimulus is expected to direct his or her gaze toward the matching visual stimulus more quickly, and to look at it for longer than the non-matching stimulus. Naigles and colleagues conducted a series of investigations of morphosyntactic comprehension using the IPL paradigm (with off-line coding) in young autistic children (ages 2–5). They evaluated three different morphosyntactic structures, each in a separate task (Candan et al., 2012; Goodwin et al., 2012; Jyotishi et al., 2017; Naigles et al., 2011; Naigles & Fein, 2017; Naigles & Tovar, 2012; Su & Naigles, 2019; Swensen et al., 2007; Tovar et al., 2015; Wagner et al., 2009). These IPL studies notably reported that some autistic children who were not producing word combinations nonetheless showed good comprehension of simple sentences, looking longer and more quickly at corresponding videos, and furthermore that children followed longitudinally showed a relative developmental advantage of comprehension before production.

The IPL paradigm has until recently been used in combination with off-line manual coding, which involves a human examiner looking at video recordings of the children's behavior during the test session, and coding manually for each time frame whether the child appears to have looked to the right, left, or away from the screen. Eye-tracking allows for an online, direct set-up including automatic coding, leading to greater reliability, and higher consistency of data processing across different conditions and participants. Manual coding is not only subjective, it is also much more labor-intensive (in both coder training and reliability checks and in coding time itself), and spatial and temporal resolution is more limited, compared to automatic coding, though manual coding may reduce data loss (Venker et al., 2020), a point to which we return in the discussion.

The aim of the present study was fundamentally methodological: to ascertain whether an IPL-with-eye-tracking task testing several different structures within the same task could be used to assess general morphosyntactic comprehension in young (autistic) children. We sought to

develop a task that is feasible from a practical point of view, has potential clinical usefulness, is linguistically relevant, and proves suitable to young autistic children. We will henceforth refer to this task by the name MOSYCAT: MORphoSYntactic Comprehension in Autistic and Typically developing children. In this task, participants watch videos of children performing actions either alone or in relation to another child, while listening to short sentences of four different structure types, both syntactically very simple and more complex. MOSYCAT is, to our knowledge, the first tool designed to assess syntactic comprehension passively, while including syntactic structures having a wide range of morphosyntactic complexity within the same task. We report here in detail how the task was created and then present results regarding its acceptability, feasibility, and its clinical and linguistic relevance.

A linguistically relevant task should explore morphosyntactic abilities over a range of structures to provide a global score of each child's morphosyntactic comprehension. Therefore, our task is composed of sentences designed to test the comprehension of four different types of linguistic structures, some simpler/early acquired and others more complex and known to cause difficulties in French-exposed children with a neurodevelopmental disorder (NDD) (see Task Creation Section and Supplementary Material). We designed the task to have potential clinical value by graduating these different levels of complexity in a single task, as short as possible, so that it would be possible to assess the morphosyntactic development of young children, including autistic children, with a wide range of language abilities. Clinical feasibility meant that the entire task had to be short enough to be both suitable for its target population, young autistic children, and incorporable into clinical practice, in settings with the requisite equipment, which we acknowledge at the outset are probably limited to specialized, research-affiliated centers. The task had to include enough items to generalize gaze behaviors to accurately reflect comprehension, while being as short as possible, to align with children's limited attention spans and high fatigability, and as attractive as possible to catch visual and auditory attention and keep interest over time. Since our aim was to arrive at a relevant, feasible, and suitable task, our operational hypotheses are fundamentally methodological. We explored feasibility, acceptability, data usability, and preliminary validity of the task. We then looked at the effect of chronological age and diagnostic group on several measured eye-gaze parameters. In a follow-up study, we investigate the morphosyntactic comprehension-production profiles identified by MOSYCAT in young autistic and non-autistic children.

An appropriate task should have high feasibility, as can be measured by a high participation rate: Are children able to engage in the task, sitting down, and watching the screen? It should likewise have high acceptability, as can be measured by completion rate: Are children able to finish the task—i.e., do they remain seated and focused on the

Table 1. Participant Characteristics.

	Autistic Participants (n = 21)	Non-Autistic Participants		
		Group 1 (n = 56)	Group 2 (n = 23)	Total Non-Autistic (n = 79)
Birth-assigned sex (n)	4 F, 17 M	23 F, 33 M	13 F, 10 M	36 F, 43 M
Age (mo.) μ ($\pm \sigma$) [min max]	54.14 (± 8.73) [42 71]	51.20 (± 11.48) [32 71]	47.78 (± 10.47) [30 64]	50.20 (± 11.24) [30 71]
VSI, WPPSI μ ($\pm \sigma$) [min max]	86.07 (± 17.90) [57 111]	103.57 (± 15.64) [71 151]	114.39 (± 10.30) [91 132]	106.72 (± 15.06) [71 151]
ADOS-2-CSS-SA μ ($\pm \sigma$) [min max]	8.00 (± 1.97) [4 10]			
ADOS-2 CSS-RRB μ ($\pm \sigma$) [min max]	6.84 (± 2.01) [4 10]			

Note. VSI = Visual Spatial Index. WPPSI = Wechsler Preschool and Primary Scale of Intelligence. CSS-SA = calibrated severity score, social affect /10, ADOS-2. CSS-RRB = calibrated severity social, restricted and repetitive behavior /10 (autism diagnostic observation schedule-2). ADOS scores were available for 19/21 participants.

screen for the duration of the task and therefore contribute a complete recording? A successful eye-tracking task should also provide a sufficient amount of usable data. Finally, we evaluated preliminary validity by examining the relationship between parameters derived from MOYSCAT and the standardized measure of morphosyntactic comprehension from the EVALO battery. We varied task length, hypothesizing that a shortened version of the task would maximize feasibility, acceptability, and data usability for further analysis. A valid task to evaluate morphosyntactic abilities in all children should yield data with some variability, and a distribution shape close to that of a normal distribution, i.e., without extreme tail behaviors, approximatively symmetric and without clear ceiling or floor effects.

Regarding chronological age, although typically developing children in this age range already show considerable morphosyntactic comprehension, we expected that performance would increase with increasing age in non-autistic children (e.g., faster orientation to the matching video or increased exploration of the matching video), since older children have faster general processing speed.

The literature shows that while some autistic children display typical morphosyntactic development, many show delayed expressive language, and IPL evidence suggests that comprehension can also be delayed. These considerations led us to expect that young autistic children, as a group, would perform somewhat below non-autistic children in the same age range, if each of the MOSYCAT parameters is a reliable indicator of morphosyntactic comprehension.

Methods

Ethics

This research was approved by the Comité d’Ethique de la Recherche Tours-Poitiers and the Comité de Protection des Personnes EST 1 (2017/23-ID RCB: 2017-A00756-47;

PROSCEA). Children were recruited in nursery schools and kindergartens in Tours, France. Autistic children were from the Centre-Val de Loire region and the child psychiatry department of the Tours University Regional Hospital. Parents and children were informed of their rights and of all aspects of the experimental design; parental consent was obtained in the form of a signed written consent form. Children’s agreement was also sought, verbally, when possible, or non-verbally, based on their behavior.

Participants and Global Experimental Protocol

We recruited children aged 2;6 to 5;11. A total of 100 participants, divided into three groups, took part in our study (Table 1). All participants had normal or corrected-to-normal vision and normal hearing, based on parental or clinical reports. Bilingual language exposure was not an exclusionary criterion, as we aimed for population samples that reflected the cultural diversity of the population consulting at the public center where we recruited the autistic children and in the surrounding community.¹ All participants had all-day monolingual exposure to French during the week in daycare centers or at school. Most children were also exposed to French at home, monolingually, or, for some, bilingually (14/56 bilingual exposure in Group 1, 4/23 in Group 2, and 15/21 in the ASD group). French was the dominant language spoken at home for all these children, except two autistic children. Nonverbal reasoning abilities were assessed using the block design and object assembly subtests (generating a visual spatial index) from the age-appropriate Wechsler Intelligence Scales (Wechsler Preschool and Primary Scale of Intelligence 4th ed., Wechsler, 2012).

The first version of MOSYCAT was administered to 56 children (23 girls; mean age: 51.20 months old (± 11.48) [32 71]) who were recruited in nursery schools and kindergartens. We henceforth refer to these children as Group 1. The second version of MOSYCAT was administered to

a second group (Group 2) of 23 children (13 girls; mean age: 47.78 months old (± 10.47) [30–64]), recruited in nursery schools and kindergartens and by word of mouth. Children in Group 1 and Group 2 had no history of neurological or psychiatric disorders or early learning disabilities (notably no NDD diagnoses).

Twenty-one autistic children (ASD group) (4 girls; mean age: 54.19 months old (± 8.53) [42–71]) were administered the second version of MOSYCAT.² Each of these children had been diagnosed with ASD by a multidisciplinary team of expert clinicians, using DSM-5 and ICD-10, and the Autism Diagnostic Observation Scale-Second Edition (ADOS-2; Lord et al., 2012) and/or the Autism Diagnosis Interview-Revised (ADI-R; Le Couteur et al., 2003). We sought to have a population sample that represented as closely as possible the wide spectrum variability in autism. To meet the objectives of our protocol, we deliberately included participants from all developmental levels. Our sample thus included children with varying levels of autistic symptomology as measured by ADOS-2 calibrated severity scores (CSS) for social affect (SA) and for restricted and repetitive behaviors (RRB) (Esler et al., 2015; Hus et al., 2014). CSS-SA scores ($M = 7.95$, $SD = 1.93$) and CSS-RRB scores ($M = 6.84$, $SD = 2.01$) both ranged from 4 to 10/10, and thus including children with concern ranges characterized in these tools as “mild-to-moderate” and “moderate-to-severe.”³ These 21 autistic children also varied in intellectual ability, as seen in both their NVIQ scores, measured by the visual spatial index of the WPPSI (range 57–111, $M = 86.07$, $SD = 17.90$) and whether or not they had accompanying intellectual disability (ID; Full Scale IQ < 70), which varied from mild to severe (8 with no ID, 5 with mild ID, 7 with moderate ID, and 1 with severe ID).⁴ In the center where all autistic children were recruited, the diagnostic process includes a clinical scale on which the center’s expert clinicians rate children’s language levels based on clinical observation and testing (including by speech-language pathologists). The expressive modality scale ranges from 0 to 5 as follows: 0—no language, 1—babbling, 2—isolated and/or juxtaposed words, 3—simple sentences, 4—embedded or coordinated sentences, 5——complex language. Our sample included 15 children who were not producing sentences (3 children rated 0, 7 rated 1, and 5 rated 2) and 6 children who were producing sentences (1 rated 3, 3 rated 4, and 2 rated 5).

MOSYCAT was part of a larger language assessment protocol. Two tasks from the standardized EVALO 2-6 battery (Coquet et al., 2009) assessed, respectively, receptive phonology and receptive morphosyntax: the EVALO-phoneme discrimination subtest and the EVALO object manipulation for morphosyntactic comprehension subtest. Language was also assessed with two language impairment testing in multilingual settings (LITMUS) tasks: LITMUS-nonword repetition (LITMUS-QU-NWR-Toddlers, Ferré et al., 2022) and LITMUS-sentence repetition (LITMUS-

SR-FR-Toddlers, Tuller & Prévost, 2022). The two LITMUS tasks assess, respectively, phonology and morphosyntax in production. The toddler versions used here are experimental tasks, adapted from tasks with a very broad empirical basis, including autistic children and adults, and are recommended for assessment of language in autism (Schaeffer et al., 2023). Only respective participation and completion rates for these tasks will be presented here, except for the *EVALO 2-6* object manipulation for morphosyntactic comprehension subtest (Coquet et al., 2009), which will be compared to MOSYCAT scores, in Group 1 non-autistic children. Administration of this subtest was discontinued in Group 2, and in the ASD group due to its length and difficulty, participants had maintaining engagement. Not only did administration time in Group 1 cause overall protocol testing time to far exceed what had been promised to families (and to the ethics committee), but also in the ASD group, participants were highly distracted by the figurines (throwing them, playing with them) leading to an inability to engage in the task and/or complete it.

Autistic children were tested in a university hospital autism center, where they receive clinical services. Children in Group 1 were tested in a quiet room at their daycare center or school; children in Group 2 were tested at the same clinical center as the autistic children. The protocol tasks were distributed into two sessions for all children, such that Part A and Part B of MOSYCAT were in separate sessions, as were the two EVALO subtests, and the two LITMUS repetition tasks. For Group 1 participants, who were given the long version of MOSYCAT, the two testing sessions took place on separate days, a couple of days to roughly a month apart. For Group 2 participants, who were given the short version of MOSYCAT, the two testing sessions were grouped on the same day, separated by a small break lasting about 30’ (except one participant whose second session was a week later). For most children in the ASD group, the two testing sessions took place over different days, to minimize child fatigability and accommodate family/clinical caregiver availability. Six autistic children had both sessions on the same day. Fourteen autistic children had the two sessions on different days, separated by about a week (6 children), a month ($n = 4$), 2 months ($n = 3$), or 4 months ($n = 1$). One child completed only one session.

Task Creation

We describe here in detail the different steps in the creation of MOSYCAT. First, we selected four syntactic structures varying in level of complexity (syntactic computation involved) and corresponding to the age of acquisition in young TD French-speaking children and to the degree of difficulty for French-speaking children with language impairment (see Supplemental Material for motivation behind structure selection). There is considerable evidence

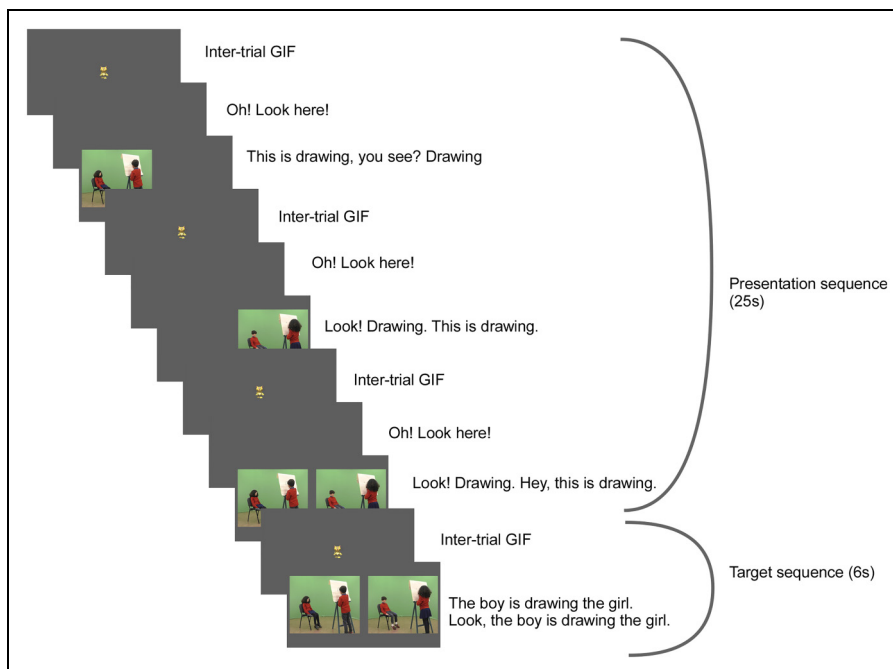


Figure 1. Example of a complete audiovisual trial. The presentation sequence (25s) comprises the action introduction and the neutral presentation (when the two videos are presented simultaneously with the action verb in its infinitival form).

that the complexity of the operations involved in generating syntactic structures for comprehension and production underlies their developmental sequences and difficulty levels, including in autistic children (Tuller et al. 2017). MOSYCAT assesses the following four structures: (a) a baseline structure consisting of intransitive verbs in simple subject + verb sentences (SV; e.g., *La fille saute* “The girl is jumping”), (b) word-order in simple (transitive) Subject + verb + object sentences, with a reversible animate agent and patient (i.e., reversing the agent and the patient noun phrases yields a grammatical sentence—e.g., *La fille dessine le garçon* “The girl is drawing the boy” vs. *Le garçon dessine la fille* “The boy is drawing the girl”) (SVO), (c) composite past tense in transitive sentences with an animate subject and an inanimate object (compared to the same sentences in the present tense) (TENSE; e.g., *La fille a mangé les gâteaux* “The girl ate the cookies” vs. *La fille mange les gâteaux* “The girl is eating the cookies”), and (d) reversible SVO sentences with an accusative clitic pronoun object, which in French undergoes syntactic movement to the left of the verb (CLIT; e.g., *La fille la chatouille* “The girl is tickling her”) (See Supplementary Material—Table S1). SV was included as a baseline, as it requires only lexical information processing but no syntactic comprehension. In contrast, each of the other three structures requires syntactic comprehension: understanding who did what to whom (identification of the agent and the patient with respect to the action), in the case of SVO and CLIT, or distinguishing the syntax of the present tense (a lexical verb only) versus that of the composite past tense (an auxiliary and a lexical

verb), in the case of TENSE. These four structures were selected because of their variable complexity levels, to ensure assessment across the wide range of morphosyntactic comprehension levels that could be expected to be found in young autistic children, particularly those with limited expressive language (see Supplementary Material).

Sentences were made up of words common in young children’s vocabulary, based on the results from the French Inventory of the Communicative Development IFDC 8–16 months and 16–30 months (Kern & Gayraud, 2010). Auditory stimuli were produced by a speech-language pathologist with considerable experience working with children with NDD and considerable acting experience in theater for children. We followed Naigles & Tovar (2012) in using exaggerated motherese prosody to foster child engagement in the task. For each structure, we recorded eight samples with different sentences, leading to a total of 32 target sentences (8 samples by 4 conditions). In addition to the target sentences, we recorded different auditory stimuli, such as *Eh, regarde ici* “Hey, look here,” or *ça c’est [action verb]* “It’s ...,” which presented each action (Figure 1). Four child actors (2 boys, 2 girls; about 10 years old) were recruited to act in the video scenes accompanying the auditory recordings (See Supplementary Material). Parent and child consents were obtained for the capture and use of their image for research and academic purposes. The visual scenes were shot at a university professional audiovisual studio. The videos were silent, with no movement of the mouth to avoid visual attention bias to faces during social interaction. The visual information needed for sentence comprehension relied on exaggerated

movements of the body, or gestures, and never merely on facial expressions. Actors performed the actions corresponding to the target sentences. All videos lasted about 6 s. Auditory sentences and visual scenes were then combined in single audiovisual stimuli using Adobe Premiere Pro (22.2) and DaVinci Resolve (v.18.1.1.7).

A single trial lasted approximately 31 s, and all trials were organized as follows (Figure 1). First, to introduce the action, the two videos were presented one after the other, on each side of the screen, conjointly with an auditory stimulus corresponding to an action verb in its infinitival form, referred to as the action introduction. Then, the two videos were displayed simultaneously on each side of the screen, conjointly with the same action verb auditory stimulus, referred to as the neutral presentation. Together they form the presentation sequence which was intended to familiarize participants with the videos and the verb, thereby avoiding potential bias due to prior lexical knowledge (or lack thereof). This sequence was also designed to present the visual configuration of the stimuli used in the target sequence and thereby reduce any novelty effect to minimize free visual exploration and to maximize exploration due to syntactic comprehension. Finally, the target sequence was presented: the two videos were presented simultaneously, on either side of the screen, as in the neutral presentation, but conjointly with the auditory target sentence. Visual scenes were separated by an inter-stimulus interval during which a GIF cat was presented in the center of the screen to center the child's gaze. Since sentences of varying complexity also varied in length, we adjusted blank and inter-stimulus time durations.

The first version of MOSYCAT was composed of 32 trials, 8 per syntactic structure (see Table 2). The first version was deliberately long, to allow for the subsequent selection of the best items for the creation of a shorter version. It lasted approximately 24 min. Trials ($n=32$) were presented in two parts. Each part started and ended with an SV trial, and trial order was otherwise randomized within each part. The two parts were given to children in random order on separate days (i.e., two sessions, a couple of days to roughly a month apart). Both parts began with a short video presentation of the actors. Part A lasted 13 min and tested all linguistic structures but CLITIC, while

Part B lasted 12 min and included all structures but SVO. A third part (part C, administered to 8 children overall), which consisted of a mix of items from parts A and B, was created to compensate for any recording issues (e.g., too few gaze recordings, difficulty in testing etc.) during the first session.

To be able to test a wide range of young autistic children, we sought to shorten Version 1, after running it on Group 1 (non-autistic) participants, reasoning that the shortest possible version would be more likely to be clinically feasible for autistic children. To do this, we selected the target sentences with the highest success rates (more than 50% of children performed correctly) based on (a) dwell time (DT) to the matching video during the target sequence, and (b) the comparison of DT to the matching video between target and neutral sequences. These different steps resulted in a final stimulus set comprising four SV, four SVO, four CLITIC, and three TENSE. Since the TENSE structure includes in fact two substructures—tense modalities (past/present), we chose to keep three target sentences per substructure, for a total of six trials for TENSE.

Across the 32 trials of Version 1, we had balanced the proportion of actions with a boy versus a girl agent, the proportion of videos in which the first display of the trial started on the right or on the left of the screen, and the proportion of trials for which the audio-matching video was on the right or on the left. In version 2, trials were no longer fully balanced within structures but were nearly balanced across the task (Table S2). We considered that preferential looking was not biased to either side of the screen due to either the gender of the agent or an over-presentation of matching videos on a particular side.

Version 2 of MOSYCAT included 18 trials: 4 SV, 4 SVO, 4 CLITIC, and 6 TENSE (3 target present tense and 3 target composite past tense). It lasted approximately 13 min (Table S2), divided into two parts; all structures were included in both parts: 7 min for part A, which included an introduction to the video characters, and 6 min for part B, in which characters were not introduced. The two parts were always presented in the same order (A, then B), but the presentation order of trials was randomized within each part, except that each part began and ended with an SV trial. When the second testing session was on a different day (see above), a laminated screenshot of the actors was used as a reminder of actor identity.

Table 2. Characteristics of MOSYCAT Version 1 (Long) and MOSYCAT Version 2 (Shortened).

	MOSYCAT Version 1		MOSYCAT Version 2	
	Part A	Part B	Part A	Part B
# Items	16	16	9	9
Duration (minutes)	12'	13'	7'	6'

Note. Items in each part began and ended with an SV structure, but were otherwise in randomized order.

Data Acquisition and Processing

Eye-tracking data from MOSYCAT were acquired at a 120 Hz sampling rate with the Tobii Pro-fusion eye-tracking device (Tobii® Technology, Sweden), in a free-movement set-up (no chin rest, and the child was not restrained in any way). Stimuli were broadcasted on a 1920 × 1080 LED presentation screen (F27T450FQR, SAMSUNG, Seoul, South Korea; refreshing rate: 75 Hz). Participants were seated approximately 70 cm away from

the screen, either alone, or on the lap of a trusted adult. Sounds were delivered through a portable speaker positioned behind the screen, equidistant from the right and left of the screen (Go 3, JBL, Northridge, USA) and set at a comfortable sound level. Lighting conditions were adjusted, according to the room where the experiment took place (lab or room at a school/daycare center), to maximize data quality in eye-tracking recordings and child comfort. Children's gaze direction was recorded continuously. There were no explicit instructions, except to look at the screen. Throughout the testing session, the experimenter monitored the eye-tracking recording, to ensure that the eye-tracking system was continuously capturing the participant's eyes. Calibration was performed using 5 or 2 target points which could either be a simple circle or an animated GIF with sound to ease calibration (Sasson & Elison, 2012). In case of poor or unavailable calibration during acquisition, we applied a correction using in-house MATLAB scripts for recordings exhibiting a systematic offset calibration, and if notes about the child's behavior during acquisition demonstrated the child was paying attention to the screen. The procedure consisted of isolating recording sections corresponding to the inter-stimulus intervals where the stimulus remained at the center of the screen and performing a translation along the X and Y axis to compensate for the constant global offset. This correction was then applied to recalibrate the gaze position of the child's complete recording (Franck et al., 2012). This likely had no influence on our results as areas of interest (AOI) were large, and our measures did not rely on the location of participant fixations but rather on participant exploration. Counting two recordings per child (the two parts of the task), the percentage of total recordings for which this correction was used was 7.6% in Group 1, 20% in Group 2, and 40% in the ASD group.

AOI and time of interest (TOI) were defined directly in the recordings using the Tobii Pro Lab software (v.1.207). AOI corresponded to the location of each video (matching and non-matching); TOI, representing the time portion variable, could either be the total length of the target sequence (Total) or correspond to two successive TOI representing halves of the target sequence (1st Half 1; 2nd Half). Data were exported using the I-VT fixation gaze filter algorithm proposed by Tobii to categorize datapoints as fixations (http://www.vinix.co.kr/ivt_filter.pdf).

Data were preprocessed using in-house MatLab scripts (9.9.0.2037887). Gaze position data were averaged across the two eyes. Missing data on time periods smaller than 200 ms were interpolated (Leppänen et al., 2015). The I-VT fixation gaze filter proposed by Tobii automatically labels events, such as saccades and fixations, and unidentified events. We measured DT separately for both the matching and non-matching video, i.e., the time spent exploring each visual scene, including saccades and fixations. Based on measures proposed in Tovar et al., 2015 (coded off-line, by hand), two types of parameters recorded during the target

were used: those based on DT, and those based on the first orientation (FO). For analyses of DT-based parameters, a trial was considered valid if the total DT in both the neutral and target sequences was superior to the DT threshold (900 ms). The DT threshold, derived from Group 1 data, corresponded to the minimum DT in the 95% most looked at trials, across matching and non-matching video, considering all conditions and children. The threshold was 680 ms for the target sequence, and 908 ms for the neutral presentation. We chose a cut-off of 900 ms, rounding the neutral presentation threshold. For analysis of FO-based parameters, a trial was considered valid if DT in the neutral sequence only was greater than the threshold.

To evaluate the FO, saccade information was extracted, using in-house MatLab scripts (Kovarski et al., 2019), only if it met the following conditions: (a) the saccades occurred 3000 ms after the onset of the cat-GIF, corresponding to the appearance of both videos on the screen and (b) gaze hit an AOI, whether matching or non-matching.

DT measured the time spent exploring the visual scene, in a pre-defined AOI, corresponding to the size of the video on the screen or, when two videos appeared side by side, the whole half of the screen. Other parameters derived from DT were calculated considering different TOI (total length of the target sequence; halved times of the target sequence - time portion), following Naigles and collaborators (Candan et al., 2012; Naigles et al., 2011; Tovar et al., 2015). To characterize children's eye-gaze behavior, two DT-based parameters were used. Exploration proportion corresponds to the proportion of DT toward the matching video with respect to the total (or halved) DT (e.g., matching + non-matching video). This proportion was calculated on a trial-by-trial basis and then averaged across trials; values above 50% indicate that the child looked more at the matching videos. A second parameter, the exploration accuracy, corresponded to the percentage of correct trials with respect to the total number of valid trials per condition. A trial was considered correct if exploration proportion was significantly higher than 50% ($p < .05$ on chi-square test), meaning that exploration accuracy was more stringent than exploration proportion.

Finally, we measured the latency and accuracy of the FO to evaluate "automatic processing" of linguistic information. Orientation latency corresponds to the latency (in milliseconds) of the first correct saccade (toward the matching video) with respect to the onset of the two videos in the target sequence. Orientation accuracy corresponds to the percentage of trials where the first saccade was made toward the matching video (e.g., a correct response) relative to the number of valid trials.

Statistical Analysis

Statistical analyses were performed following our hypotheses. The feasibility of the different tasks was assessed with a measure of participation. "Participation" refers to

the percentage of participants who engaged in the task, meaning they complied with task instructions and had the necessary skills to perform the task as instructed, regardless of data quality, completion rate, and accuracy scores. Acceptability of the different tasks was explored by evaluating completion rate, i.e., the percentage of participants able to finish the task according to any stopping criteria. For MOSYCAT, participation refers to the percentage of participants who participated in both task sessions (Part A and Part B), and completion refers to the percentage of participants who completed both of these sessions, meaning they remained seated and focused on the screen for the duration of the task providing complete recordings. MOSYCAT data usability was measured as the percentage of children providing usable data, i.e., with at least one valid trial per linguistic structure and at least 50% of valid trials overall.

Analysis of the eye-tracking parameters was performed using data recorded for the trials included in Version 2 of MOSYCAT ($n = 18$); as some items were not present in Version 1, only data from trials common to Group 1 and Group 2 and ASD were used for analysis of Group 1 ($n = 15$). The effect of task length on DT and number of valid trials was evaluated by means of Mann–Whitney non-parametric tests comparing the two groups of non-autistic children using only items common to both groups ($n = 15$). Preliminary validity was assessed by evaluating the correlation of our results with those of a standardized assessment of language comprehension. Only participants from Group 1 were included in this analysis. A principal component analysis (PCA) was used to identify the most relevant eye-gaze parameters. Eye-tracking data, averaged across structure, were z-transformed within parameters and across participants. Loadings on the first two components of the PCA were correlated with raw scores obtained on the EVALO morphosyntactic comprehension subtest using Spearman correlations.

To evaluate data validity, we measured moments of the distribution shape of each parameter and tested whether the distribution was normal using a Shapiro–Wilks test in the autistic participants ($n = 21$) and in the non-autistic participants ($n = 79$). To test our hypotheses regarding language development in non-autistic children, we tested the effect of age on data from the whole target sequence. All parameters met the assumptions of the linear model (no outliers, residual normality, homoscedasticity) and were therefore analyzed by means of a linear model with age as a continuous variable. To test our hypotheses regarding the effect of diagnosis on children’s behavior, we took into account potential differences between the first and second halves of the target sequence (Naigles et al., 2011), analyzing exploration proportion and exploration accuracy with a two factor mixed design ANOVA with diagnosis (two levels: ASD/non-ASD) and time portion (two levels: Half1/Half2) as within-subject factors. If the ANOVA

assumptions were not met, we relied on robust ANOVA. Effects of diagnosis on orientation latency and accuracy were evaluated by means of two-sample t-tests or their non-parametric equivalent.

Results

Feasibility and Acceptability

Beginning with MOSYCAT acceptability and feasibility, Table 3 presents participation and completion rates for each of the tasks in our experimental protocol. Shortening the task yielded higher feasibility and acceptability of the task even in non-autistic children (i.e., comparison of GRP1 and GRP2). Out of all tasks in the protocol, MOSYCAT presented the highest participation and completion rates for the autistic children.

Data Usability

Data usability was assessed by measuring the percentage of participants with sufficient valid data, i.e., enough trials meeting the defined minimum DT thresholds (see Methods Section). A child’s valid data was considered to be sufficient when (a) there was at least one valid trial per condition and (b) at least 50% of the child’s total trials were valid; henceforth, data for the remaining children were averaged over a minimum of 9 trials. We note with regard to (a) that the number of participants having only one valid trial for any given structure was very low: 8/56 (14.29%) in Group 1, 2/23 (8.69%) in Group 2, and 1/21 (4.76%) in the ASD group. Since our objective was a global score for morphosyntactic comprehension, we reasoned that (a) and (b) were both needed to ensure that children’s performance would indicate their ability to understand sentences. We found that more than 85% of children in each group provided sufficient data (Table 4).

The effect of task length was evaluated by comparing the two TD groups on the percentage of valid trials and mean exploration duration using the 15 items common to both versions of the task. The percentage of valid trials did not differ between Group 1 ($M = 97.9$, $SD = 3.84$) and Group 2 ($M = 98.18$, $SD = 5.88$) (Mann–Whitney $U = 496$, $n_1 = 51$, $n_2 = 22$, $p = .28$ two-tailed, effect size = 0.13). Mean exploration duration during the target sequence was significantly shorter in Group 1 ($M = 5087$ ms, $SD = 566$ ms) than in Group 2 ($M = 5363$ ms, $SD = 670$ ms) (Mann–Whitney $U = 345$, $n_1 = 51$, $n_2 = 22$, $p = .0096$ two-tailed, effect size = 0.30).

Participants who did not provide usable data, whether because they did not participate, or complete the task, or because data was of poor quality, were removed from further analyses: 5/56 from Group 1, 1/23 from Group 2 and 3/21 from the ASD group.

Table 3. Comparison of Task Participation and Completion Rates (%) for Each Participant Group.

Task	Group 1 (Non-Autistic)		Group 2 (Non-Autistic)		ASD	
	Participated % /56	Completed % /56	Participated % /24	Completed % /24	Participated % /21	Completed % /21
MOSYCAT (IPL + eye-tracking task)	87.50	80.36	95.65	95.65	95.24	85.71
EVALO morphosyntactic Comprehension	94.6	92.9				Discontinued
EVALO-phoneme discrimination	100	100	100	100	28.6	10
LITMUS-sentence repetition	92.9	92.9	100	95.7	42.9	38.1
LITMUS-nonword repetition	96.4	91.1	100	100	42.9	42.9
WPPSI-block design	100	100	100	100	66.7	52.4
WPPSI-object assembly	100	100	100	100	95.2	61.9

Note. Participated = child agreed to participate (in each session of) the task proposed. Completed = child completed (each session of) the task. ASD = autism spectrum disorder; IPL = intermodal preferential looking.

Table 4. MOSYCAT Participation, Completion, and Valid Trial Rates for Each Participant Group (*n*/Group Total and %).

Group	Participated in Each MOSYCAT Session	Completed Each MOSYCAT Session	Sufficient Number of Valid Trials
Group 1	49/56 = 87.50%	45/56 = 80.36%	51/56 = 91.07%
Group 2	22/23 = 95.65%	22/23 = 95.65%	22/23 = 95.65%
ASD	20/21 = 95.24%	20/21 = 95.24%	18/21 = 85.71%

Note. Sufficient number of valid trials = number and percentage of participants having a minimum of 1 valid trial per condition (syntactic structure) and for whom overall 50% of task trails yielded valid data.

ASD = Autism Spectrum Disorder.

Preliminary Validity

Data from the non-autistic children in Group 1 were included in a PCA run over the four eye-tracking measures. Two principal components (PCs) accounted for 86.5% in data variability (63.2% by PC1; 23.3% by PC2) (Figure 2A). PC1 was better explained by DT for matching videos, with contributions of exploration proportion and exploration accuracy of 31.08% and 31.76%, respectively. PC2 was better explained by orientation latency (57.56%). Spearman correlations with scores on the EVALO morphosyntactic comprehension subtest (Figure 2B) revealed a strongly significant positive correlation with PC1 (DT for matching video) ($\rho=0.48$, $p<.001$) and a non-significant negative correlation with PC2 (orientation latency) ($\rho=-0.26$, $p=.06$).

Data Distribution in Each Group

We expected a valid task to assess morphosyntactic abilities to produce relatively normally distributed parameters in our population samples. Here we report the Shapiro–Wilk tests and moments (skewness and kurtosis) of the distribution for each parameter in both groups (Figure 3). Exploration proportion followed a normal distribution in both groups (non-autistic: $W=0.98$, $p=.33$; autistic: $W=0.94$, $p=.34$). The distribution of exploration proportion was approximately

symmetric and close to the normal distribution in non-autistic children with minimal skew (0.42) and kurtosis (3.1) in the expected range for a normal distribution. For autistic children, the distribution was symmetric (skewness = -0.24) with slightly fatter tails (moderate negative excess kurtosis = -1.06).

Exploration accuracy followed a normal distribution in the autistic group ($W=0.91$, $p=.09$), but not in the non-autistic group ($W=0.96$, $p=.04$). The distribution of exploration accuracy was approximately symmetric in the non-autistic group (non-autistic = 0.044) but was left skewed in the autistic group (skewness = -1). Kurtosis was in the normal range for both groups (autistic group: excess kurtosis = 0.74; non-autistic: -0.12), although slightly peakier in the autistic group.

Orientation latency followed a normal distribution in the autistic group ($W=0.92$, $p=.13$) but not in the non-autistic group ($W=0.92$, $p<.001$). Latency of the FO did not follow a normal distribution; it had a very peaky shape (excess kurtosis = 1.16) and a strong rightward asymmetry (skewness = 1.09) in non-autistic participants. In autistic participants, the distribution had a slightly right-skewed normal distribution (excess kurtosis = 0.07; skewness = 0.71). This demonstrates the presence of two outliers in the non-autistic groups with delayed first correct responses.

Finally, orientation accuracy followed a normal distribution in the non-autistic group ($W=0.98$, $p=.26$) but not in

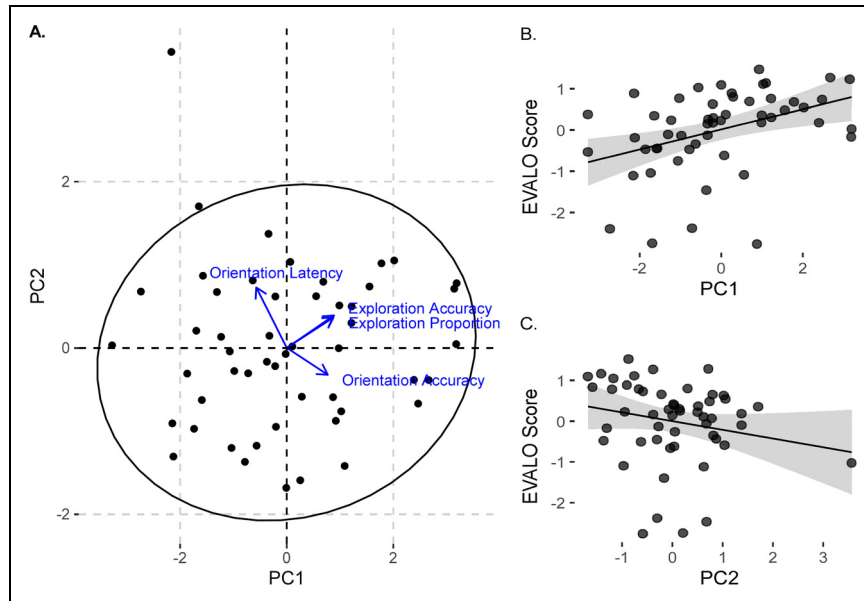


Figure 2. Preliminary validity. A) Results of the principal components analysis (PCA). B) Correlation analyses of PC1 (left) and PC2 (right) with the scores at EVALO morphosyntactic comprehension subtest. Dots represent individual Group I participants ($n = 56$). PCA = principal component analysis; PC = principal component.

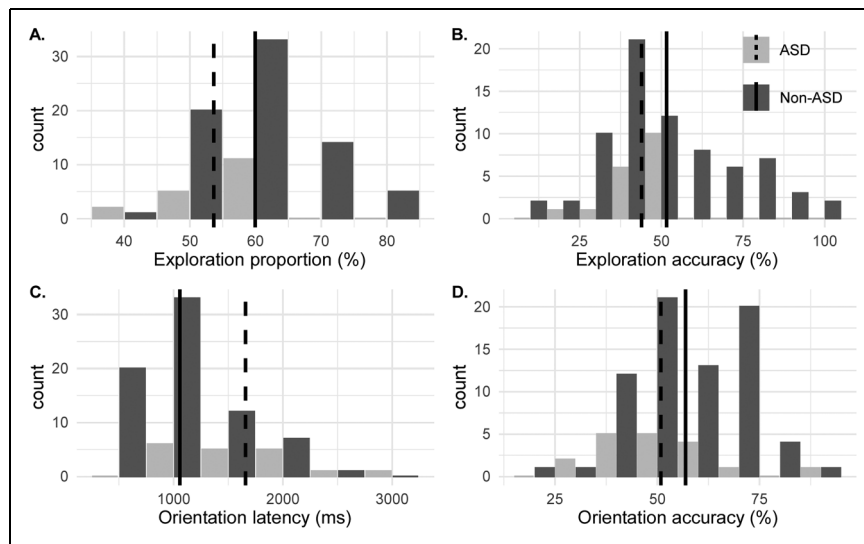


Figure 3. Distribution of the eye-tracking parameters in autistic and non-autistic participants (non-autistic ($n = 79$): dark gray ($n = 21$); autistic: light gray). A) Exploration proportion; B) exploration accuracy; C) orientation latency; D) orientation accuracy.

the autistic group ($W = 0.89, p = .04$). The accuracy of the first saccade presented a distribution with moments in the range of the normal distribution in the non-autistic group (skewness = -0.09 ; excess kurtosis = -0.11). In the autistic group, moments were not in the range of the normal distribution. The distribution was rightly skewed (skewness = 1.23) and peakier (excess kurtosis = 1.44) than a normal distribution, driven by some children with higher performance.

Age Effect in the Non-Autistic Participants

For exploration proportion, a linear regression in non-autistic participants ($n = 73$), indicated a significant positive linear relationship between age and exploration proportion ($\beta = .302, F(1,71) = 12.02, p < .001$; Figure 4A); the model explained 13% of inter-subject variance (adjusted $R^2 = 0.13$). A significant positive linear relationship was also observed between age and exploration accuracy ($\beta = .51,$

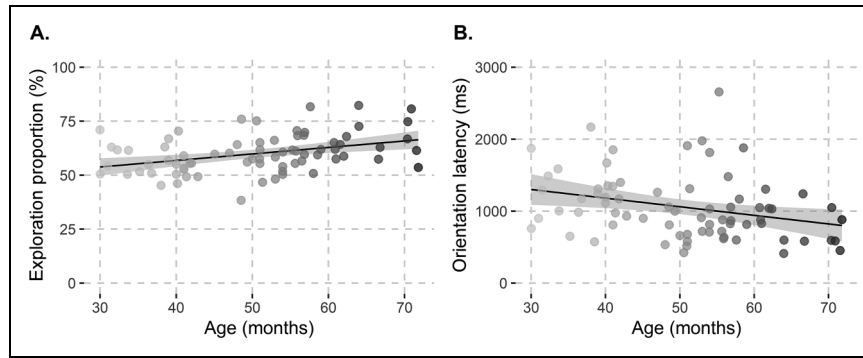


Figure 4. MOSYCAT exploration proportion and orientation latency in non-autistic participants according to age (X-axis). Individual participants ($n = 79$) are represented as dots. A) Mean exploration proportion (Y-axis). B) Orientation latency to the matching video.

$F(1,71) = 6.78$, $p = .012$), although the model explained only a small proportion of the inter-subject variance (adjusted $R^2 = 0.07$). Exploration proportion and accuracy increased with increasing age.

There was a significant negative linear relationship between age and orientation latency with a decrease in the latency of the first correct saccades toward the matching video with increasing age ($\beta = -11.97$, $F(1,71) = 6.79$, $p = .011$; Figure 4B), although the model explained only a small proportion of inter-subject variance (adjusted $R^2 = 0.07$). Orientation accuracy did not show a significant linear relationship with age ($\beta = 0.15$, $F(1,71) = 1.24$, $p = .27$, $R^2 = 0.003$).

Effect of Diagnosis

The effect of diagnosis was evaluated based on the performance split according to the first and second halves of the target sequence. Exploration proportion was analyzed by means of a two-way mixed ANOVA. Exploration proportion was significantly lower ($F(1,89) = 6.77$, $p = .011$) in the ASD group ($M = 53.93$, $SE = 8.64$) than in the non-ASD group ($M = 59.78$, $SD = 10.66$). Time portion ($F(1,89) = 10.27$, $p = .002$) also affected exploration proportion but with no interaction with diagnosis ($F(1,89) = 0.008$, $p = .93$). Children, regardless of their diagnosis, looked more at the matching video during the second half ($M = 60.86$, $SD = 11.81$) than during the first half of the trial ($M = 56.39$, $SD = 8.57$) (Figure 5A).

Similar results were obtained for exploration accuracy with a robust two-way mixed ANOVA. Autistic children had lower performance ($M = 43.36$, $SD = 13.15$) than non-autistic children ($M = 51.18$, $SD = 16.93$; $F(1, 17.22) = 5.498$, $p = .031$). All children performed better in the second half ($M = 53.62$, $SD = 17.34$) than in the first half ($M = 45.65$, $SD = 14.69$) of the trial ($F(1, 15.37) = 5.46$, $p = .033$).

Orientation latency was higher in autistic children ($M = 1658$ ms, $SD = 524$ ms) than in non-autistic children ($M =$

1057 ms, $SD = 453$) (Mann–Whitney $U = 1083$, $n_1 = 73$, $n_2 = 18$, $p < .001$ two-tailed; effect size = 0.44) (see Figure 5B). Orientation accuracy did not differ between autistic ($M = 50.89$, $SD = 14.27$) and non-autistic ($M = 56.91$, $SD = 12.61$) children ($t(89) = -1.77$, $p = .08$).

Discussion

Our aim was to determine whether an IPL task testing several different syntactic structures could be used to assess morphosyntactic comprehension in autistic children. We sought to develop a task suitable for young autistic children that has practical feasibility, potential clinical usefulness, and linguistic relevance. To that aim, we used an IPL paradigm combined with eye-tracking to assess morphosyntactic comprehension. In designing MOSYCAT, we considered practical, methodological, and theoretical constraints in the creation of a first and a second version of the task, with future clinical application in mind. In our exploration of the validity of this tool to assess morphosyntactic comprehension, we formulated several hypotheses regarding the distribution of the parameters that we measured and the effects of age and diagnosis on these parameters.

Our first consideration was to create a protocol that was as short as possible to allow easier assessment of autistic children and assessment of a wider range of autistic children across the spectrum. We designed a long version of the task and then reduced it based on the selection of items on which performance was the highest in a group of young non-autistic children (Group 1). As expected, shortening the task increased (a) the percentage of participants who provided usable data and (b) the time spent exploring the visual scenes during the target sequence, without impacting the percentage of valid trials. To increase acceptability and feasibility, the task was split into two sessions that could be held on either the same or different days. Running the two sessions on the same day as in Group 2 could yield larger fatigue and decrease task performance, yet Group 2

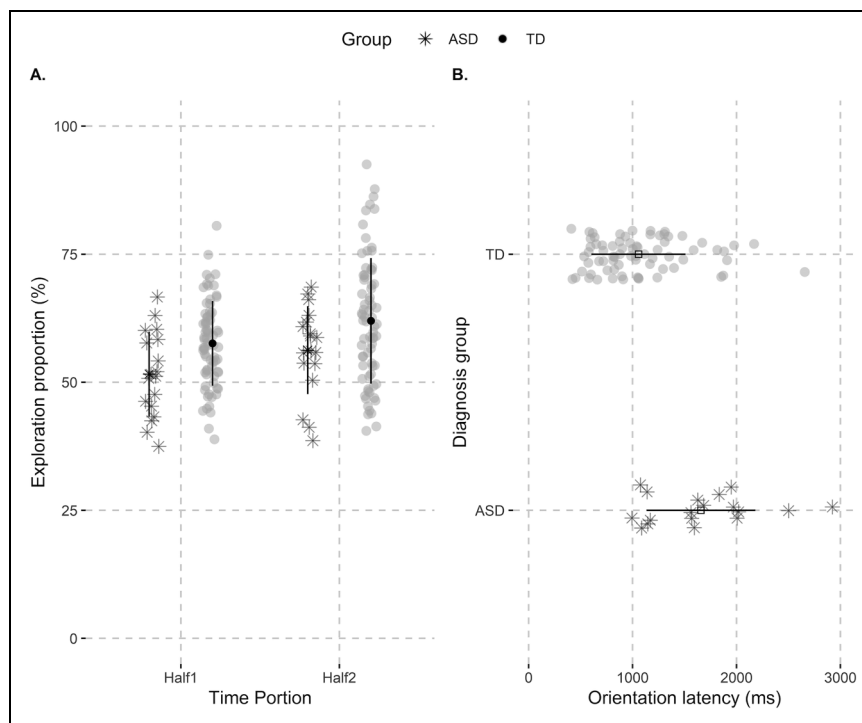


Figure 5. Effect of diagnosis. A) Exploration proportion (Y-axis) according to time portion (X-axis) and diagnostic group (shapes). B) Latency (X-axis) of the FO toward the matching video, according to diagnostic group (Y axis). Latency 0 corresponds to the appearance of the two videos on the screen. For (A) and (B), gray asterisks: autistic participants ($n = 21$); gray dots: non-autistic participants ($n = 79$). Black symbols are the mean per diagnostic group. Error bars are standard deviation.

children had the highest participation and completion rates, and provided more usable data, suggesting that break duration had no influence on performance. On the other hand, children from Group 1, who had long breaks but also longer sessions, had decreased participation rates, suggesting that increasing break durations (of more than a couple of days) led to greater chances that children would not complete the second session. In other words, shortening the task seems to have had more effect on children's performance than break duration. Shortening the task so that it could be administered to autistic children improved visual attention in the non-autistic children. All autistic children included in the study were able to take part in MOSYCAT, except one (who participated in/completed only one of the two sessions), showing the high acceptability of the task by autistic children. Out of all the tasks in our experimental protocol (Table 3), this was the only language task in which so many autistic children were able to participate regardless of their level of required support, NVIQ, or expressive language level. It moreover was the task with the best overall completion rate for the autistic children. The appropriateness of this task for these autistic children is notable since they represented the well-known variability across the autism spectrum: some of them required very substantial support for their level of autism traits, had low NVIQs, and/or had very minimal expressive language

skills. These results suggest that this task meets the imperatives and constraints of acceptability, practical feasibility, and suitability for autistic children that directed its creation. The MOSYCAT protocol allowed for the assessment of morphosyntactic comprehension, confirming it is a powerful tool to establish receptive language profiles in a wide range of children using a passive, undemanding task (Houston-Price et al., 2007; Plesa Skwerer et al., 2016; Tager-Flusberg & Kasari, 2013).

MOSYCAT utilizes eye-tracking and thus although it was designed to assess morphosyntactic abilities, it also clearly relies on some non-linguistic skills (e.g., processing speed, visual attention). Following a number of studies on language in autism (Schaeffer et al., 2023), we noted that many standardized morphosyntactic comprehension tasks, such as the EVALO act-out task that was part of our protocol, rely on non-linguistic skills that make testing many autistic children entirely impossible. As anticipated, we found that nearly every autistic child we studied was able to complete MOSYCAT, whereas the act-out task, which required volitional actions and strong inhibition, among other skills, had to be abandoned because of inordinate testing time and associated behavioral issues. In other words, whereas task demands on traditional morphosyntactic assessment tasks preclude their use for many children, non-linguistic skills tapped by IPL-with-eye-tracking do not preclude its use

for most children. As expected, given the number of autistic children in our sample who were minimally speaking, participation and completion rates for the language production tasks in our protocol (nonword repetition and sentence repetition) were very low (<50%).

We expected a valid tool to evaluate morphosyntactic comprehension to yield parameters with normal or close to normal distributions in both population samples. Data normality and moments of the distribution demonstrated that the parameters measured overall yielded normal distributions that were approximatively symmetric with a reasonable spread of values and without extreme outliers, and no distortion, suggesting that MOSYCAT can be used to assess morphosyntactic abilities in all children. We expected a valid tool to evaluate morphosyntactic comprehension to show sensitivity to chronological age. Consistent with our expectations, our parameters were sensitive to chronological age, reflecting that in the non-autistic population, syntactic comprehension matures with age (Candan et al., 2012).

Our PCA analysis showed that different language profiles were captured by a combination of orientation latency and exploration proportion or exploration accuracy, and that these were linked to standardized evaluation scores. In other words, a combination of parameters appears to be more appropriate to describe children's morphosyntactic comprehension. The PCA and associated correlations supported MOSYCAT's validity and confirmed its clinical suitability, since results from this task correlated with an independently validated measure of morphosyntactic comprehension, the EVALO object manipulation task. It should be noted that orientation accuracy did not explain much of children's performance variability, nor did it increase with age or differ between groups, suggesting that orientation accuracy may measure a skill of children less related to language comprehension.

Finally, as expected, autistic children spent less time exploring the videos than non-autistic children, resulting in a reduced number of valid items used in further statistical analyses. In line with the literature, autistic children at the group level performed less well than the non-autistic children: they spent less time looking at the matching video, and they directed their gaze toward the matching video with a delay. Yet they directed their first gaze toward the matching video with the same accuracy as non-autistic participants. A delayed orientation to the matching video was expected, and therefore, we evaluated performance separately in the first and second half of the target sequence, expecting better performance in the second half (Naigles et al., 2011), especially for autistic children (Su & Naigles, 2019). This hypothesis was confirmed, but, in contrast to Naigles et al. (2011), we found that this effect was not group-specific: both autistic and non-autistic children showed the same pattern. This result suggests that performance improved in the second half for both autistic and non-autistic children. The longer orientation latency in autistic

children might in fact be the result of disrupted anticipated responses, in line with the predictive coding model of autism (Chao et al., 2024; Gomot & Wicker, 2012; Van De Cruys et al., 2014). Delayed orientation may suggest that the familiarization to the videos in the "Neutral Presentation" is less beneficial for autistic children, compared to non-autistic children. Hence, the "Target Sequence" would appear as new to autistic children, thus preventing anticipatory or faster orientation. Non-autistic children, on the other hand, benefited from the neutral presentation, understood that the videos remained on the same side of the screen, and oriented faster to the matching videos. This could be tested by measuring anticipatory orientation and checking whether there was a difference between autistic and non-autistic participants. This could not be done in our study, as orientation latency was measured after the video onset.

In this study, our aim was to make a methodological contribution to the study of morphosyntactic understanding in autism, by developing a tool that effectively assesses morphosyntactic comprehension in young children from across the spectrum, whatever their strengths and challenges may be, including notably children with little or no expressive language. In other words, we have reported here on the feasibility of assessing an autistic child's morphosyntactic comprehension with MOSYCAT. The next step in our study of MOSYCAT was to tackle the heterogeneity inherent in autism (level of required support, intellectual ability, overall language level), especially in young autistic children, through exploration of morphosyntactic comprehension-production profiles generated by MOSYCAT paired with LITMUS-SR, a maximally inclusive task for assessing expressive morphosyntax (see Michel et al., submitted), profiles which could contribute to refinement of clinical profiles for young autistic children. Envisaging full transfer of MOSYCAT to clinical practice warrants further study, notably with the aims of reducing the cost of eye-tracking equipment, extending atomization of data processing and analysis, as well as development of robust profile thresholds.

Limitations

We believe that this study demonstrates that eye-tracking can be used for the assessment of general morphosyntactic comprehension in autistic and young typically developing children. However, we would like to underline some important limitations, which we offer in the spirit of furthering discussion on how experimental tasks such as MOSYCAT can be meaningfully transferred to clinical settings. Throughout the study, we grappled with finding a balance between the desire to develop a robust, controlled task and the desire to assess a wide range of autistic children. First of all, task administration was not entirely standardized due to constraints linked to the clinical environment

in which testing took place and our wish to assess as broad a range of autistic children as possible, including those requiring considerable support. Lighting conditions were not standardized, not only because of differing environments, but also to ensure child comfort. To accommodate children and caregivers, break durations both within and between sessions were variable, which certainly could introduce variability in task engagement and performance. Second, despite our following recommended guidelines to maximize the calibration procedure (Sasson & Elison, 2012), calibration remained low or impossible for a number of children. We therefore used a posteriori calibration to correct eye-gaze trajectories (Franck, 2012) for a number of recordings (7.6% in Group 1, 20% in Group 2 and 40% in the ASD group). Although this procedure appears to be applied regularly, few studies have described it in detail or reported the number of recordings concerned. All of these adjustments were made to increase feasibility, acceptability, and data usability in difficult-to-access children. It remains to be determined how any or all of these adjustments may have affected children's performance, negatively or positively.

Our IPL task was combined with automated processing of eye-tracking data, something that is probably not available in most clinical settings. This raises the important question of whether manual coding of MOSYCAT might increase its clinical usability. Venker et al. (2020) compared manual coding and automatic coding of a task testing the lexical processing of single words. While the two methods yielded similar results, with children (ages 2 to 3) looking significantly more at the target image (e.g., a hat) when it corresponded to the auditory stimulus (*Look at the hat!*) than when it was only semantically related to the auditory stimulus (e.g., *Look at the pants!*), though overall accuracy was significantly larger for manual compared to automatic gaze coding and automatic coding led to more data loss. A further step would be to compare the use of MOSYCAT with automatic (eye-tracker) coding, as in this study of children ages 2 to 5, with (off-line) manual coding. Such a study would be useful in advancing eye-gaze methodology for young autistic children, especially since MOSYCAT differed from the Venker et al. task in several ways. Whereas the Venker et al. study was based on a lexical processing task (i.e., single words), in which trials lasted approximately 6.5 s, MOSYCAT assesses sentence comprehension, with trials lasting 30 s. Finally, MOSYCAT parameters were linked to an independent measure of morphosyntactic comprehension in the non-autistic children, providing preliminary validity of the task. However, we were unable to perform a similar analysis in the autistic children due to these children's inability to engage in the standardized morphosyntax act-out task. A future study could determine MOSYCAT validity in autistic children through comparison with clinical/parental assessment of children's sentence comprehension.

Nonetheless, in our follow-up study (Michel et al., submitted), we used MOSYCAT and an autism-friendly task of morphosyntactic production to identify comprehension-production profiles for autistic children.

Our population sample of young autistic children contained very few girls, and several bilingually exposed children. Numbers did not allow for exploration of either of these variables in relation to MOSYCAT feasibility. These weaknesses are unfortunately common in studies on language in autism, and particularly in young children (see Prévost & Tuller, 2022; Schaeffer et al., 2025).

Conclusion


We developed a short, passive task allowing for assessment of syntactic comprehension based on several syntactic structures of varying complexity: MOSYCAT. Young children, both non-autistic and autistic, including many not yet producing sentences, were able to engage in the task. It moreover was the task with the highest completion rate for autistic children. The parameters measured were normally distributed in both population samples and correlated with standardized language assessment in the non-autistic children, demonstrating the relevance of MOSYCAT for evaluating comprehension in young autistic children, often unable to perform other types of tasks assessing language comprehension. Finally, MOSYCAT assesses morphosyntactic comprehension of several linguistic structures, making it linguistically relevant, and serving potentially to distinguish an autistic child able to understand a very simple structure from a child with no or complete morphosyntactic comprehension.

In sum, MOSYCAT is suitable for young autistic children and may constitute an important contribution to the assessment of language comprehension in autistic children from across the spectrum. As such, it could ultimately serve in speech-language pathology, allowing for the establishment of individual language profiles, even for children who may currently be considered "unassessable," though clinical transfer is clearly not ready at this time.

Acknowledgments

We are grateful to the children who took part in this study, as participants or actors; their parents; staff at the University of Tours autism center; our colleagues from the Autism and Neurodevelopment research team; Zinaida Tamiato, our voice-over specialist; and Letitia Naigles for sharing her IPL expertise with us.

ORCID iD

Marianne Latinus  <https://orcid.org/0000-0002-4406-2962>

Ethics Approval

This research was approved by the Comité d'Éthique de la Recherche Tours-Poitiers the 2021-09-03 and the Comité de Protection des Personnes EST I the 2017-04-20 (2017/23-ID

RCB: 2017-A00756-47; PROSCA). Children were recruited in nursery schools and kindergartens in the region of Tours (France). Autistic children were from the Centre-Val de Loire region and the child psychiatry department of Tours University Hospital.

Consent to Participate

Parents and children were informed of their rights and of all aspects of the experimental design; parental consent was obtained in the form of a signed written consent form. Children's agreement was also sought, verbally, when possible, or non-verbally, based on their behavior.

Consent for Publication

Parental and child consents were obtained for the capture, use, and publication of their images for research and academic purposes. All authors approved the publication of the manuscript.

Author Contributions

LM contributed to methodology, investigation, data curation, formal analysis, visualization, and writing—original draft, review and editing. ML and LT contributed to conceptualization, methodology, funding acquisition, project administration, supervision, and writing—original draft, review and editing. CB contributed to methodology, investigation, and writing—review and editing. FBB, JM, and AR contributed to investigation, resources, and writing—review and editing. All authors read and approved the final manuscript.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: LM was supported by a doctoral scholarship from Region Centre-Val de Loire and by an end-of-doctoral grant from the Fédération Hospitalo-Universitaire Autisme et Troubles du Neurodéveloppement Exac-t.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability and Material

The datasets generated and/or analyzed for the current study are available at: <https://doi.org/10.6084/m9.figshare.32268348>.

Code Availability

Analysis codes are available at: <https://doi.org/10.6084/m9.figshare.32268348>.

Supplemental material

Supplemental material for this article is available online.

Notes

1. Given the age range, and based on information collected from families, the bilingual children all had either simultaneous bilingual exposure (most), or very early sequential exposure. All autistic children were receiving services (including speech-language therapy) exclusively in French. The number of bilingual children in each group did not allow for investigation of whether task feasibility or outcome was related to bilingualism. For a review of bilingualism in language acquisition in autistic children, see Prévost and Tuller, 2022.
2. The gender imbalance in the ASD Group reflects the fact that ASD continues to be diagnosed later in girls than in boys, if not missed altogether (Lockwood Estrin et al., 2021), and thus diagnosed children in this age range are predominantly boys. As is typical for early-diagnosed girls, all four were non-speaking and had associated intellectual disability; two of them were bilingually exposed. The limited number of girls precluded examination of gender in this study.
3. Means were calculated over scores of all participants assessed with the ADOS-2 ($n = 19$), either Toddler Module, Module 1, or Module 2.
4. The approximate IQ range is, for mild ID, 50–69, for moderate ID, 36–49, and for severe ID, 20–35 (DSM-IV).

References

- Anderson, D. K., Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., Welch, K., & Pickles, A. (2007). Patterns of growth in verbal abilities among children with autism spectrum disorder. *Journal of Consulting and Clinical Psychology, 75*(4), 594–604. <https://doi.org/10.1037/0022-006X.75.4.594>
- Candan, A., Küntay, A. C., Yeh, Y., Cheung, H., Wagner, L., & Naigles, L. R. (2012). Language and age effects in children's processing of word order. *Cognitive Development, 27*(3), 205–221. <https://doi.org/10.1016/j.cogdev.2011.12.001>
- Chao, Z. C., Komatsu, M., Matsumoto, M., Iijima, K., Nakagaki, K., & Ichinohe, N. (2024). Erroneous predictive coding across brain hierarchies in a non-human primate model of autism spectrum disorder. *Communications Biology, 7*(1), 1–16. <https://doi.org/10.1038/s42003-024-06545-3>
- Chen, Y., Siles, B., & Tager-Flusberg, H. (2024). Receptive language and receptive-expressive discrepancy in minimally verbal autistic children and adolescents. *Autism Research, 17*(2), 381–394. <https://doi.org/10.1002/aur.3079>
- Coquet, F., Ferrand, P., & Roustit, J. (2009). *EVALO 2-6 Une nouvelle approche méthodologique de l'évaluation*. Ortho Edition.
- Esler, A. N., Bal, V. H., Guthrie, W., Wetherby, A., Weismer, S. E., & Lord, C. (2015). The autism diagnostic observation schedule, toddler module: Standardized severity scores. *Journal of Autism and Developmental Disorders, 45*(9), 2704–2720. <https://doi.org/10.1007/s10803-015-2432-7>
- Ferré, S., dos Santos, C., & Tuller, L. (2022). *LITMUS-QU-NWR-Toddlers*. Ms., University of Tours.

- Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the development of social attention using free-viewing. *Infancy, 17*(4), 355–375. <https://doi.org/10.1111/j.1532-7078.2011.00086.x>
- Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14*(1), 23–45. <https://doi.org/10.1017/S030500090001271X>
- Gomot, M., & Wicker, B. (2012). A challenging, unpredictable world for people with Autism Spectrum disorder. *International Journal of Psychophysiology, 83*(2), 240–247. <https://doi.org/10.1016/j.ijpsycho.2011.09.017>
- Goodwin, A., Fein, D., & Naigles, L. R. (2012). Comprehension of wh-questions precedes their production in typical development and autism spectrum disorders. *Autism Research, 5*(2), 109–123. <https://doi.org/10.1002/aur.1220>
- Grandgeorge, M., Hausberger, M., Tordjman, S., Deleau, M., Lazard, A., & Lemonnier, E. (2009). Environmental factors influence language development in children with Autism Spectrum disorders. *PLOS ONE, 4*(4), e4683. <https://doi.org/10.1371/journal.pone.0004683>
- Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *Journal of Child Language, 34*(4), 701–724. <https://doi.org/10.1017/S0305000907008124>
- Hus, V., Gotham, K., & Lord, C. (2014). Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. *Journal of Autism and Developmental Disorders, 44*(10), 2400–2412. <https://doi.org/10.1007/s10803-012-1719-1>
- Jyotishi, M., Fein, D., & Naigles, L. (2017). Investigating the grammatical and pragmatic origins of wh-questions in children with autism spectrum disorders. *Frontiers in Psychology, 8*, 319. <https://doi.org/10.3389/fpsyg.2017.00319>
- Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research, 6*(6), 479–493.
- Kern, S., & Gayraud, F. (2010). *Inventaire Français du Développement Communicatif (IFDC)* (Cigale). de la Cigale.
- Kovarski, K., Siwiaszczyk, M., Malvy, J., Batty, M., & Latinus, M. (2019). Faster eye movements in children with autism spectrum disorder. *Autism Research, 12*(2), 212–224. <https://doi.org/10.1002/aur.2054>
- Le Couteur, A., Lord, C., & Rutter, M. (2003). The Autism Diagnostic Interview-Revised (ADI-R). *Western Psychological Services*.
- Leppänen, J. M., Forssman, L., Kaatiala, J., Yrttiaho, S., & Wass, S. (2015). Widely applicable MATLAB routines for automated analysis of saccadic reaction times. *Behavior Research Methods, 47*(2), 538–548. <https://doi.org/10.3758/s13428-014-0473-z>
- Lockwood Estrin, G., Milner, V., Spain, D., Happé, F., & Colvert, E. (2021). Barriers to autism spectrum disorder diagnosis for young women and girls: A systematic review. *Review journal of autism and developmental disorders, 8*(4), 454–470.
- Lord, C., Rutter, M., Dilavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule, Second Edition*. Western Psychological Services.
- Muller, K., & Brady, N. (2016). Assessing early receptive language skills in children with ASD. *Perspectives of the ASHA Special Interest Groups, 1*(1), 12–19. <https://doi.org/10.1044/persp1.SIG1.12>
- Naigles, L. R., & Fein, D. (2017). Looking through their eyes: Tracking early language comprehension in ASD. In *Innovative investigations of language in autism spectrum disorder* (pp. 49–69). American Psychological Association. <https://doi.org/10.1037/15964-004>
- Naigles, L. R., Kelty, E., Jaffery, R., & Fein, D. (2011). Abstractness and continuity in the syntactic development of young children with autism. *Autism Research, 4*(6), 422–437. <https://doi.org/10.1002/aur.223>
- Naigles, L. R., & Tovar, A. T. (2012). Portable intermodal preferential looking (IPL): Investigating language comprehension in typically developing toddlers and young children with autism. *Journal of Visualized Experiments, 70*, 4331. <https://doi.org/10.3791/4331>
- Plesa Skwerer, D., Jordan, S. E., Brukilacchio, B. H., & Tager-Flusberg, H. (2016). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with autism spectrum disorders. *Autism: The International Journal of Research and Practice, 20*(5), 591–604. <https://doi.org/10.1177/1362361315600146>
- Prevost, P., & Tuller, L. (2022). Bilingual language development in autism. *Linguistic Approaches to Bilingualism, 12*(1), 1–32.
- Sasson, N. J., & Elison, J. T. (2012). Eye tracking young children with autism. *Journal of Visualized Experiments, 61*, e3675. <https://doi.org/10.3791/3675> (2012)
- Schaeffer, J., Abd El-Raziq, M., Castroviejo, E., Durrleman, S., Ferré, S., Grama, I., Hendriks, P., Kissine, M., Manenti, M., Marinis, T., Meir, N., Novogrodsky, R., Perovic, A., Panzeri, F., Silleresi, S., Sukenik, N., Vicente, A., Zebib, R., Prévost, P., & Tuller, L. (2023). Language in autism: Domains, profiles and co-occurring conditions. *Journal of Neural Transmission, 130*, 433–457. <https://doi.org/10.1007/s00702-023-02592-y>
- Schaeffer, J., Novogrodsky, R., Perovic, A., Prévost, P., & Tuller, L. (Eds.). (2025). *Language in autism*. John Wiley & Sons. ISBN: 978-1-394-18038-7
- Su, Y. I. & Naigles, L. R. (2019). Online processing of subject-verb-object order in a diverse sample of Mandarin-exposed preschool children with Autism Spectrum disorder. *Autism Research, 12*(12), 1829–1844. <https://doi.org/10.1002/aur.2190>
- Swensen, L. D., Kelley, E., Fein, D., & Naigles, L. R. (2007). Processes of language acquisition in children with autism: Evidence from preferential looking. *Child Development, 78*(2), 542–557. <https://doi.org/10.1111/j.1467-8624.2007.01022.x>
- Tager-Flusberg, H. (1999). The challenge of studying language development in children with autism. In L. Menn & N. B. Ratner (Eds.), *Methods for studying language production* (pp. 313–332). Psychology Press.

- Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research, 6*(6), 468–478. <https://doi.org/10.1002/aur.1329>
- Tomasello, M., & Mervis, C. B. (1994). The instrument is great, but measuring comprehension is still a problem. *Monographs of the Society for Research in Child Development, 59*(5), 174–179. <https://doi.org/10.1111/j.1540-5834.1994.tb00186.x>
- Tovar, A. T., Fein, D., & Naigles, L. R. (2015). Grammatical aspect is a strength in the language comprehension of young children with Autism Spectrum disorder. *Journal of Speech, Language, and Hearing Research, 58*(2), 301–310. https://doi.org/10.1044/2014_JSLHR-L-13-0257
- Tuller, L., Ferré, S., Prévost, P., Barthez, M. A., Malvy, J., & Bonnet-Brilhault, F. (2017). The effect of computational complexity on the acquisition of French by children with ASD. In L. Naigles (Ed.), *Innovative investigations of language in autism spectrum disorder* (pp. 115–140). Walter de Gruyter GmbH and Co KG.
- Tuller, L., & Prévost, P. (2022). *LITMUS-SR-FR-Toddlers*. Ms., University of Tours.
- Van De Cruys, S., Evers, K., Van Der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in Autism. *Psychological Review, 121*(4), 649–675. <https://doi.org/10.1037/a0037665>
- Venker, C. E., Haebig, E., Edwards, J., Saffran, J. R., & Ellis Weismer, S. (2016). Brief report: Early lexical comprehension in young children with ASD: Comparing eye-gaze methodology and parent report. *Journal of Autism and Developmental Disorders, 46*(6), 2260–2266. <https://doi.org/10.1007/s10803-016-2747-z>
- Venker, C. E., Pomper, R., Mahr, T., Edwards, J., Saffran, J., & Ellis Weismer, S. (2020). Comparing automatic eye tracking and manual gaze coding methods in young children with autism spectrum disorder. *Autism Research, 13*(2), 271–283. <https://doi.org/10.1002/aur.2225>
- Wagner, L., Swensen, L. D., & Naigles, L. R. (2009). Children's early productivity with verbal morphology. *Cognitive Development, 24*(3), 223–239. <https://doi.org/10.1016/j.cogdev.2009.05.001>
- Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence (4th ed.)*. Pearson.