124 (2025) 202589

Contents lists available at ScienceDirect

Research in Autism

journal homepage: www.elsevier.com/locate/rasd

Investigating sex differences in narrative production of autistic pre-adolescents and adolescents

Marie Belenger [a,*], Charlotte Dumont [a], Philippine Geelhand [a], Mikhail Kissine [a,b,c]

[a] ACTE, LaDisco and ULB Neuroscience Institute, Université libre de Bruxelles, Avenue F. D. Roosevelt, 50/175, 1050 Brussels, Belgium
[b] Department of Philosophy, Classics, History of Art and Ideas, University of Oslo, Norway
[c] Department of Linguistics and Comparative Cultural Studies, 'Ca Foscari University of Venice, Italy

ARTICLE INFO

Keywords:
Autism
Adolescents
Narratives
Sex differences
Coherence
Discourse analysis

ABSTRACT

Autism is less frequently diagnosed in females, and autistic females are often diagnosed later in life than males. The sex imbalance in autism could be partly due to a communicative advantage in autistic females. To better understand sex differences in language of autistic individuals in late-childhood and adolescence, we compared narrative coherence of autistic females, autistic males, non-autistic females and non-autistic males. Narrative production was elicited from a total of 113 participants (mean age 12.32), based on a wordless picture book. Relying on a thorough coding scheme, we analysed the following categories: story grammar, connectives, references to characters and internal state language. Independently of sex, autistic individuals produced less coherent narratives than non-autistic individuals. Narratives by autistic adolescents included more comments unrelated to the story and more image descriptions, less causal connectives, less internal state language and fewer mentions of main characters. Autistic participants also used more indefinite expressions to refer to story characters than their non-autistic peers. No significant sex differences were found between autistic males and females. Based on a sex-balanced sample, this study confirms prior results on narrative production by autistic individuals and provides new insights into referential expression choices. Importantly, this study finds no evidence for lower atypicality of communicative profiles of autistic females, as measured by a narrative task.

Preliminary note

In this study, we focused on participants' sex, i.e., assigned sex at birth, as reported by their caregivers, rather than on their gender identity. This decision was taken because our sample includes children and adolescents who, in most cases, did not have the opportunity to self-report their gender. We recognize that this approach does not account for individuals who identify as transgender, non-binary, or gender diverse, and we acknowledge the limitations inherent in this method.

Introduction

Autism is more frequently diagnosed in males than in females, with sex ratios hovering around 4:1 or 3:1 (Fombonne, 2009; Loomes

* Correspondence to: ACTE, Université libre de Bruxelles, Avenue F. D. Roosevelt, 50/175, 1050 Brussels, Belgium.
E-mail address: Marie.Belenger@ulb.be (M. Belenger).

https://doi.org/10.1016/j.reia.2025.202589
Received 22 August 2024; Received in revised form 18 November 2024; Accepted 30 March 2025
Available online 10 April 2025

et al., 2017). There is also growing evidence that autistic females are less likely to be identified, especially when they do not display cooccurring intellectual disabilities (Dworzynski et al., 2012; Loomes et al., 2017). This sex imbalance may be due, in part, to autism manifesting differently in females compared to males (Hull et al., 2020). Several candidate aspects of autism presentation that may vary across sexes have been singled out in the literature (Beggiato et al., 2017; Cruz et al., 2024; Harrop et al., 2018; Rynkiewicz et al., 2016; Sedgewick et al., 2016), the most relevant of which for our purposes being language and communication.

Previous findings on potential effects of sex on language and communication in autism are, in fact, somewhat mixed. A 2014 meta-analysis found no difference in the domain of communication between autistic males and females, but the authors also noted that females with cognitive abilities within the typical range were underrepresented in the studies they analysed (Van Wijngaarden-Cremers et al., 2014). Moreover, standardized language assessments often used in these studies are probably insufficiently sensitive to reveal sex differences in cognitively able individuals (Sturrock, Yau et al., 2019). Such tests primarily assess language at the sentence level and do not capture the dynamic of discourse-level language use in interactional contexts. When more complex aspects of language, such as semantics, pragmatics and narrative abilities, are examined, sex differences do emerge (den Hartog et al., 2023). For example, autistic females tend to perform better than autistic males on pragmatic and semantic tasks (Sturrock, Adams et al., 2021), produce more social words (Cola et al., 2022), use more plural pronouns to talk about social groups (Song et al., 2021), and use filled pauses differently (Parish-Morris et al., 2017). Taken together, these results suggest that autistic females have fewer communication difficulties. This, in turn, may significantly impact how they are perceived, leading for example, to better first impressions during naturalistic conversations for autistic females than autistic males (Cola et al., 2020) but also likely lead to under-detection, contributing to the imbalanced sex ratio in autism diagnoses.

Despite this potential 'communicative advantage', autistic females still experience communication and social difficulties compared to non-autistic individuals (Burton et al., 2020; den Hartog et al., 2023). Furthermore, while fewer language and communication difficulties are being observed for autistic females, their perceived difficulties, as reported by caregivers or by the individual themselves, remain as significant as for autistic males (Sturrock et al., 2019; Sturrock et al., 2021). This discrepancy between reported and observed difficulties could be due to higher social expectations for females (Holtmann et al., 2007), but it may also demonstrate that self- or caregiver reports are more accurate than language assessments when it comes to identifying subtle communicative difficulties (Sturrock, Adams et al., 2021). Additionally, autistic individuals with cognitive abilities within the typical range tend to perform better in structured tasks, with explicit and clear instructions (Volden & Phillips, 2010). Consequently, language assessment in controlled setting may not indicate their conversational flexibility nor to what extent they may experience real-life communication difficulties (Volden et al., 2017).

In sum, there is a need for a deeper investigation of the language abilities of autistic females to understand the gap between reported and observed measures. This will help to determine whether autistic females truly exhibit better communicational abilities than males, as tends to be observed in research settings, or whether true communication abilities of autistic females are comparable to those of autistic males.

One interesting approach for evaluating communication abilities is the use of narratives, which provide a way to collect language data that are both fluid and more naturalistic and easily comparable between groups of participants (Stirling et al., 2014). Narratives are a complex form of discourse, characterized by recounting a series of interconnected actions or events that unfold over time (Stirling et al., 2014), and rely on a range of linguistic, cognitive, and pragmatic abilities (Norbury et al., 2014). Narratives are essential to human experience: most of our daily communication occurs in the form of narratives (Lê et al., 2011), our understanding of the world is shaped as narrations, and narratives are used to communicate with others about the world (Bruner, 1991). Accordingly, the ability to produce a coherent narrative is crucial to effective communication, self-advocacy, social development, and peer engagement as well as academic success (Dean et al., 2013; Petersen, 2011; Sturrock et al., 2021). Given their naturalistic qualities and the central role they play in everyday communication, narratives may be particularly useful in revealing subtle sex-related differences in language and pragmatics among autistic individuals, which may otherwise go unnoticed in traditional assessments.

Let us turn now to the kind of dependent variables that can be drawn from narratives. One widespread measure of narrative coherence is story grammar, which targets the presence of key story elements (Stein & Glenn, 1979). However, simply counting the main events fails to fully capture the narrative organization: a protracted but poorly organized narrative may still contain more key story elements than a short, but well-structured one. This highlights the importance of evaluating not just the quantity, but also the quality and organization of the narrative elements (Harvey et al., 2023). Additionally, assessing cohesion is crucial to assess whether there is a meaningful link between sentences (Norbury & Bishop, 2003) and if the referents in the narrative can be identified unambiguously (Liles, 1985). Story grammar, along with the quality, organization, and cohesive adequacy of the narrative, all contribute to narrative coherence.

Overall, the narratives of autistic children tend to exhibit reduced coherence, as compared to those of their non-autistic peers. Autistic children make less use of the gist of the story, i.e. the key story grammar elements, to structure their narratives (Baixauli et al., 2016). In older children or adults, however, some studies found no differences between autistic and non-autistic participants in the inclusion of story grammar elements (Geelhand et al., 2020; Kauschke et al., 2016). That said, even when autistic adults adhere to the gist of the story, they tend to include more extraneous comments about the story or about the task itself (Geelhand et al., 2020). Moreover, the narratives of autistic individuals tend to be less causally connected than those of non-autistic ones and to feature more ambiguous pronouns (Baixauli et al., 2016; Geelhand et al., 2020).

In addition to using accurate references, it is also important to express the point of view of the characters in the story by including Internal State Language elements, viz. linguistic descriptions of feelings, desires, beliefs, intentions, and other internal states (Stirling et al., 2014). References to the story protagonists' internal states help the listener understand who is doing what to whom and render the narrative more coherent (Harvey et al., 2023). Overall, narratives of autistic individuals tend to contain less internal state language

elements than those of non-autistic individuals, particularly those of autistic individuals with higher cognitive abilities (Baixauli et al., 2016). Furthermore, autistic individuals tend to provide fewer causal explanations for the emotions or mental states they describe using Internal State Language (Beaumont & Newcombe, 2006; Losh & Capps, 2003).

To recap, analyzing narratives in autism may bring to light difficulties in discourse coherence, which may be obfuscated by coarser standardized tests but also reflect challenges in real-world communication. Assessing narrative abilities targets linguistic, cognitive, and pragmatic abilities, and thus could be particularly useful to unveil subtle communication difficulties and potentially shedding light on sex differences in autism. To date, only four articles have directly examined sex differences in oral narratives of autistic children and adolescents, with mixed results (Boorse et al., 2019; Conlon et al., 2019; Kauschke et al., 2016; Sturrock, Yau et al., 2019). These studies are not directly comparable because of different methodological choices, i.e., elicitation tasks, discourse analysis, sample sizes, and comparison groups, but Table 1 provides a summary of different results. Overall, these four studies suggest that narratives by autistic females may be more similar to those by non-autistic individuals than those by autistic males. It seems that autistic females may produce more coherent narratives than autistic males by including more salient elements (Conlon et al., 2019) and producing more terms related to Internal State Language (Boorse et al., 2019; Conlon et al., 2019; Kauschke et al., 2016). Seminal as they are, these studies have also several limitations, the chief of which is a small number of participants. Furthermore, not all studies conducted a systematic analysis of narrative coherence (Boorse et al., 2019; Sturrock, Adams et al., 2021), and there was a lack of a consistently applied coding guide. Finally, two of these studies did not include a comparison group of non-autistic individuals, making it difficult to fully contextualize the results (Conlon et al., 2019; Kauschke et al., 2016).

Current study

This study addresses the gaps delineated in the foregoing by (i) systematically analyzing narrative coherence in (ii) autistic females and males (ii) in comparison to their non-autistic female and male peers. We assessed narrative coherence in autistic and non-autistic

Table 1
Summary of previous studies on sex differences in narrative production of autistic individuals.

Reference	Sample	Elicitation task	Measures	Key findings
Kauschke et al. (2016)	11 autistic females, 11 autistic males, and 11 non-autistic females (age 8 −19).	Story telling from a wordless picture book: *Tuesday* (Wiesner, 1991). 32 pages.	Detailed discourse analysis: Story length; References to story's character; Orientation (space and time); Core events); Irrelevant details; Connectives; Evaluative devices; ISL.	Overall, comparable narrative abilities between groups except for ISL. Autistic females produced more ISL (Cohen's $d = 0.84$) and more causal explanation for ISL (Cohen's $d = -0.45$) than autistic males. Non-autistic females produced more emotion terms than autistic females (Cohen's $d = -1.45$)
Conlon et al. (2019)	13 autistic females and 13 autistic males (age 8 −8:11).	Narrative assessment instrument: Expression, Reception, Recall of Narrative Instrument (ERRNI;Bishop, 2004). 15 pictures.	Scores at ERRNI: Inclusion of relevant information (initial & recall); comprehension questions. Detailed discourse analysis: Pragmatic index; Semantic index; Formulation index. Semantic enhancement analysis: ISL (emotion and mental state words, intentional utterances); Adverb; Character speech.	Autistic females included more salient story elements than autistic males (Glass' $\Delta = 1.47$). Autistic females semantically enhanced more their story than autistic males (Glass' $\Delta = 2.08$) with more utterances referring to intentions (Glass' $\Delta = 1.49$).
Sturrock, Yau et al. (2019)	13 autistic females, 13 autistic males, 13 non-autistic females, and 13 non-autistic males (age 8:11 −11:6)	Story telling from a wordless picture book: *A boy, a dog and a frog* (Mayer, 2003). 32 pages.	Story length; Temporal and causal connectives; Coherence of the narrative (qualitative evaluation); ISL (emotion words).	Autistic females did not differ from autistic males. Autistic participants produced less causal connectives ($\eta^2 = 0.124$) and a narrower range of temporal connectives ($\eta^2 = 0.103$) and causal connectives ($\eta^2 = 0.175$). Autistic participants produced less coherent narratives ($\eta^2 = 0.260$). No group, sex or interaction effect was found for the production of emotion words.
Boorse et al. (2019)	21 autistic females, 41 autistic males, 19 non-autistic females, and 21 non-autistic males (age 5:8 −15:2)	Elicitation task from the ADOS−2 module 3: *The fisherman and the cat* (Lord et al., 2012). 6 pictures.	Number of nouns and number of cognitive process words.	No significant diagnosis by sex interaction on number of nouns. Autistic participants produced more nouns than non-autistic participants (SMD = 1.08). Females produced more cognitive process words than males (SMD = 0.76). Autistic females produced more cognitive process words than autistic males (SMD = 1.07).

This table presents previous studies investigating sex differences in narrative production of autistic individuals. For each reference, the sample, the task, the measures as well as the key findings are provided.

children and adolescents, matched on sex. Based on the literature, we expect differences in the coherence of narrative between autistic and non-autistic participants, including differences in the story grammar as well as in the inclusion of other events (e.g., extraneous comments), causal connectives, references to characters and Internal State Language.

Furthermore, we expect the narratives of autistic females to differ from those of autistic males, as well as from those of non-autistic males and females. Specifically, we anticipate that autistic females will produce more coherent narratives than autistic males, but less coherent than their non-autistic peers. Based on previous studies, we expect autistic females to: (i) include more main events than autistic males, (ii) include more internal state language than autistic males with more physiological, event modalities and emotions terms as well as more cognitive process words for autistic females, and (iii) include more causal explanation for Internal State Language than autistic males.

Method

Participants

113 French-speaking participants aged between 9 to 16 years (M= 12.32; SD= 2.29) took part in this study: 54 autistics and 59 non-autistics. Participants were divided into four groups according to sex and diagnosis: autistic females (n = 24), autistic males (n = 30), non-autistic females (n = 30) and non-autistic males (n = 29). Autistic participants were recruited through the lab's volunteer database, schools, associations, or social media. Non-autistic participants were recruited through schools, word of mouth and social media. Recruitment initially took place in the French-speaking part of Belgium; but due to challenges in recruiting verbally fluent autistic female adolescents, recruitment was extended to France.

Inclusion criteria were: to be aged between 9 and 16 at the time of recruitment, to have non-verbal IQ above 70, and to have sufficient verbal fluency to perform a narrative task. All autistic participants had a formal autism diagnosis previously issued by a specialized multi-disciplinary team. For non-autistic participants, autism was ruled out using the Lifetime form of the Social Communication Questionnaire, filled in by participants' parents (SCQ; Rutter et al., 2003).

Procedure

This study was approved by the Ethics Committee of Erasme Hospital and written informed consent was obtained from each of the study's participants and their parents.

This study was conducted over two sessions, either at participants' homes, participants' schools or in the lab. During the first session, after a short get-to know, participants completed a narrative task. Next, they completed a behavioral task for another study. Lastly, the CELF-5 was administrated. Total time of this session was around 1 h. During the second session, participants completed two behavioral tasks for another study. Then, the WISC-V was administered. The total duration for this session was around 90 min.

Psychometric measures & preliminary questionnaires

Cognitive abilities were assessed using the fifth version of the Wechsler Intelligence Scale for Children (WISC-V; Wechsler, 2016). Core language abilities were assessed using the French version of the Clinical Evaluation of Language Fundamental–Fifth version (CELF-5; Wiig et al., 2019).

Along with the SCQ, parents completed three additional questionnaires. First, a questionnaire collecting general information including the socio-economic status of participants, based on the Revised Family Affluence Scale (FAS III; Torsheim et al., 2016). This questionnaire is administered to all participants taking part in our lab studies. Second, the Children's Communication Checklist-2 (CCC-2; Norbury et al., 2004), a 70-item questionnaire that assesses aspects of communication that are often overlooked in traditional language assessments, including pragmatic elements. Third, the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997), a 25-item behavioral questionnaire screening for difficulties (internalized and externalized), as well as prosocial behavior. As the CCC-2 and the SDQ scores were initially collected for another study, we report group comparisons in participant's characteristics for completeness, but these scores were not included in analysis of narrative.

Narrative assessment material & Data collection.

Frog where are you? (Mayer, 1969), a 32-pages wordless picture book, was used to elicit a narrative from each participant. Narratives were video-taped using a Sony DSC-RX100 III video camera and were then converted to audio files. Due to technical issues with the video-recorder, two recordings were lost. The total amount of recording was 7 h, 23 min, and 55 s. Mean length of recording per participant was 3 min, 56 s (SD= 1 min, 35 s). The experimenter (the first author) always told the first page of the story and then remained silent until the end of the story. In some cases, when needed, the experimenter provided some encouragement to pursue the story in the form of facial expression or backchannelling.

Narrative measures & data coding

Data preparation and coding procedure was largely based on the procedure described in Geelhand et al. (2020). All recordings were analyzed in Praat (Boersma & Weenink, 2022). Trained MA students in linguistics orthographically transcribed the audio recordings and the narratives were then annotated. Some categories were manually coded, while others were automatically extracted (see Supp. Mat. 1 for the complete coding manual).

Narrative coherence was evaluated by analyzing the transcribed stories according to the story grammar and the other events included, connectives, references to characters, and Internal State Language (ISL). Table 2 presents an overview of the examined variables and examples.

Events included in the story

Our coding scheme included a measure of story grammar, using counts of main events. The different elements coded for this category are adapted from Ma et al. (2017) and the script of the story is available on Salt Software (Miller & Nockerts, 2023). The story setting was not included in the analysis, as it was always introduced by the experimenter. Table 3 summarizes the different story elements coded for this category.

Five other categories were added to capture whether and how participants added elements that are not part of the story scheme: *dog events, additional story events, image descriptions, additional elements, extraneous comments* (see Table 2 for examples). As there is a parallel story in which the main protagonist is the dog (Ma et al., 2017), events related to the dog are coded as *Dog Events*. *Additional story events* encompass elements that can be found in the book but not directly related to the central plot (i.e. cannot be found in the script). *Image description* was assigned when participants provided descriptions of book elements without referencing actions or psychological states. Finally, two additional categories were coded to capture disruptive elements hypothesized to interrupt narrative flow and coherence: *Additional element* was coded when participants referenced events or elements that were not present in the book, and *Extraneous comment* was coded when participants made comments not related to the story or comments related to the narrative task, rather than to the depicted events.

Connectives

Connectives are discourse structuring devices, used to link utterances together in various forms. A list of connectives was compiled based on those annotated in Geelhand et al. (2020). Python 3.12.0 (Python Software Foundation, 2023), with the os (os: Python Software Foundation, 2023) and spaCy packages (spaCy: Explosion AI, 2023), was used to automatically identify connectives from the transcription. To validate accuracy, a subset of ten audio samples was manually coded and compared with the automated results, updating the connective list as needed. This process was repeated iteratively with five additional samples manually coded and compared after each cycle, for a total of seven cycles. By the end of the seventh iteration, all connectives identified manually were accurately extracted by the automated coding. The final updated list of connectives is available in the Supp. Mat 1.

Referential expressions

Referential expressions were coded across three dimensions: (i) the accurate identification of the character being referenced, (ii) the selection of referential expression, and (iii) the presence or absence of ambiguity in the reference to the character. There are three

Table 2
Variables for the analysis of narratives.

Categories Subcategories	Examples / Descriptions
Events included	
Main events	Any events that is part of the narrative scheme
Dog events	Le chien cherche la grenouille derrière la pierre (*The dog is looking for the frog behind the rock*) – Autistic male participant
Additional story events	il se met assis dans l'eau et il voit un trou (*He sits down in the water and sees a hole)* – Non-autistic female participant
Image description	La forêt était très dense *(The forest was very dense)* – Non-autistic male participant
Additional elements	Mais sa maman lui disait qu'il devait aller au lit *(But his mom told him he had to go to bed)* – Non-autistic female participant
Extraneous comments	On dirait le chien de ma grand-mère *(He looks like my grand-mother's dog)* – Non-autistic female participant Je raconte bien hein (*I am a good storyteller, right)* – Autistic male participant
Connectives	
Additive	Et (*and*)
Temporal	Après (*after*)
Causal	Parce que (*because*)
Contrastive	Mais (*but*)
References	
Main characters	The boy, the dog, the frog
Definite referential expression	Tom (*Tom*), la petite grenouille (*the little frog*), son animal de compagnie (*his pet*), Cette grenouille (*this frog)*
Indefinite referential expression	Un petit garçon (*a little boy*), Certains (*some*), Quelqu'un (*somebody*)
Pronominal referential expression	Ils (*they*), elle (*she*)
Ambiguous pronoun	In case of competing referents for a pronominal referential expression
Internal State Language	
Emotion	Terms referring to emotions / behaviors linked to emotions. Être triste (*to be sad*), malheureux (*unhappy*)
Cognition	Terms referring to mental / cognitive states, knowledges, beliefs and memories. Decider (*decide*), r ê ver (*dream*)
Physiology	Terms referring to biological / physiological perception and sensations. Voir (*see*), entendre (*hear*).
Modality	Terms referring to intention, obligations, and volitions. Vouloir (*want*), devoir (*must*).
Intention	Terms implying an intention but not directly referring to intention. S'enfuir (*to escape*), se cacher (*to hide*).
Explanation	Verbalization of the causes / consequences of ISL. Ils sont contents parce que Lila a eu des enfants (*they are happy because Lila had children*).

Each coding category and subcategory is illustrated with examples from collected corpus.

Table 3
Main story elements of the wordless picture book Frog where are you?.

Story elements	Description
Initiating events	At night, while the boy and the dog are asleep, the frog escapes.
	The boy and the dog wake up and see that the frog is gone.
Sequence of events	They are looking for the frog.
	The boy calls the frog through the window
	The boy picks up the dog.
	They keep looking outside / in the forest. The boy calls the frog.
	The boy calls down a hole.
	A gopher bites the boy.
	The boy climbs a tree and look into a tree hole.
	An owl comes out of the tree hole causing the fall of the boy.
	The owl chases the boy all the way to a large rock.
	The boy climbs the rock, hold on some branches, call for the frog.
	Branches were actually deer antlers. The boy is on the deer head.
	The deer starts running with the boy on his head.
	The deer stops and drops the boy on the edge of a cliff, they fall over the edge of the cliff.
	The boy and the dog land into water.
Decisive events	They hear a sound.
	The boy tells the dog to be quiet.
	They look behind a log.
Conclusion	They find the pet frog. The frog is with another frog. And some baby frogs.
Coda	The boy waves goodbye and goes back home with a baby frog

Script of the main events of the narrative task.

main characters in *Frog where are you* coded as follow: (1) the boy, (2) the dog, and (3) the frog. All secondary characters are collectively coded as (4) other. Choices of referential expressions were adapted from Geelhand et al. (2020).

Ambiguous pronoun use was coded in instances where competing referents arose for pronominal definite expression. Referent becomes ambiguous if the last character mentioned in a non-pronominal expression was not the same one. For example, in the sentence "le petit garçon et le chien aperçoivent un cerf qui **le** prend avec ses bois'' (*the little boy and the dog spot a deer who take **him** with his antlers*) "le'' can either refer to "le petit garçon'' or "le chien'' as in French, we use "il" to refer both to animate subjects (including animals) as well as inanimate objects.

Internal State Language (ISL)

The coding categories ISL consists of references made to the internal states of the story's characters. The ISL categories, listed in Table 2, were adapted from previous narrative studies (Geelhand et al., 2020; Kauschke et al., 2016; Norbury et al., 2014).

The coding of ISL is often highly subjective, making it sometimes difficult to determine which words belong to which categories. To bring more objectivity into coding, we decided to create a dictionary. First, we compiled a list of ISL, based on those annotated in Geelhand et al. (2020). This list was extracted using Praat scripts (Boersma & Weenink, 2022). To address the frequent overlap and ambiguity in the *Cognition* and *Modality* categories, we decided to refine our approach with the introduction the category, *Intention* (adapted from Norbury et al., 2014), which includes verbs implying a character's intention without explicitly stating it. We also made the decision not to code the category *Evaluation* as we found it challenging to distinguish between the narrator's evaluation and the character's evaluation, leading to potential ambiguity in our coding process.

To complete the ISL dictionary, five audios were initially coded independently by a trained MA student in linguistics and the first author to identify and add any missing terms. Any disagreements were discussed and resolved. Five additional audios were then independently coded using the same procedure, confirming that no new terms or synonym were found. The remaining audios were subsequently coded. The ISL dictionary can be found in the Supp. Mat.1.

Inter-rater agreement

Reliability of manual coding was measured by double-coding 10 % of the transcripts. The interrater agreement was assessed using

Table 4
Cohen's κ per coding category.

Category	κ	*Interpretation* (Landis & Koch, 1977)
Story structure & other elements	.79	*Substantial agreement*
Connectives	NA - Automated extraction	
Referential expressions		
Identification of character	.95	*almost perfect agreement*
Choice of referential expression	.86	*almost perfect agreement*
Ambiguous pronoun use	.81	*almost perfect agreement*
Internal State Language	NA – Extraction based on a closed list	

Cohen's kappa and its interpretation are reported for each relevant category.

Cohen's κ statistics (Cohen, 1960). Cohen's κ were calculated in R (R Core Team, 2023) using the *kappa2* function from the *irr* package (Gamer and Lemon, 2019). See Table 4 for Cohen's kappa for each category.

Analysis

All statistical analyses were conducted in R (R Core Team, 2023). As the linguistic measures extracted from narratives were count data, we initially planned to use Generalized Linear Models with a Poisson family distribution. To ensure the validity of the Poisson regression, we tested the likelihood ratio for over-dispersion using the odTest function from the *pscl* package (Jackman, 2024). If the assumption of equidispersion was met, we proceeded with general linear models using the glm function (family = poisson). In cases of overdispersion, we used negative binomial models using the glm.nb function from the *MASS* package (Ripley, 2024). Results of likelihood ratio for over-dispersion for each model is available in Supp. Mat. 2.

For each linguistic measure, we proceeded by stepwise forward model comparison. Each baseline model included a fixed factor accounting for the length of the narrative or, when examining categories within a measure, the total number of elements for that measure (e.g. for the type of connectives, we included the total number of connectives produced in the narrative as a fixed factor). Subsequent models were iteratively expanded by adding predictor variables: category (when relevant), group, sex, as well as their interactions. Model selection was guided by the likelihood ratio tests, evaluating the significance of predictors and the fit of the models (see Supp. Mat. 3 for detailed model comparisons). The results reported are from the model that best fits the data. Post-hoc comparisons were conducted with Tukey adjustments using the emmeans functions from the *emmeans* package (Searle et al., 2024). Whenever we found a significant effect, we further controlled for IQ and language by augmenting, separately, the maximal model with FS-IQ and CELF scores.

Effect sizes were then calculated with the help of RcountD (Stefany Coxe, 2024), which allows to compute the effect sizes of count regression models, and are reported as Standardized Mean Differences (SMD).

Results

Participant characteristics and matching statistics are provided in Table 5.

For counts of main events, the addition of the group and sex did not improve the fit of the baseline model. Turning to other event counts, the best fitting model included the main effects of categories and group as well as an interaction effect between type of events and group. Post hoc comparisons revealed that autistic individuals produced more extraneous comments than non-autistic individuals ($\beta = -.92$, SE =.19, $p < .001$, SMD = −.796). Autistic individuals also produced more image descriptions than non-autistic individuals ($\beta = -1.72$, SE =.37, $p < .001$, SMD = −1.084). These differences remained significant when FS-IQ and CELF-5 scores were added to

Table 5
Descriptive statistics of participant's characteristics by group.

	Autistic males (Au-M)		Autistic females (Au-F)		Non-autistic males (NonAu-M)		Non-autistic females (NonAu-F)		Anova post hoc p-values			
	n	Mean (sd) range	N	Mean (sd) range	N	Mean (sd) range	N	Mean (sd) range	Au-M vs. Au-F	NonAu-M vs. NonAu-F	Au-M vs. NonAu-M	Au-F vs. NonAu-F
Age	30	12.2 (2.2) 9 −16	24	12.16 (2.21) 9 −16	29	12.34 (2.4) 9 −16	30	12.53 (2.41) 9 −16	.99	.99	.99	.95
SCQ	23	22.48 (6.16) 11 −33	23	20.52 (4.45) 12 −31	25	3.68 (3.19) 0 −10	25	2.8 (2.48) 0 −8	.41	.88	**< .001**	**< .001**
NVI	29	98.55 (17.43) 70 −134	23	97.32 (15.12) 74 −132	27	106.5 (11.94) 78 −125	24	111.71 (14.11) 82 −139	.99	.6	.2	**.007**
FSIQ	30	96.23 (17.39) 70 −129	24	96.13 (19.38) 64 −128	28	111.39 (12.45) 82 −139	30	111.53 (14.2) 73 −137	.99	.99	**.003**	**.003**
CLS	29	81.07 (18.06) 53 −125	23	81.7 (21.95) 44 −124	29	103.68 (11.45) 82 −127	30	101.77 (12.17) 70 −122	.99	.97	**< .001**	**< .001**
SES	24	7.48 (2.33) 3 −12	23	7.04 (1.54)4 −10	25	8.88 (2.3)4 −12	26	9.15 (2.11)4 −12	.55	.99	**.017**	**< .001**
GCC	24	29.33 (14.27) 2 −59	23	33 (15.22) 10 −68	25	72.8 (15.25° 42 −96	26	82.04 (14.15) 34 −98	.83	.13	**< .001**	**< .001**
SDQ	24	18.7 (6.55)6 −30	23	17.91 (6.03) 4 −32	25	7.12 (5.55)0 −22	26	5.46 (3.35)1 −13	.96	.71	**< .001**	**< .001**

Participants informations by group and sex. N indicates the number of participants for whom data were available. For each measure, mean, standard deviation and range are reported. Anova post-hoc p-values with Tukey adjustment are reported for group differences.
Abbreviations: SCQ = Social Communication questionnaire, NVI = Non-Verbal Index, FS-IQ = Full Scale Intelligence Quotient, CLS = Core Language Score, SES = Socio-Economic Status, GCC = General Communication Composite measured by the Children's Communication Checklist −2, SDQ = Strengths and Difficulties questionnaire

the model. There was no other significant interaction between categories and group (all $p > .19$). See Table 6 for descriptive statistics for the story structure and other events and Fig. 1 for fitted counts of other events per type and per group.

For total counts of connectives, the addition of the main effects of group and sex did not improve the fit of the baseline model. For type of connectives, the best fitting model included main effects of type of connectives and group, as well as an interaction effect between type and group. Post hoc comparisons revealed that autistic individuals produced less causal connectives than non-autistic individuals ($\beta = .76$, SE =.14, $p < .001$, SMD = 1.741). This difference remained significant when FS-IQ and CELF-5 scores were added to the model. There were no other significant interaction effects between categories and group (all $p > .2$). See Table 7 for descriptive statistics for the connectives and Fig. 2 for fitted counts of connectives per category and group.

Regarding mention of main characters, the best fitting model included the main effect of group, with non-autistic participants producing more references to main characters than autistic participants ($\beta = -.121$, SE =.04, $p = .003$, SMD = −.433). This difference remained significant when FS-IQ and CELF-5 scores were added to the model. See Fig. 3 for fitted counts of mention to main characters per group.

For the choice of referential expressions, the best fitting model included the main effects of type of referential expression and group, as well as the interaction effect between type and group. Post-hoc comparisons revealed that autistic individuals produced more indefinite referential expression than non-autistic individuals, ($\beta = -.26$, SE =.089, $p = .003$, SMD = −.616). This difference remained significant when FS-IQ and CELF-5 scores were added to the model. There were no other significant interaction effects between type and group (all $p > .062$). See Fig. 4 for fitted counts of referential expression per type and group.

Regarding production of ambiguous pronouns, the best fitting model included a main effect of group, with autistic individuals producing significantly more ambiguous pronouns than non-autistic individuals ($\beta = .467$, SE =.155, $p = .003$, SMD =.659). However, this result did not remain significant once CELF-5 scores was added to the model ($p = .068$). See Table 8 for descriptive statistics for the references to characters.

For production of ISL, the best fitting model included the main effect of group, with non-autistic participants producing more ISL than autistic participants ($\beta = .27$, SE =.046, $p < .001$, SMD = −.756). This difference remained significant when FS-IQ and CELF-5 scores were added to the model. See Fig. 5 for fitted counts of ISL per group.

For type of ISL, the addition of the group and sex did not improve the fit of the model. Regarding the production of causes and consequences of ISL, the best fitting model included a main effect of group, with autistic individuals producing significantly less explanations of ISL than non-autistic individuals ($\beta = -.643$, SE =.229, $p = .005$, SMD = −.252). This remained significant when FS-IQ and CELF-5 scores were added to the model. See Table 9 for descriptive statistics for ISL and Fig. 6 for fitted counts of explanation of ISL per group.

Discussion

This study aimed to further delineate the communicative profile of autistic females. We assessed the coherence of narrative productions by autistic children and adolescents of both sexes and compared them to those by their sex-matched non-autistic peers. We hypothesized that narratives of autistic females would be more coherent than narratives of autistic males, but also less coherent than narratives of non-autistic individuals. This is so because we anticipated that communication difficulties experienced by autistic females

Table 6
Counts of events by diagnosis group and sex: mean, sd & proportion (%).

	Autistic group		Non-autistic group		Sex overall	
	Counts	*%*	*Counts*	*%*	*Counts*	*%*
Main events						
Females	17.88 (3.14)	26.88	19.27 (2.03)	28.47	18.65 (2.65)	27.77
Males	18.38 (2.92)	27.19	18.69 (2)	29.01	18.53 (2.49)	28.08
Group overall	18.15 (3)	27.05	18.98 (2.02)	28.73		
Additional story events						
Females	5.23 (3.17)	6.22	4.93 (2.77)	7.29	5.07 (2.93)	7.56
Males	4.21 (3.37)	6.22	4.97 (3.06)	7.71	4.59 (3.22)	6.95
Group overall	4.68 (3.29)	6.97	4.95 (2.89)	7.49		
Dog events						
Females	6.63 (4.21)	9.96	6 (2.3)	8.87	6.28 (3.27)	9.35
Males	7.14 (3.42)	10.56	6.28 (2.43)	9.74	6.71 (2.97)	10.16
Group overall	6.91 (3.77)	10.29	6.14 (2.35)	9.29		
Extraneous comments						
Females	2.13 (2.46)	3.2	1.1 (1.27)	1.63	1.56 (1.94)	2.32
Males	2.1 (2.09)	3.11	0.66 (1.11)	1.02	1.38 (1.81)	2.09
Group overall	2.11 (2.24)	3.15	0.88 (1.2)	1.33		
Image description						
Females	0.83 (2.55)	1.25	0.13 (0.35)	0.2	0.44 (1.73)	0.66
Males	0.79 (2.37)	1.17	0.17 (0.54)	0.69	0.48 (1.73)	0.73
Group overall	0.81 (2.43)	1.21	0.15 (0.44)	0.23		

Production of events by group and sex. For each event, mean, standard deviation and proportion in percentage are reported. Proportions are calculated relatively to the total number of events.

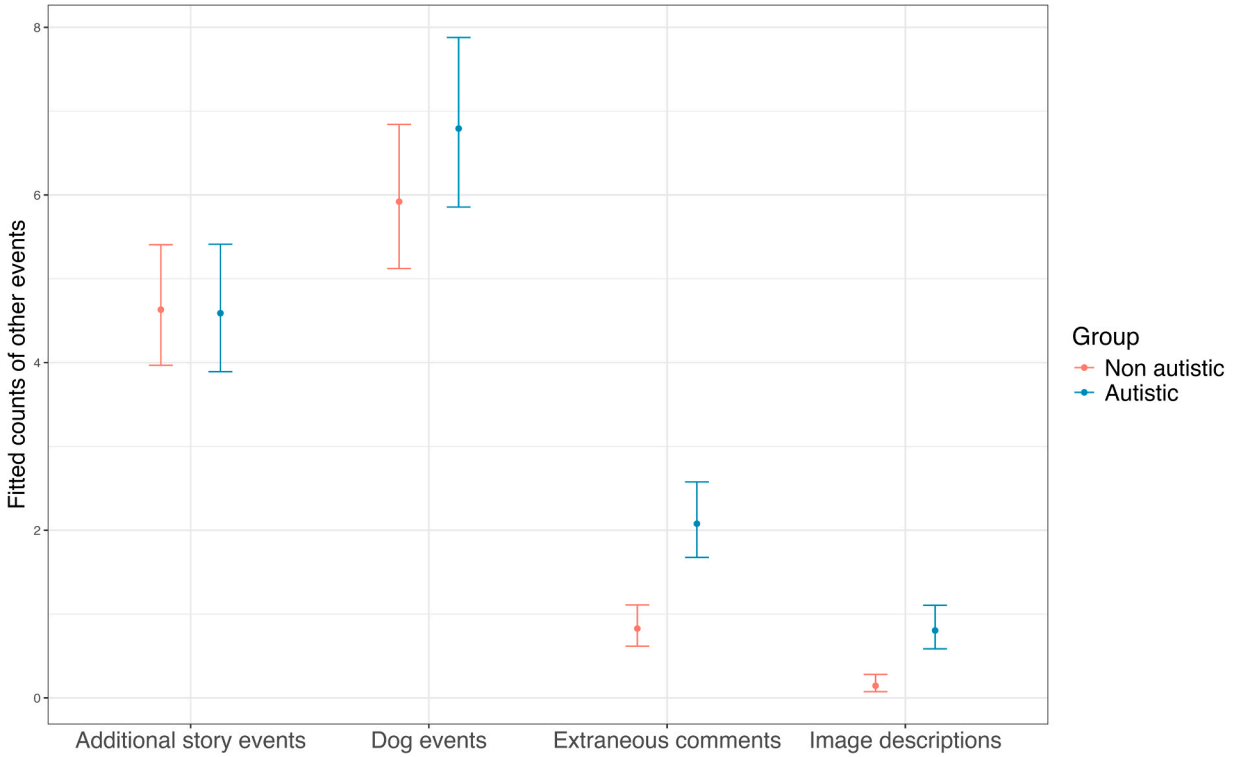

Fig. 1. Fitted counts of mentioned events per category and per group. Note: Figure shows the fitted counts (on the y-axis) for "Additional story events", "Dog events", "Extraneous comments", and "Image descriptions" (on the x-axis). The dots represent the estimated mean counts for each event type within each group, with error bars indicating the standard errors.

Table 7
Counts of connectives by diagnosis group and sex: mean, sd & proportion (%).

	Autistic group		Non-autistic group		Sex overall	
	Counts	*%*	*Counts*	*%*	*Counts*	*%*
Connectives total						
Females	19.36 (13.35)	4.63	20.67 (9.69)	4.35	20.07 (11.4)	4.47
Males	25.93 (12.91)	5.86	19.21 (11.52)	4.37	22.57 (12.59)	5.11
Group overall	22.89 (13.4)	5.31	19.95 (10.56)	4.36		
Additive link						
Females	19.36 (13.35)	43.88	20.67 (9.69)	42.49	20.07 (11.4)	43.09
Males	25.93 (12.91)	49.15	19.21 (11.52)	40.66	22.57 (12.59)	45.14
Group overall	22.89 (13.4)	46.94	19.95 (10.56)	41.6		
Temporal link						
Females	7.08 (6.8)	16.05	8.83 (5.79)	18.16	8.04 (6.27)	17.25
Males	9.97 (8.04)	18.89	8.9 (7.78)	18.83	9.43 (7.76)	18.86
Group overall	8.63 (7.56)	17.7	8.86 (6.78)	18.49		
Causal link						
Females	3.2 (5.76)	7.25	7.03 (6.46)	14.46	5.29 (6.39)	11.36
Males	3.13 (3.66)	6.27	7.24 (8.75)	15.33	5.28 (6.94)	10.55
Group overall	3.26 (4.7)	6.68	7.14 (7.6)	14.88		
Contrastive link						
Females	3.96 (4.23)	8.98	3.33 (2.97)	6.85	3.62 (3.55)	7.77
Males	3.52 (3.95)	6.67	2.59 (2.37)	5.47	3.05 (3.26)	6.1
Group overall	3.72 (4.05)	7.63	2.97 (2.67)	6.19		

Production of connectives by group and sex. For each subcategory, mean, standard deviation and proportion in percentage are reported. Proportions are calculated relatively to the total number of words in the narrative for total number of connectives and to the total number of connectives for types of link.

might be more subtle compared to those experienced by autistic males. While we replicated previous findings on narrative coherence of autistic individuals, we did, however, not find any evidence for sex differences in narratives of autistic individuals. Analyzing the coherence of narratives did not provide any evidence that the autistic females differ from autistic males in their linguistic profile.

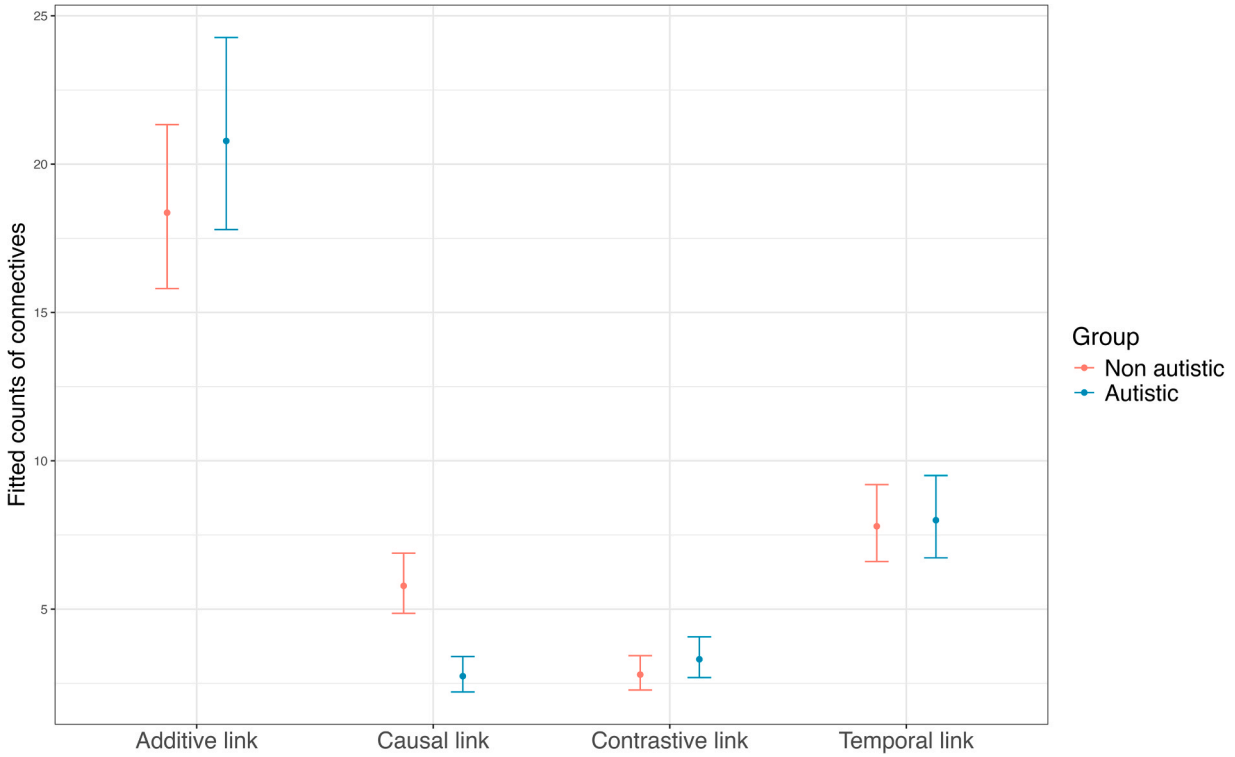

Fig. 2. Fitted counts of connectives per category and per group. Note: Figure shows the fitted counts (on the y-axis) for "Additive link", "Causal link", "Contrastive link", and "Temporal link" (on the x-axis). The dots represent the estimated mean counts for each event type within each group, with error bars indicating the standard errors.

Overall, narratives of autistic male and female our participants were comparable, but differed in several respects from those of non-autistic individuals, male and female. We will begin by discussing group differences, and turn next to possible reason for the absence of significant sex effects.

Starting with story grammar, there were no significant differences between groups: autistic and non-autistic participants included a similar amount of main story events, additional story events and dog events. These results suggest that both autistic and non-autistic participants might be similarly sensitive to the importance of the gist of the story (Diehl et al., 2006; Geelhand et al., 2020) and organize their discourse according to the story grammar. Interestingly, autistic participants produced more image descriptions compared to non-autistic participants, which suggests an increased attention to details (Happé & Frith, 2006). Moreover, autistic participants were also more likely to produce extraneous comments. This suggests that despite having no difficulties in identifying the main story events, autistic participants still included disruptive elements that reduce the overall coherence of their narratives (Geelhand et al., 2020).

Turning to connectives, autistic individuals produced less causal connectives than non-autistic individuals. These findings are consistent with previous studies that report differences between autistic and non-autistic groups (Baixauli et al., 2016; Geelhand et al., 2020). Reduced use of causal connectives by autistic individuals may make it harder for listeners to follow the storyline, as storytelling becomes a mere list of events rather than a structured narrative (Diehl et al., 2006).

Concerning references to characters, autistic individuals mentioned less frequently the main characters of the story. Additionally, groups differed in their choice of referential expressions, with autistic individuals producing more indefinite referential expressions. The choice of referential expression carries a pragmatic function: the speaker must represent what context is already known by the interlocutor and determine what new information needs to be provided (Colle et al., 2008). The use of a pronominal or definite expression assumes shared knowledge between interlocutors whereas the use of an indefinite expression assumes least shared knowledge and is typically used to introduce a new character in the story (Van Der Lely, 1997). We found that autistic individuals keep using indefinite expression even when the character had already been introduced. The use of more indefinite referential expressions could result in reduced narrative fluency (Colle et al., 2008) and contribute to a more rigid and overall less coherent discourse structure.

Autistic individuals also struggled with cohesive adequacy by producing more ambiguous pronominal references (i.e., pronouns with competing referent), However, when we controlled for the CELF-5 this difference was not significant anymore, suggesting that in our sample, the production of ambiguous pronouns was better predicted by language abilities than by diagnostic status. This result is surprising given that studies that included well-matched participants with language abilities within the typical range still find an effect of the group (Banney et al., 2015; Colle et al., 2008; Novogrodsky, 2013). However, one study with similarly well-matched participants

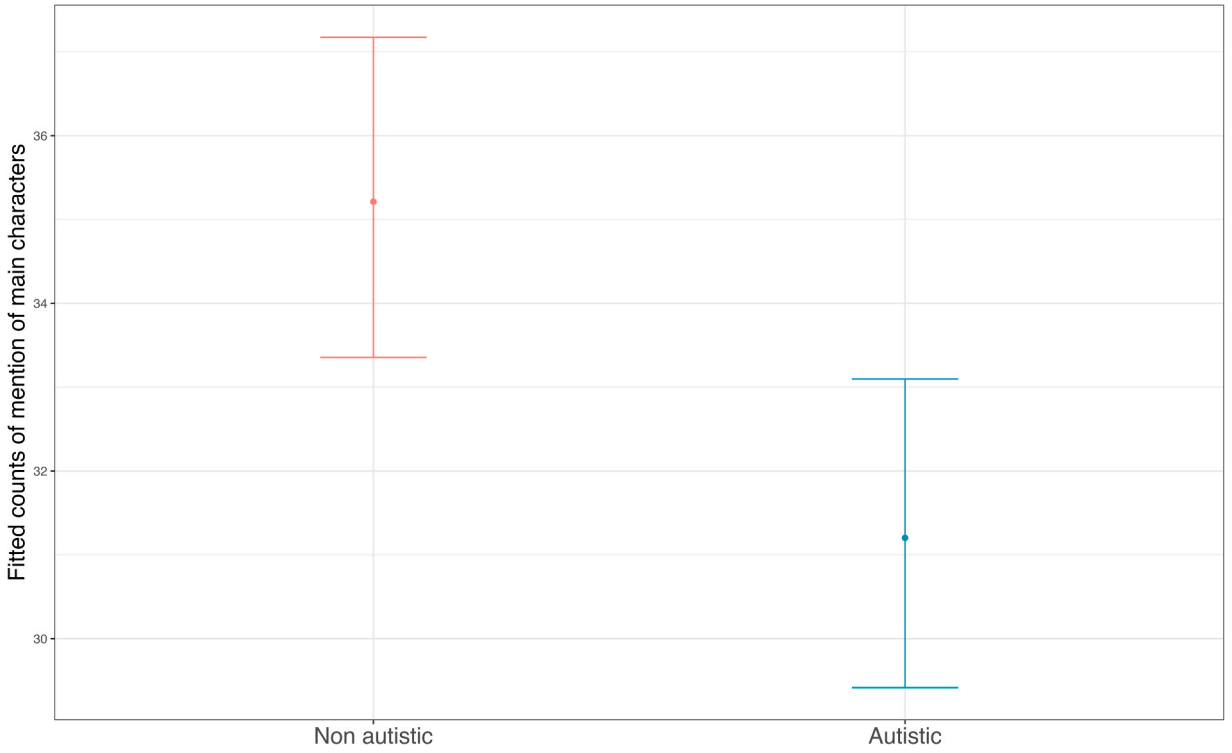

Fig. 3. Fitted counts of mention of main characters. Note: Figure shows the fitted counts (on the y-axis) for "Non-autistic" and "Autistic" groups (on the x-axis). The dots represent the estimated mean counts for each event type within each group, with error bars indicating the standard errors.

(Zane & Grossman, 2023) reports no differences between the autistic and non-autistic groups. This highlights the importance of conducting cross-linguistic studies to determine how abilities in different languages may influence the production of ambiguous pronouns.

Finally, in line with previous studies (Baixauli et al., 2016), autistic participants produced fewer references to Internal State Language than non-autistic participants. Moreover, they motivated Internal State Language less frequently than non-autistic participant, including less explanation of the causes and consequences of the character's internal state. Consequently, not only did autistic participants differed from non-autistic participants in the way they referred to the story characters, but they also faced difficulties in building strong characters by expressing the point of view of the characters in the story less often.

While this study is mostly in line with previous work on narrative coherence in autism, it did not replicate previous findings on sex differences in narratives of autistic individuals. We expected sex differences to emerge in autistic participant with respect to the story grammar and the production of Internal State Language, but this prediction was not borne out. For the story grammar, our results mirror previous findings by Kauschke et al. (2016) who found no evidence of a difference between autistic males and females. In contrast with our results, Conlon et al. (2019) reported that autistic females included more salient story elements than autistic males. This discrepancy between findings highlights the importance of methodological choices in evaluating narrative structure (Harvey et al., 2023). Both Kauschke et al. (2016) and the current study used a quantitative approach, whereas Conlon et al. (2019) used a qualitative approach. That is, in our coding scheme, the mention of an event would always be counted as the inclusion of one event, whereas a rating score system would give each event a score (from 0 to 2) according to the completeness of the information. In our study, we decided to use a counting measure, given that other aspects of the narrative were also being evaluated, and qualitative information might be influenced by elements assessed in other measures (e.g., causal connectives, choice of referential expression).

As already pointed out, various methodological choices across studies could explain the inconsistent results. Studies investigating sex differences in narratives of autistic participant differ in every step, including experimental design, comparison groups, variable of interest, assessment of those variables, and statistical choices. These differences in methodology can significantly impact the findings and interpretations, leading to variability in the results observed across different studies. Our analyses are based on a comprehensive and easily replicable coding manual (available in Supp. Mat. 1). This coding manual helps to provide a clear and detailed framework for data analysis. By providing access to this manual, we aim to ensure a better understanding of our methodology and facilitate comparison with other studies. Future research should systematically include comparison groups and provide access to their coding manuals. Doing so would enable researchers to replicate analyses under different frameworks and potentially reveal how different coding systems, such as using a counting system versus a rating score system, could affect the results.

One reason for the absence of differences between autistic females and autistic males in our sample could be that the autistic females in our study have profiles more similar to those of autistic males in general–in contrast to more distinct profiles previous

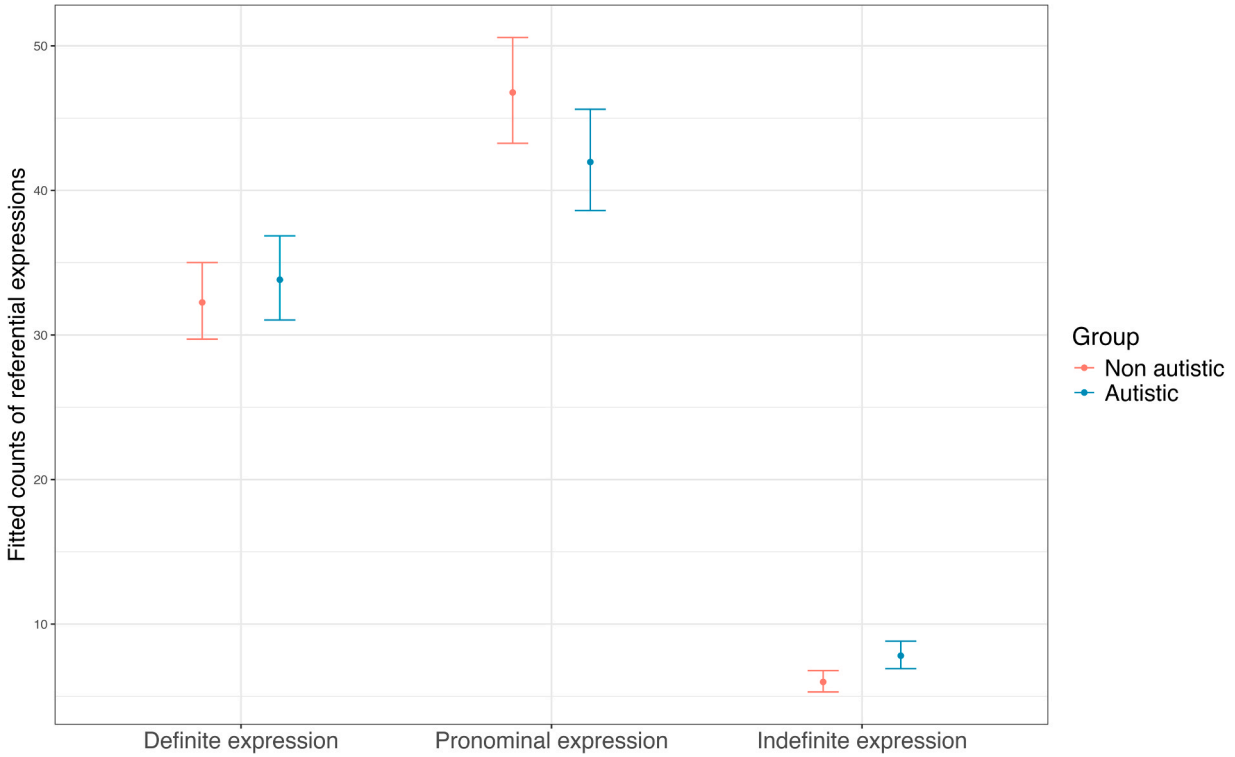

Fig. 4. Fitted counts per category per group for referential expressions. Note: Figure shows the fitted counts (on the y-axis) for "Definite expression", "Pronominal expression" and "Indefinite expression" (on the x-axis). The dots represent the estimated mean counts for each event type within each group, with error bars indicating the standard errors.

Table 8
Counts of references by diagnosis group and sex: mean, sd & proportion (%).

	Autistic group		Non-autistic group		Sex overall	
	Counts	*%*	*Counts*	*%*	*Counts*	*%*
Main characters						
Females	30.96 (10.4)	7.41	37.97 (12.99)	7.98	34.78 (12.92)	7.74
Males	32.1 (11.57)	7.25	36.52 (14.91)	8.3	34.31 (13.41)	7.78
Group overall	31.57 (10.96)	7.32	37.25 (13.86)	8.13		
Definite expression						
Females	34.4 (14.71)	42.85	35.7 (12.35)	37.4	35.11 (13.36)	39.64
Males	32.38 (14.61)	37.95	33.21 (9.23)	35.34	32.79 (12.14)	36.58
Group overall	33.31 (14.55)	40.15	34.47 (10.93)	36.39		
Indefinite expression						
Females	8.04 (4.34)	10.1	6.73 (2.92)	7.05	7.33 (3.66)	8.27
Males	7.14 (3.86)	8.37	6.45 (2.68)	6.86	6.79 (3.31)	7.58
Group overall	7.56 (4.08)	9.11	6.59 (2.79)	6.96		
Pronominal expression						
Females	37.84 (22.19)	47.14	53.03 (23.84)	55.55	46.13 (24.13)	52.08
Males	45.79 (19.84)	53.68	54.31 (33.18)	57.8	50.05 (27.44)	55.84
Group overall	42.11 (21.14)	50.75	53.66 (28.57)	56.65		
Ambiguous pronoun						
Females	9.84 (8.57)	26	9.73 (8.73)	18.35	9.78 (8.58)	21.21
Males	13.72 (8.89)	29.97	9.62 (11.69)	17.71	11.67 (10.5)	23.32
Group overall	11.93 (8.88)	28.32	9.68 (10.2)	18.04		

Production of references by group and sex. For each subcategory, mean, standard deviation and proportion in percentage are reported. Proportions are calculated relatively to total number of words for references to main characters, total number of references for type of referential expressions, and total number of pronominal expressions for ambiguous pronouns

studies found in autistic girls and women. Autistic females and males in our sample were very similar in terms of cognitive abilities, language abilities as well as SCQ scores. In particular, autistic participants had mean CELF-5 scores below the typical ranges, with huge heterogeneity within the group, ranging from very low language abilities to score well above the typical range. By contrast, in all

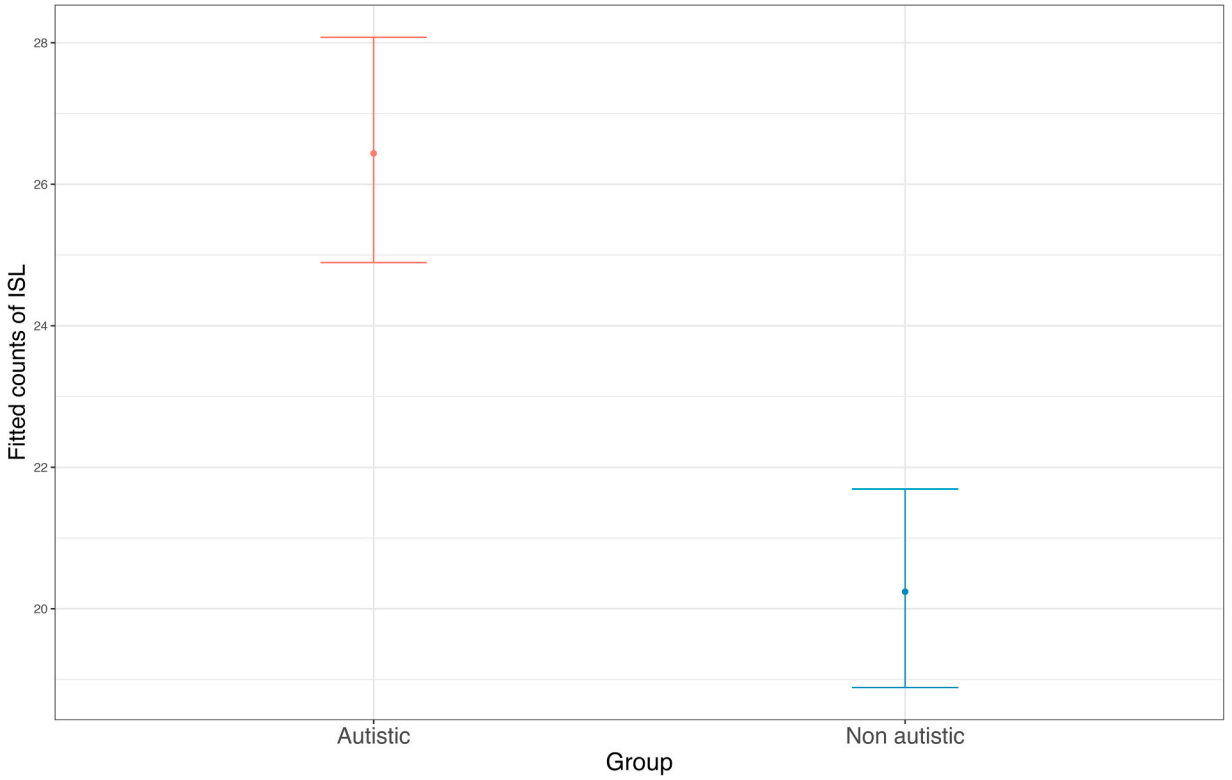

Fig. 5. Fitted counts of Internal State Language per group. Note: Figure shows the fitted counts (on the y-axis) for "Non-autistic" and "Autistic" groups (on the x-axis). The dots represent the estimated mean counts for each event type within each group, with error bars indicating the standard errors.

previous studies on gender and language in autism, autistic females had core language abilities or verbal IQ within the typical range (Boorse et al., 2019; Conlon et al., 2019; Kauschke et al., 2016; Sturrock, Yau et al., 2019).

Furthermore, only seven females of our sample were diagnosed in late childhood / adolescence (i.e. after 8). Most of them were identified as autistic before the age of five (n = 17), which can be considered as the average age of diagnosis for autistic individuals (van 't Hof et al., 2021). The fact that females in our sample were diagnosed at an expected age suggests they may have been identified early because they resemble autistic males, making them less representative of females who are diagnosed late or misdiagnosed (Hiller et al., 2014). If the purpose of studying sex differences in the language of autistic individuals is to understand the reason why we are failing to identify autistic females, looking at individuals with such a relatively early diagnosis might not be that informative. A next step would be to investigate this question in a sample of females without an autism diagnosis but at high likelihood of being diagnosed as adults (i.e., siblings of autistic children).

A clear strength of the present study is the relatively large sample of verbal autistic females as well as its systematic analysis of narratives, but it also has several limitations. Some of them have already been discussed, such as the representativity of our autistic female sample. To this, one can add that while a semi-structured task narrative task, can provide valuable insights, it does not account for other important aspects of everyday conversation, such as turn-taking. Furthermore, difficulties in narrative abilities might be underestimated during a task based on a book and accompanied by clear instructions (Losh & Gordon, 2014). Accordingly, it would be interesting to compare semi-structured narratives, elicited with a wordless picture book, with more spontaneous, personal discourse samples.

Another point is that, while our study does include more participants than previous ones, the number of participants might still not be large enough. Small sample sizes can lead to overestimated effect sizes and increased false positives (Kühberger et al., 2014), or insufficient power to detect true effects, as suggested in Sturrock et al. (2019). This creates a paradox: small samples may yield significant but inflated effects or fail to find true effects due to low statistical power, leading to poor reproducibility (Button et al., 2013). Therefore, studies with more autistic females are needed for reliable results. However, here comes the vicious cycle: we need to investigate sex differences to better diagnose autistic females, but to do so, we need more females in our studies to detect these differences. Indeed, as autistic females are underdiagnosed, it makes them difficult to recruit for research, and they remain underdiagnosed precisely because they are not adequately represented in research (Cruz et al., 2024; D'mello et al., 2022; Hull et al., 2020).

Table 9
Counts of ISL by diagnosis group and sex: mean, sd & proportion (%).

	Autistic group		Non-autistic group		Sex overall	
	Counts	*%*	*Counts*	*%*	*Counts*	*%*
ISL total						
Females	20.36 (9.5)	4.87	28.17 (11.19)	5.92	24.62 (11.08)	5.48
Males	20.97 (8.11)	4.74	28.69 (12.18)	6.52	24.83 (10.97)	5.63
Group overall	20.69 (8.7)	4.8	28.42 (11.59)	6.21		
Physiological terms						
Females	4.8 (2.36)	23.58	5.83 (3.9)	20.71	5.36 (3.3)	21.79
Males	5 (3.28)	23.85	6.86 (4.72)	23.92	5.93 (4.14)	23.89
Group overall	4.91 (2.87)	23.72	4.91 (2.87)	22.3		
Emotion terms						
Females	2.72 (2.62)	13.36	4.23 (3.41)	15.03	3.55 (3.14)	14.4
Males	2.97 (2.47)	14.14	3.9 (3.04)	13.58	3.43 (2.79)	13.82
Group overall	2.85 (2.52)	13.79	4.07 (3.21)	14.31		
Cognitive terms						
Females	1.76 (2.44)	8.64	2.53 (2.89)	8.99	2.18 (2.69)	8.86
Males	1.21 (1.37)	5.76	1.62 (2.09)	5.65	1.41 (1.77)	5.69
Group overall	1.46 (1.94)	7.07	2.08 (2.55)	7.33		
Modality terms						
Females	2.28 (2.14)	10.86	2.53 (1.94)	8.99	2.29 (2.35)	9.31
Males	2 (2.78)	9.82	2.52 (2.5)	8.77	2.4 (2.31)	9.65
Group overall	2.15 (2.44)	10.38	2.52 (2.22)	8.88		
Intention terms						
Females	8.44 (4.98)	41.45	11.27 (3.76)	40	9.98 (4.54)	40.55
Males	8.97 (4.15)	42.76	11.69 (5.13)	40.75	10.33 (4.82)	41.6
Group overall	8.72 (4.52)	42.17	11.47 (4.45)	40.37		
Explanation of ISL						
Females	20.97 (8.11)	3.14	28.17 (11.19)	6.27	24.62 (11.08)	5.1
Males	20.36 (9.5)	2.63	28.69 (12.18)	7.33	24.83 (10.97)	5.35
Group overall	20.69 (8.7)	2.86	28.42 (11.59)	6.8		

Production of ISL by group and sex. For each subcategory, mean, standard deviation and proportion in percentage are reported. Proportions are calculated to the total number of words for total number of ISL, and total number of ISL for ISL subcategories and explanation of ISL.

Conclusion

The present study analyzed sex differences in narrative abilities of autistic individuals in a sample larger than most previous studies. We relied on an open, easily reproducible coding manual, making future replications or study comparisons easier. The narratives of autistic individuals were overall less coherent than those of non-autistic individuals. Autistic individuals included more extraneous comments and more image descriptions, deviating from the story grammar. Additionally, the events in their narratives were less causally connected and they used more indefinite referential expressions. They also built weaker characters by mentioning less often the main characters of the story and referring less to their mental states. Crucially, no significant interaction effect between group and sex was found. While examining autistic females identified at a young age provides valuable insights, it may not fully explain why some are diagnosed later in life. Our study suggests that focusing on these early-identified individuals may not provide the insights needed to address the issue of late diagnosis in autistic females.

CRediT authorship contribution statement

Dumont Charlotte: Investigation, Writing – review & editing. **Geelhand Philippine:** Methodology, Writing – review & editing. **Belenger Marie:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Writing – original draft. **Kissine Mikhail:** Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing.

Ethics approval

Ethical approval was received for the study from the Erasme-ULB ethics committee in accordance with the Declaration of Helsinki. Participants' parents signed a written consent for their children to be enrolled in this study after being informed of their rights and all aspects of the experimental design. Written consent was obtained from children enrolled in this study.

Funding

The project was supported by both Margueritte-Marie Delacroix fundings [grant number RVC/B-476] and Fonds de la Recherche Scientifique [grant number 4003675].

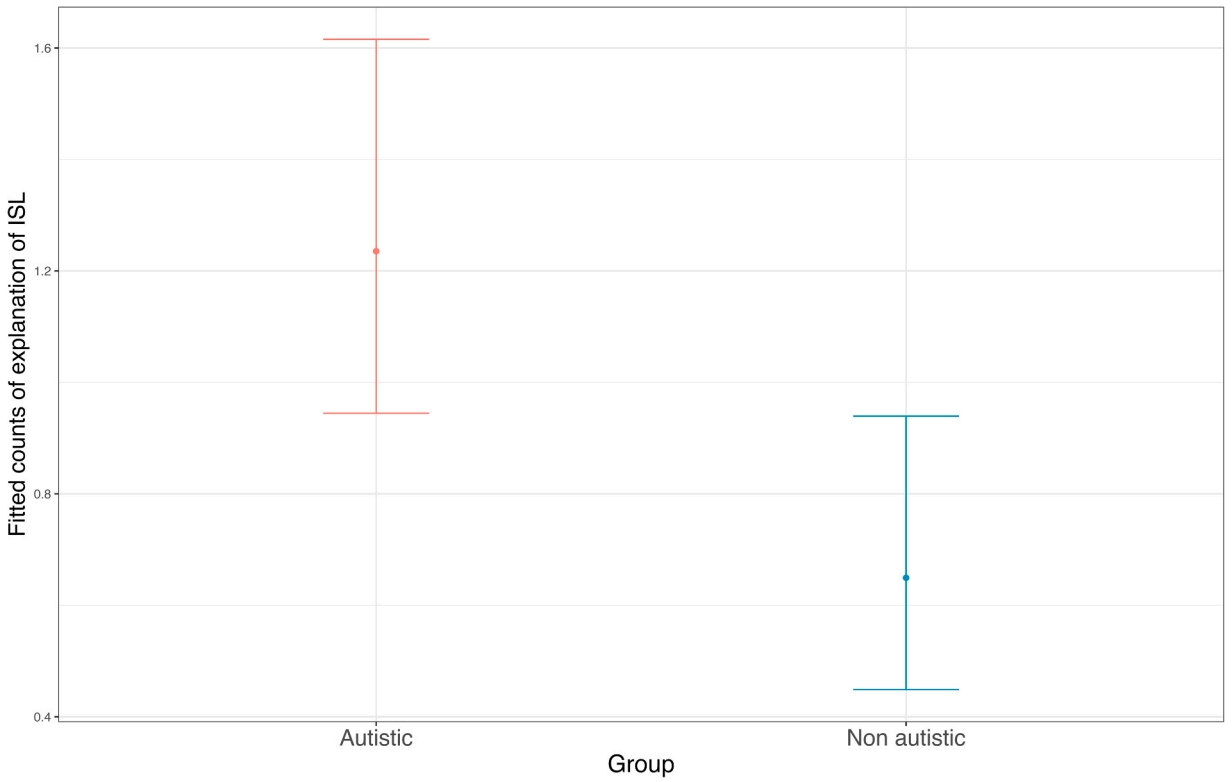

Fig. 6. Fitted counts of explanation of Internal State Language per group. Note: Figure shows the fitted counts (on the y-axis) for "Non-autistic" and "Autistic" groups (on the x-axis). The dots represent the estimated mean counts for each event type within each group, with error bars indicating the standard errors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are grateful to all the participating children, parents, and schools that collaborated with us. We would like to thank the students who transcribed and annotated the audio and in particular Alizée Thion for all the work she has put on this study.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.reia.2025.202589.

Data availability

Data will be made available on request.

References

Baixauli, I., Colomer, C., Roselló, B., & Miranda, A. (2016). Narratives of children with high-functioning autism spectrum disorder: A meta-analysis (Elsevier Inc) *In Research in Developmental Disabilities, 59*, 234–254. https://doi.org/10.1016/j.ridd.2016.09.007.

Banney, R. M., Harper-Hill, K., & Arnott, W. L. (2015). The Autism Diagnostic Observation Schedule and narrative assessment: Evidence for specific narrative impairments in autism spectrum disorders. *17*(2), 159–171. https://doi.org/10.3109/17549507.2014.977348

Beaumont, R. B., & Newcombe, P. (2006). Theory of mind and central coherence in adults with high-functioning autism or Asperger syndrome. *Autism: The International Journal of Research and Practice, 10*(4), 365–382. https://doi.org/10.1177/1362361306064416

Beggiato, A., Peyre, H., Maruani, A., Scheid, I., Rastam, M., Amsellem, F., Gillberg, C. I., Leboyer, M., Bourgeron, T., Gillberg, C., & Delorme, R. (2017). Gender differences in autism spectrum disorders: Divergence among specific core symptoms. *Autism Research, 10*(4), 680–689. https://doi.org/10.1002/AUR.1715

Bishop, D. (2004). *Expression, reception and recall of narrative instrument.* Harcourt assessment,.
Boersma, P., & Weenink, D. (2022). Praat: Doing Phonetics by Computer (6.2.15).
Boorse, J., Cola, M., Plate, S., Yankowitz, L., Pandey, J., Schultz, R. T., & Parish-Morris, J. (2019). Linguistic markers of autism in girls: Evidence of a "blended phenotype" during storytelling. *Molecular Autism, 10*(1). https://doi.org/10.1186/s13229-019-0268-2
Bruner, J. (1991). *The Narrative Construction of Reality*. ⟨https://about.jstor.org/terms⟩.
Burton, J. M., Creaghead, N. A., Silbert, N., Breit-Smith, A., Duncan, A. W., & Grether, S. M. (2020). Social communication and structural language of girls with high-functioning autism spectrum disorder. *Language, Speech, and Hearing Services in Schools, 51*(4), 1139–1155. https://doi.org/10.1044/2020_LSHSS-20-00004
Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience 2013, 14*(5), 365–376. https://doi.org/10.1038/nrn3475
Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104
Cola, M. L., Plate, S., Yankowitz, L., Yankowitz, L., Petrulla, V., Bateman, L., Zampella, C. J., De Marchena, A., De Marchena, A., Pandey, J., Pandey, J., Schultz, R. T., Schultz, R. T., Parish-Morris, J., & Parish-Morris, J. (2020). Sex differences in the first impressions made by girls and boys with autism. *Molecular Autism, 11*(1). https://doi.org/10.1186/S13229-020-00336-3
Cola, M., Yankowitz, L. D., Tena, K., Russell, A., Bateman, L., Knox, A., Plate, S., Cubit, L. S., Zampella, C. J., Pandey, J., Schultz, R. T., & Parish-Morris, J. (2022). Friend matters: Sex differences in social language during autism diagnostic interviews. *Molecular Autism, 13*(1). https://doi.org/10.1186/S13229-021-00483-1
Colle, L., Baron-Cohen, S., Wheelwright, S., & Van Der Lely, H. K. J. (2008). Narrative discourse in adults with high-functioning autism or Asperger syndrome. *Journal of Autism and Developmental Disorders, 38*(1), 28–40. https://doi.org/10.1007/s10803-007-0357-5
Conlon, O., Volden, J., Smith, I. M., Duku, E., Zwaigenbaum, L., Waddell, C., Szatmari, P., Mirenda, P., Vaillancourt, T., Bennett, T., Georgiades, S., Elsabbagh, M., & Ungar, W. J. (2019). Gender differences in pragmatic communication in school-aged children with Autism Spectrum Disorder (ASD). *Journal of Autism and Developmental Disorders, 49*(5), 1937–1948. https://doi.org/10.1007/s10803-018-03873-2
Stefany Coxe. (2024). RcountD. Effect size for Poisson regression. ⟨Https://Stefany.Shinyapps.Io/RcountD/⟩.
Cruz, S., Zubizarreta, S. C. P., Costa, A. D., Araújo, R., Martinho, J., Tubío-Fungueiriño, M., Sampaio, A., Cruz, R., Carracedo, A., & Fernández-Prieto, M. (2024). Is there a bias towards males in the diagnosis of autism? A systematic review and meta-analysis. *Neuropsychology Review 2024, 1–24*. https://doi.org/10.1007/S11065-023-09630-2
D'mello, A. M., Frosch, I. R., Li, C. E., Cardinaux, A. L., Gabrieli, J. D. E., Tan, H. E., Lisa, K., Center, Y., Mello, A. M. D.', & Yang, K. L. (2022). Exclusion of females in autism research: Empirical evidence for a "leaky" recruitment-to-research pipeline. *Autism Research, 15*(10), 1929–1940. https://doi.org/10.1002/AUR.2795
Dean, M., Adams, G. F., & Kasari, C. (2013). How narrative difficulties build peer rejection: A discourse analysis of a girl with autism and her female peers. *Discourse Studies, 15*(2), 147–166. https://doi.org/10.1177/1461445612471472
den Hartog, A., de la Roche, L., Derby, B., Psaradellis, E., & Kelley, E. (2023). A scoping review of sex/gender differences in pragmatic language and friendship characteristics in autistic children and adolescents. *Research in Autism Spectrum Disorders, 108*, Article 102229. https://doi.org/10.1016/J.RASD.2023.102229
Diehl, J. J., Bennetto, L., & Young, E. C. (2006). Story recall and narrative coherence of high-functioning children with autism spectrum disorders. *Journal of Abnormal Child Psychology, 34*(1), 87–102. https://doi.org/10.1007/S10802-005-9003-X
Dworzynski, K., Ronald, A., Bolton, P., & Happé, F. (2012). How different are girls and boys above and below the diagnostic threshold for autism spectrum disorders? *Journal of the American Academy of Child and Adolescent Psychiatry, 51*(8), 788–797. https://doi.org/10.1016/j.jaac.2012.05.018
Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatric Research, 65*(6), 591–598. https://doi.org/10.1203/PDR.0B013E31819E7203
Gamer M., & Lemon J. (2019). irr: Various Coefficients of Interrater Reliability and Agreement. ⟨https://cran.r-project.org/web/packages/irr/index.html⟩.
Geelhand, P., Papastamou, F., Deliens, G., & Kissine, M. (2020). Narrative production in autistic adults: A systematic analysis of the microstructure, macrostructure and internal state language. *Journal of Pragmatics, 164*, 57–81. https://doi.org/10.1016/j.pragma.2020.04.014
Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 38*(5), 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x
Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 36*(1), 5–25. https://doi.org/10.1007/S10803-005-0039-0/TABLES/1
Harrop, C., Jones, D., Zheng, S., Nowell, S. W., Boyd, B. A., & Sasson, N. (2018). Sex differences in social attention in autism spectrum disorder. *Autism Research, 11*(9), 1264–1275. https://doi.org/10.1002/aur.1997
Harvey, A., Spicer-Cain, H., Botting, N., Ryan, G., & Henry, L. (2023). Assessing 'coherence' in the spoken narrative accounts of autistic people: A systematic scoping review. *Research in Autism Spectrum Disorders, 102*, Article 102108. https://doi.org/10.1016/J.RASD.2023.102108
Hiller, R. M., Young, R. L., & Weber, N. (2014). Sex differences in Autism Spectrum Disorder based on DSM-5 Criteria: Evidence from clinician and teacher reporting. *Journal of Abnormal Child Psychology, 42*(8), 1381–1393. https://doi.org/10.1007/S10802-014-9881-X/TABLES/4
Holtmann, M., Bölte, S., & Poustka, F. (2007). Autism spectrum disorders: Sex differences in autistic behaviour domains and coexisting psychopathology. *Developmental Medicine and Child Neurology, 49*(5), 361–366. https://doi.org/10.1111/j.1469-8749.2007.00361.x
Hull, L., Petrides, K. V., & Mandy, W. (2020). The female autism phenotype and camouflaging: A narrative review (Springer) *Review Journal of Autism and Developmental Disorders, 7*(4), 306–317. https://doi.org/10.1007/s40489-020-00197-9.
Jackman, S. (2024). Political Science Computational Laboratory [R package pscl version 1.5.9]. ⟨https://CRAN.R-project.org/package=pscl⟩.
Kauschke, C., van der Beek, B., & Kamp-Becker, I. (2016). Narratives of girls and boys with Autism Spectrum Disorders: Gender differences in narrative competence and internal state language. *Journal of Autism and Developmental Disorders, 46*(3), 840–852. https://doi.org/10.1007/s10803-015-2620-5
Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLOS ONE, 9*(9), Article e105825. https://doi.org/10.1371/JOURNAL.PONE.0105825
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159. https://doi.org/10.2307/2529310
Lê, K., Coelho, C., Mozeiko, J., & Grafman, J. (2011). Measuring goodness of story narratives. *Journal of Speech, Language, and Hearing Research, 54*(1), 118–126. https://doi.org/10.1044/1092-4388(2010/09-0022)
Liles, B. Z. (1985). Production and comprehension of narrative discourse in normal and language disordered children. *Journal of Communication Disorders, 18*(6), 409–427. https://doi.org/10.1016/0021-9924(85)90030-9
Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis (Elsevier Inc) *Journal of the American Academy of Child and Adolescent Psychiatry, 56*(6), 466–474. https://doi.org/10.1016/j.jaac.2017.03.013.
Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnosis observation schedule, second edition (ADOS-2)*. Western Psychological Services,.
Losh, M., & Capps, L. (2003). Narrative ability in high-functioning children with autism or Asperger's Syndrome. *Journal of Autism and Developmental Disorders, 33*(3), 239–251. https://doi.org/10.1023/A:1024446215446
Losh, M., & Gordon, P. C. (2014). Quantifying narrative ability in autism spectrum disorder: A computational linguistic analysis of narrative coherence. *Journal of Autism and Developmental Disorders, 44*(12), 3016–3025. https://doi.org/10.1007/S10803-014-2158-Y
Ma, S., Anderson, R. C., Lin, T. J., Zhang, J., Morris, J. A., Nguyen-Jahiel, K., Miller, B. W., Jadallah, M., Scott, T., Sun, J., Grabow, K., Latawiec, B. M., & Yi, S. (2017). Instructional influences on English language learners' storytelling. *Learning and Instruction, 49*, 64–80. https://doi.org/10.1016/J.LEARNINSTRUC.2016.12.004
Mayer, M. (1969). Frog, where are you?
Mayer, M. (2003). A boy, A dog, and A frog. Dial books.
Miller, J.F., & Nockerts, A. (2023). September 4). C-Unit Segmentation Rules.
Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language and Communication Disorders, 38*(3), 287–313. https://doi.org/10.1080/1368203100001081333

Norbury, C. F., Gemmell, T., & Paul, R. (2014). Pragmatics abilities in narrative production: A cross-disorder comparison. *Journal of Child Language, 41*(3), 485–510. https://doi.org/10.1017/S030500091300007X

Norbury, C. F., Nash, M., Baird, G., & Bishop, D. V. M. (2004). Using a parental checklist to identify diagnostic groups in children with communication impairment: A validation of the Children's Communication Checklist—2. *International Journal of Language & Communication Disorders, 39*(3), 345–364. https://doi.org/10.1080/13682820410001654883

Novogrodsky, R. (2013). Subject pronoun use by children with autism spectrum disorders (ASD). *Clinical Linguistics and Phonetics, 27*(2), 85–93. https://doi.org/10.3109/02699206.2012.742567

os: Python Software Foundation. (2023). os Module [Software]. ⟨https://docs.python.org/3/library/os.html⟩.

Parish-Morris, J., Liberman, M. Y., Cieri, C., Herrington, J. D., Yerys, B. E., Bateman, L., Donaher, J., Ferguson, E., Pandey, J., & Schultz, R. T. (2017). Linguistic camouflage in girls with autism spectrum disorder. *Molecular Autism, 8*(1). https://doi.org/10.1186/s13229-017-0164-6

Petersen, D. B. (2011). A systematic review of narrative-based language intervention with children who have language impairment. *Communication Disorders Quarterly, 32*(4), 207–220. https://doi.org/10.1177/1525740109353937

Python Software Foundation. (2023). Python 3. Version 12.0 [Software]. ⟨https://www.python.org⟩.

R Core Team. (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. ⟨https://www.R-project.org⟩.

Ripley, B. (2024). Support Functions and Datasets for Venables and Ripley's MASS [R package MASS version 7.3–60.0.1]. ⟨https://CRAN.R-project.org/package=MASS⟩.

Rutter, M., Bailey, Anthony, & Lord, Catherine (2003). *Social Communication Questionnaire (SCQ)*, 23.

Rynkiewicz, A., Schuller, B., Marchi, E., Piana, S., Camurri, A., Lassalle, A., & Baron-Cohen, S. (2016). An investigation of the "female camouflage effect" in autism using a computerized ADOS-2 and a test of sex/gender differences. *Molecular Autism, 7*(1), 1–8. https://doi.org/10.1186/s13229-016-0073-0

Searle, S. R., Speed, F. M., & Milliken, G. A. (2024). Estimated marginal means, aka least-squares means [R package emmeans version 1.10.1. *American Statistician, 34*(4), 216–221. https://doi.org/10.1080/00031305.1980.10483031

Sedgewick, F., Hill, V., Yates, R., Pickering, L., & Pellicano, E. (2016). Gender differences in the social motivation and friendship experiences of autistic and non-autistic adolescents. *Journal of Autism and Developmental Disorders, 46*(4), 1297–1306. https://doi.org/10.1007/s10803-015-2669-1

Song, A., Cola, M., Plate, S., Petrulla, V., Yankowitz, L., Pandey, J., Schultz, R. T., & Parish-Morris, J. (2021). Natural language markers of social phenotype in girls with autism. *Journal of Child Psychology and Psychiatry, 62*(8), 949–960. https://doi.org/10.1111/JCPP.13348

spaCy: Explosion AI. (2023). spaCy: Industrial-strength Natural Language Processing in Python [Software]. ⟨https://spacy.io⟩.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. *New Directions in Discourse Processing, 2*, 53–120.

Stirling, L., Douglas, S., Leekam, S., & Carey, L. (2014). Communication in autism. In *Trends in language acquisition research, 11*. John Benjamins,.

Sturrock, A., Adams, C., & Freed, J. (2021). A subtle profile with a significant impact: Language and communication difficulties for autistic females without intellectual disability. *0*, 3400 *Frontiers in Psychology*. https://doi.org/10.3389/FPSYG.2021.621742.

Sturrock, A., Marsden, A., Adams, C., & Freed, J. (2019). Observational and reported measures of language and pragmatics in young people with autism: A comparison of respondent data and gender profiles. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-019-04288-3

Sturrock, A., Yau, N., Freed, J., & Adams, C. (2019). Speaking the same language? A preliminary investigation, comparing the language and communication skills of females and males with high-functioning autism. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-019-03920-6

Sturrock, A., Chilton, H., Foy, K., Freed, J., & Adams, C. (2021). In their own words: The impact of subtle language and communication difficulties as described by autistic girls and boys without intellectual disability.doi:10.1177/13623613211002047.

Torsheim, T., Cavallo, F., Levin, K. A., Schnohr, C., Mazur, J., Niclasen, B., & Currie, C. (2016). Psychometric validation of the revised family affluence scale: A latent variable approach. *Child Indicators Research, 9*(3), 771–784. https://doi.org/10.1007/S12187-015-9339-X

van 't Hof, Tisseur, M., van Berckelear-Onnes, C., van Nieuwenhuyzen, I., Daniels, A., Deen, A. M., Hoek, M., H. W, & Ester, W. A. (2021). Age at autism spectrum disorder diagnosis: A systematic review and meta-analysis from 2012 to 2019. *Autism, 25*(4), 862–873. https://doi.org/10.1177/1362361320971107/ASSET/IMAGES/LARGE/10.1177_1362361320971107-FIG3.JPEG

Van Der Lely, H. K. J. (1997). Narrative discourse in grammatical specific language impaired children: A modular language deficit? *Journal of Child Language, 24*(1), 221–256. https://doi.org/10.1017/S0305000996002966

Van Wijngaarden-Cremers, P. J. M., Van Eeten, E., Groen, W. B., Van Deurzen, P. A., Oosterling, I. J., & Van Der Gaag, R. J. (2014). Gender and age differences in the core triad of impairments in autism spectrum disorders: A systematic review and meta-analysis. *Journal of Autism and Developmental Disorders, 44*(3), 627–635. https://doi.org/10.1007/S10803-013-1913-9

Volden, J., Dodd, E., Engel, K., Smith, I. M., Szatmari, P., Fombonne, E., Zwaigenbaum, L., Mirenda, P., Bryson, S., Roberts, W., Vaillancourt, T., Waddell, C., Elsabbagh, M., Bennett, T., Georgiades, S., & Duku, E. (2017). Beyond sentences: Using the expression, reception, and recall of narratives instrument to assess communication in school-aged children with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research, 60*(8), 2228–2240. https://doi.org/10.1044/2017_JSLHR-L-16-0168

Volden, J., & Phillips, L. (2010). Measuring pragmatic language in speakers with autism spectrum disorders: Comparing the children's communication checklist-2 and the test of pragmatic language. *American Journal of Speech-Language Pathology, 19*(3), 204–212. https://doi.org/10.1044/1058-0360(2010/09-0011)

Wechsler, D., Dannay-Penhouët, Caroline., & Éditions du Centre de psychologie appliquée (Paris). (2016). Échelle d'intelligence de Wechsler pour enfants: WISC-V.

Wiesner, D. (1991). Tuesday (Clarion Books). ⟨https://books.google.com/books/about/Tuesday.html?hl=fr&id=gGWvDUUi58cC⟩.

Wiig, E.H., Semel, E., & Secord, W.A. (2019). CELF 5: Évaluation des fonctions langagières et de communication (Éditions du Centre de psychologie appliquée (Paris)., Ed.).

Zane, E., & Grossman, R. B. (2023). Analysis of noun phrase ambiguity in narratives reveals differences in referential establishment but not cohesion for older autistic children. *Journal of Speech, Language, and Hearing Research, 66*(8), 2802–2820. https://doi.org/10.1044/2023_JSLHR-22-00630