UNIVERSITAT DE BARCELONA

Comprehension and production of referential expressions across Autism Spectrum Conditions

Kristen Schroeder

Comprehension and production of referential expressions across Autism Spectrum Conditions

Kristen Schroeder

A Doctoral Dissertation Submitted
in Partial Fulfillment of the
Requirements of the Degree of
Doctor of Philosophy
for the Doctoral Program
Cognitive Science and Language
Department of Catalan Philology and General Linguistics
Universitat de Barcelona

under the supervision of

Dr. Wolfram Hinzen
ICREA/Universitat Pompeu Fabra

Dr. Joana Rosselló Ximenes
Universitat de Barcelona

UNIVERSITAT DE BARCELONA

June 2019

Comprehension and production of referential expressions across Autism Spectrum Conditions

Abstract

Reference is a phenomenon both fundamental to communication and inherent to every act of language use. Autism spectrum conditions (ASC) are marked by problems of both communication and language. This thesis presents three studies of production and comprehension of reference in ASC.

Study 1 profiled the use of referential devices in narratives as part of a broader profiling of narrative language in ASC. Studies 2 and 3 tracked comprehension of grammatically different referential devices in a picture book task (the 'book task') and an interactive task which specifically tested for the comprehension of definite determiners (the 'box task'). Study 1 investigated autism without intellectual disabilities (ASC–ID, *n*=18) in comparison to typically developing controls (TD, *n*= 18) while in studies 2 & 3 recruitment was broadened to include the entirety of the verbal spectrum, including individuals with intellectual disability (ASC–ID, *n*= 34; ASC+ID, *n*=14) as well TD (*n*=34) controls and a control group of children with intellectual disability without ASC (ID, *n*= 9).

Results of study 1 showed that individuals with ASC–ID more frequently misused referential devices that imply specificity in comparison to TD controls, while the quantitative distribution of Noun Phrase (NP) types did not distinguish groups. The broader grammatical profile showed differences in grammatical complexity as well as rate of lexical errors while mental state content did not differ. Study 2, the book task, showed a relative weakness in comprehension of inflectional pronouns and clitics among participants with ASC. While the differences between the ASC–ID and TD groups only approached statistical significance, both ASC groups(+/-ID) showed a similar pattern of having significantly poorer comprehension of pronouns/clitics in comparison to non-anaphoric reference types, a pattern not found in the TD controls or the control

group with ID without ASC. The ASC+ID group showed a significantly poorer comprehension of inflectional pronouns/clitics in relation to both the ASC–ID group as well as the ID group, suggesting that such increased difficulties may not be explainable by presence of intellectual disability alone. Study 3, the box task, assessed the comprehension of definite anaphora in an interactive task using pairs of objects that belonged to the same kind category. In this task, the participant and experimenter interacted with an object (say, a ball) and then, after a moment of having their eyes closed, the participant pulled an object out of a box (either the same object or a noticeably different objects belonging to the same kind category, e.g. a different ball). The test condition assessed whether the participant interpreted correctly that in this context the definite-anaphoric NP *the ball* necessarily links to the ball that was interacted with previously rather than a novel exemplar of the same kind. The results showed that the ASC–ID group had a significantly poorer performance in the definite-anaphoric condition in comparison to TD controls. The ASC+ID group performed significantly worse than the ASC–ID group, yet similar to the ID group. These results show across comprehension and production that handling referential devices is an area of vulnerability in autism against the background of a broader spectrum of linguistic impairments.

Resum

La referència és un fenomen fonamental en la comunicació i inherent a qualsevol ús del llenguatge. La condició de l'espectre autista (CEA) es caracteritza per problemes de comunicació i de llenguatge. Aquesta tesi presenta tres estudis de producció i comprensió de la referència en la CEA.

L'estudi 1 caracteritza l'ús dels mecanismes referencials en narrativa com a part d'una caracterització més àmplia del llenguatge narratiu en la CEA. Els estudis 2 i 3 ressegueixen la comprensió de mecanismes referencials diversos des del punt de vista gramatical en un llibre il·lustrat sense gairebé text (la 'tasca del llibre') i una tasca interactiva que específicament provava la comprensió dels articles determinats definits (la 'tasca de la capsa'). L'estudi 1 investigava l'autisme sense discapacitat intel·lectual (CEA–DI, *n*=18) en comparació amb controls de desenvolupament típic (DT, *n*=18) mentre que en els estudis 2 i 3, per tal de cobrir íntegrament la CEA verbal, el reclutament va ampliar-se amb la inclusió d'individus amb discapacitat intel·lectual (CEA–DI, *n*=34; CEA + DI, *n*=14) així com de controls amb DT (*n*=34) i un grup de control d'infants amb discapacitat intel·lectual sense CEA (DI, *n*=9).

Els resultats de l'estudi 1 van mostrar que els individus amb CEA–DI usaven malament els mecanismes referencials que impliquen especificitat de forma més freqüent que no pas els controls amb DT mentre que la distribució quantitativa dels tipus de sintagma nominal (SN) no distingia els dos grups. El perfil gramatical restant rastrejat a l'estudi mostrava diferències en complexitat gramatical així com en la proporció d'errors lèxics mentre que el lèxic d'estat mental no diferia. L'estudi 2, la tasca del llibre, va fer palesa una relativa feblesa en la comprensió dels pronoms flexius i dels clítics entre els participants amb CEA. Mentre que les diferències entre la CEA–DI i el grup amb DT van fregar la significació estadística, tots dos grups CEA (+/–DI) van mostrar un patró similar pel que fa a una significativament més pobre comprensió dels pronoms/clítics en

comparació amb altres tipus d'SN referencial no anafòric, un patró inexistent tant entre els controls amb DT com en els controls amb DI sense CEA. El grup CEA+DI va mostrar una comprensió significativament més pobre dels pronoms flexius/clítics en relació tant amb el grup CEA–DI com amb el grup DI, la qual cosa suggereix que aquesta més gran dificultat no pot explicar-se per la presència de DI exclusivament. L'estudi 3, la tasca de la capsa, va avaluar la comprensió de l'anàfora definida en una tasca interactiva fent servir parells d'objectes pertanyents a la mateixa categoria pel que fa a la classe. En aquesta tasca el participant i l'experimentador interactuaven amb un objecte (una bolla, posem per cas) i aleshores, després d'estar amb els ulls tancats un instant, el participant treia un objecte de la capsa (que era o bé el mateix objecte previ o bé un altre objecte clarament diferent pertanyent a la mateixa classe, e.g. una bolla diferent). El que es testava era si el participant interpretava correctament que en aquest context, el nominal anafòric, *la bolla*, s'associava necessàriament amb la bolla amb què s'havia interaccionat prèviament més que no pas amb un nou exemplar de la mateixa classe. Els resultats van mostrar que el grup CEA–DI ho feia significativament pitjor que no pas els controls amb DT en la condició del nominal anafòric. El grup CEA+DI ho va fer significativament pitjor que el grup CEA–DI tot i que de forma similar al grup amb DI. Aquests resultats mostren a través de tasques de comprensió i producció que, entre d'altres dèficits lingüístics, la utilització dels mecanismes referencials és una àrea específicament vulnerable en l'autisme.

Acknowledgements

I would like to thank my parents Cynthia and Theodore Schroeder for your continual support throughout my education. Without both your emotional and at times financial support, I would never have been able to have come so far. I would also like to give my sincerest thanks to my partner, Otávio Mattos, who has cheered me on along the way and who has always been convinced that I am capable of great things. Specifically, I would like to thank him for all of the academic discussions and helpful feedback related to my research throughout my PhD, its not common to be blessed with both emotional and professional support in just one person!

I'd like to sincerely thank my supervisors Wolfram Hinzen and Joana Rosselló for your constant support throughout my PhD. I think I have truly been lucky to have learned from you and collaborated together for these years. I know I have grown so much as researcher since our first conversations at the onset of my PhD and I really attribute this growth to your mentorship.

I'd also like to thank all of my friends and flat mates who have supported me along the way and really have made my time here in Barcelona some of the best of my life (Dagmar Alvarez, Erika Antiche, Max Palmer, Natalia Fernández, Vincenzo Reina li Crapi, Marc Estape, Sànder Celma, Morteza Yazdani, Meritxell Ayguasanosa, Sophyann Raminosoa, Cristina Galusca, Cristina Hogan, Adrien Daube, Elise Allain, Guadalupe Vega, Ellie Jones, and Shayra Taveras, most all of you being a flat mate at some point!). I'd also like to thank all of the fellow members of the Grammar and Cognition Lab (Elisabet Vila, Dominika Slušná, Antonia Tovar, Clara Soberats, Kayla Chapin, Gabriel Sevilla, Andrea Rodríguez, Txuss Martín, Katia Tarantino and Cati Morey – it has really been a privilege working with you all and to see how our little group that began 6 years ago has now grown and advanced. It is not always a given to be able to develop your PhD research as a member of such a supporting, friendly and kind group of very smart people!

I'd especially like to give a big thanks to Eli, my partner in science! We've put in so many hours together and I know I really would not have reached this point without your support (also thank you Eli, as well as Martí, Munchies and Nit for hosting me for so many weeks during our data recruitment)! I would also like to thank my colleagues in the Linguistics Department at the University of Barcelona, including my Stel officemates, including Montse Nofre, Ruoyang Shi, Yuliang Sun, Qing Zhang, Maryam Ghaderi and Venelin Kovachev as well as faculty members Mariona Taulé, Antonia Martí and Itziar Aduriz for their support. I would also like to recognize the psychologists and research collaborators who helped me with passing assessments, transcription and annotations– thank you to Annabel Delgado, Miriam Garcia, and Inas Asaf. I would also like to thank Flavia Nicoletti for her illustrations that formed part of my experimental design in Study 2. A very meaningful thank you to the Aina Garí, Raimon Salvador and Derya Cokal for guiding me with my statistical analyses.

I would like to deeply thank all of the collaborating centers and administrators, psychologists, teachers and speech pathologists who made the recruitment of participants possible: COADI, Fundació Orienta, Escola Fàsia- Eixample, Escola Sants Innocents, Escola Tramuntana, and Escola Xarxa.

Finally, I would like to extend my deepest thanks to the children and adolescents who participated in this research as well as their families. It would not have been possible without you.

♦♦♦

This thesis was funded by the Spanish Ministry of Science, Innovation and Universities (BES-2014-069483) as a researcher in training for the project entitled: language, deixis, and mental disorders (FFI2013-40526-P).

Dedication

I dedicate this thesis to my parents Cynthia and Theodore Schroeder.

Table of Contents

Comment on Terminology

Throughout this thesis, I use the term Autism Spectrum Conditions (ASC) to refer to the autism spectrum. I use this term as synonymous to 'Autism Spectrum Disorder,' which is used in the Diagnostic and Statistical Manual, 5th edition (American Psychological Association, 2013), and to 'Autistic Disorder' which is the term currently used in the ICD-10 (World Health Organization) as well as to the general term 'autism'. This terminological choice follows the philosophy of researchers such as Simon Baron-Cohen of the Autism Research Center (Cambridge, UK) to recognize autism as a both a disability (i.e. with respect to communication and social interaction skills) and a neurocognitive difference that can also present certain strengths, such as attention to detail, etc.

I also refrain from using terms such as 'high-functioning' autism and 'low-functioning autism', when referring to sections of the autism spectrum that may or may not have intellectual disability (ID). As functioning levels may also refer to how well an individual is able to function independently in society, I considered this terminology a less precise term as an indicator of concurrent intellectual disability since many individuals with ASC without ID may require substantial support with regard to daily living skills. Within this thesis, I favor the terms ASC–ID for individuals with ASC without intellectual disability and ASC+ID for individuals with ID.

List of Abbreviations

Abbreviated term	Definition
ADHD	Attention Deficit Hyperactivity Disorder
ADI-R	Autism Diagnostic Interview- Revised
ADOS	Autism Diagnostic Observational Scale
APA	American Psychological Association
AS	Asperger's syndrome
ASC	Autism Spectrum Conditions (*synonymous to ASD*)
ASC–ID	Autism Spectrum Conditions without Intellectual Disability
ASC+ID	Autism Spectrum Conditions with Intellectual Disability
ASD	Autism Spectrum Disorder (*synonymous to ASC*)
BAP	Broad autism phenotype
CDC	Center for Disease Control and Prevention (US Governmental Agency)

CI	Confidence Interval
DSM-V	Diagnostic and Statistical Manual, 5th edition
ICD-10	International Classification of Diseases -10th edition
ID	Intellectual disability
MVNA	Minimally verbal or non-verbal autism
NP	Noun Phrase
OO	Optimal Outcome
PLI	Pragmatic Language Impairment
PPVT-III	Peabody Picture Vocabulary Test- 3rd edition
Pro	Inflectional pronoun/pro
SLI	Specific Language Impairment
TD	Typically Developing
TROG-D	Test for the Reception of Grammar (German version)
VMA	Verbal Mental Age
VP	Verb Phrase
WHO	World Health Organization

WISC-IV/V	Wechsler Intelligence Scale for Children $4^{th}/5^{th}$ edition
WLI	Wrong Lexical Item

List of tables

Chapter 2

Chapter 3

Study 2: Book task

Study 3: Box task

List of figures and illustrations

Chapter 2

Chapter 3

Study 2: Book task

Study 3: Box task

Chapter 1: General Introduction

This general introduction serves to introduce the topic of language in Autism Spectrum Conditions (ASC) in broad terms. This will provide a contextualization of my specific research questions in the two experimental chapters to follow, and specifically introduce the topic of reference, which is the main foundational concept on which these studies are centered.

Autism is a spectrum of neurodevelopmental conditions which (1) affect how individuals engage in and respond to social communication and social interaction and (2) involve repetitive behaviors and/or deep, yet restricted interests (Diagnostic and Statistical Manual, fifth edition [DSM-V], American Psychological Association, 2013). An ASC diagnosis is established through observation of behaviors that correspond to these two core symptom categories and may be reliable and stable from as early as 14 months of age (Pierce et al., 2019), though indicators of ASC may be observable from earlier in infancy. Several of these early indicators for ASC that raise concern among caregivers and physicians relate to falling behind on early language milestones, including a lack of canonical babbling, poor response to name, and delay in the onset of words, as well as an absence of gestural referential communication such as a lack of declarative pointing (Center for Disease control and prevention [CDC], 2019), which in typical development relates to early language development (Colonnesi et al., 2010).

The etiology of ASC is complex with a strong genetic component that is modulated by pre- and perinatal environmental factors (Gardener et al., 2011). Research on first degree relatives who show sub-clinical ASC traits, also known as the broad autism phenotype (BAP) is illustrative of inheritable components of ASC and aids in identifying endophenotypes which, in turn, can inform underlying neurocognitive markers and help bridge the divide from biology to behavior (Viding & Blakmore, 2007). Research on first degree relatives of children with ASC shows subtle

differences with regard to the parent or sibling's language profiles both in terms of developmental history and in language processing tasks. For example, in a study assessing 139 parents of children with ASC, poorer performance on language-related tests in their own childhood, as assessed through standardized test records, best predicted the ASC-symptom severity in their children as well as whether the parents would have a BAP profile in adulthood (Losh et al., 2017). Parents of children with ASC tend to have slower response time in rapid automated naming tasks, which are argued to tap into fluid cognitive and linguistic processes that underlie complex language skills (Nayar et al., 2018). Regarding siblings of children with ASC, language delay and language impairment is also one of the most robust indicators for BAP in siblings (Elsabbagh & Johnson, 2007) and, as a group, younger siblings of children with ASC tend to have reduced receptive language in early childhood in comparison to low risk peers (Garrido et al., 2017). Together this research supports the genetic liability to ASC as well as tunes our attention to the fact that language may play an important role in understanding the neurocognitive markers of ASC.

One of the striking characteristics of ASC is the heterogeneity that presents across individuals with ASC diagnoses, both in terms of the precise nature and severity of ASC symptoms as well as the wide range of symptoms that may commonly occur in ASC yet fall outside of the core diagnostic criteria. In a sample of size of over 2,500 children with ASC in the United States, 83% had a secondary developmental condition diagnosis or specification, such as language disorder, intellectual disability, sensory processing disorder, among others– the most prominent of which being language disorder (Levy et al., 2010). The definition of language disorder in ASC in the DSM-V is coarsely grained, however, such as not being able to speak in full sentences. Language related symptoms, on the other hand, such as atypical prosody, stereotyped language and difficulty using language in context are distributed across diagnostic categories. The clinical

distinction of defining ASC either with or without language disorder may therefore not capture more fine-grained differences in the individual's language profile or potentially how atypical language development in ASC may factor into ASC symptomatology.

Research on language in ASC indicates that language capacities are widely variable across the spectrum. As much as 30% of the spectrum do not acquire language and remain minimally verbal or non-verbal (MNVA) (Tager-Flusberg & Kasari, 2013) and may rely on alternative communication systems, such as pictograms to express needs. For these individuals, communication predominantly relates to imperative requests rather than declarative, referential communication (Maljaars et al., 2011). Individuals with MNVA may learn some words for fixed phrases, however rarely use language flexibly or outside of a habitual context (Kasari et al., 2014). Word learning among this low-verbal population has been shown to involve atypical learning patterns, such as associative mechanisms in word-picture-object mapping tasks, unlike in typically developing children who intuit the referential nature of words and generalize word meaning to novel referents rapidly (Hartley & Allen, 2015; Preissler, 2008; Allen Preissler & Carey, 2004). For verbal individuals with ASC and intellectual disability (ASC+ID), language impairment is pervasive both in terms of lexical development and grammar (Boucher et al., 2012), though see Silverini et al. (2017) for a discussion of individuals with ASC and nonverbal IQ below norms, without formal language impairment. Among individuals with ASC+ID (including MVNA), language comprehension tends to be weaker than productive language skills, contrary to the pattern in typical development (Maljaars et al., 2012). Although MNVA and ASC+ID constitute more than half of the autism spectrum (Baio, 2014), relatively little attention has been dedicated to exploring language profiles in this population in a fine-grained way. This in part is motivated by the relative difficulty to assess these populations as well as the conceptualization that by looking

at ASC without ID (ASC–ID), we would be able to more accurately assess ASC–specific language impairments that may not be attributable to secondary concurrent conditions such as intellectual disability. However, provided the neurodevelopmental nature of ASC, which shows early atypicalities in language development alongside early atypical communication and social interaction, it may be highly informative to our understanding of ASC to attend to the kinds of variability that presents across the entire autism spectrum. That is to say, by assessing language profiles across the spectrum, we may gain a deeper understanding of the nature of language development and processing in ASC as a whole.

Among children with ASC without intellectual disability (ASC–ID), research on language production and comprehension has been more widespread, although, overall, linguistic research on ASC remains relatively scarce. In this ASC–ID population, language capacities are also variable, though severity may relate in part to early language development profiles. Noterdaeme et al. (2010) assessed language capacities including receptive and productive language, intonation, echolalia and pronoun reversal among 57 children with ASC–ID who specifically did not have a history of language delay, also known as Asperger's syndrome[1] (AS) (mean age: 11;2), and 55 children with ASC–ID (mean age: 10;6). Expressive and receptive language was assessed through standardized assessments such as the Test for the Reception of Grammar (TROG-D) and rating of spontaneous speech while the measures of echolalia, intonation and pronoun reversal were assessed through the Autism Diagnostic Interview- Revised (ADI-R) and observation. The two groups both had full-scale IQ scores within the normal range (above 80) yet differed in terms of

[1] Asperger syndrome has been incorporated into the Autism Spectrum Disorder diagnostic category in the DSM-V, however it remains a differentiated diagnosis in the ICD-10. Asperger's syndrome is differentiated from Autism Spectrum Disorder in that there is no history of cognitive or language delay.

history of language delay that was present in the ASC–ID group. The ASC–ID group presented significant deficits in expressive and receptive language, with only 23% falling within the normal range. The ASC–ID had significantly greater deficits in all language areas assessed in comparison to the AS group, except for abnormal intonation, which was impaired nearly universally in both groups (94.7% of the AS group and 92.7% of the ASC–ID group). Although the AS group had significantly better language performance than the ASC–ID group, as much as 30% still had substantial deficits in receptive language.

In a similar vein, Durrleman et al. (2015) assessed comprehension of complex syntax, particularly relative clauses in adults with ASC–ID either with language delay (n= 10) or without (i.e. an AS profile) (n=18), finding weakness in comprehension of object relatives in both groups in comparison to 28 age-matched typically developing adults. The authors also identified differences between ASC–ID groups in that the adults with ASC–ID with a developmental history of language delay also had significantly poorer comprehension of subject relatives in comparison to the group with an AS profile. In development, subject relatives such as *the dog that is biting the cat* are acquired earlier than object relatives, such as *the cat that the dog is biting*. As such we see that both ASC groups have lingering difficulty in adulthood in the later acquired and more complex object relatives while adults with ASC–ID and a history of language delay demonstrate greater weakness in a wider swath of grammatical constructions. Another recent study assessing sentence comprehension using fMRI found that ASC–ID participants (aged 8-21), who no longer showed symptomatic language impairment utilized atypical neural networks both through greater activation of regions implicated in language processing as well as right hemisphere homologues (Eigsti et al., 2016; Moseley et al., 2016; see also Mizuno et al., 2011 for similar findings comprehension of deictic expressions in adults with ASC). This suggests that even individuals

with ASC–ID who achieve high levels of language competence may be arriving at this achievement by using greater cognitive resources rather than a typical trajectory.

While the exact distribution and nature of structural language impairments in ASC remains an active area of investigation, there is wide agreement that the contextual and social use of language in interpersonal interactions is universally impaired across the spectrum (e.g. Volden, 2017), both among individuals with and without specification for language disorder or intellectual disability. Much of the research within this broad topic focuses on how language is used in free conversation or other verbal interactions and therefore relates more closely to the communicative deficits, which form part of ASC diagnosis. Conversational analyses show that children with ASC tend to have greater difficulty providing an adequate contingent response that 'meshes' with the previous conversational turns (Adams et al., 2002), and although individuals with ASC may be sensitive to the need to engage in conversational repair to ensure that they understand and are understood, children with ASC tend to be less successful in doing so (Volden, 2004). Children with ASC also have difficulty flexibly changing their language use to adapt to the listeners' needs. For example, when prompted to rephrase their speech as their interlocutor was said to not understand (e.g. being told, '[the puppet] still does understand, say it really simply'), the children with ASC did tend to adjust their speech, however they did so less adeptly than typically developing peers (Volden et al., 2007).

A recent study by Wiklund & Laakso (2019) assessed communication breakdown between preadolescent peers with AS. The peer-to-peer conversation was assessed both with attention to grammatical errors and disfluencies that arose in the participants' speech as well as how and when communicative breakdown occurred and whether the participants engaged in repair or clarification requests. Interestingly, the authors found that the participants were adept at requesting clarification

but that the participants presented significant grammatical and referential errors (e.g. case-marking errors, inaccurate tenses and ambiguous pronouns) as well as speech disfluencies, triggering comprehension failure among their interlocutor peers. The authors suggest that the AS participants in their study may have substantial language difficulties that they try to resolve and self-repair while speaking, thereby generating less coherent morpho-syntactic and disfluent sentences. This line of research suggests an interplay between difficulties in structural language and communicative competence, which should be explored further.

In this regard, a crucial feature of conversation is that it involves a triadic form of communication, which in addition to the dyadic dynamics between two individuals involves an element of referential meaning, where referents can be objects, events, or ideas. Referential meaning connects communication to grammar, as differences in referential meaning are indexed by and depend on different grammatical configurations involved in generating such forms of meaning (Hinzen & Sheehan, 2015). That is to say, the way in which entities are referred to, e.g. via indefinite NPs, definite NPs, or pronouns, index different kinds of referential meaning and function. This can include distinguishing novel entities through indefinite NPs (e.g. I saw *a* food truck) from those in the 'common ground' through definite NPs or by pronouns (*the food truck* /*it* looked popular), or entities of one type in general (e.g. *food trucks were everywhere*). As such we may view the topic of reference in ASC as bridging discussions of formal aspects of language in ASC and the communicative and contextual context in which they are deployed. An important concept under which reference has already been studied so far in ASC is cohesion. Studies of cohesion in ASC have assessed how sentences coordinate with each other in a discourse. In one early study (Baltaxe & D'Angiola, 1992), individuals with ASC were found to link their sentences through fewer cohesive ties relative to controls, as well as to make more cohesive errors that

indicated poor integration of sentences in a discourse. However, thus far there is no comprehensive assessment of reference in ASC from a linguistic perspective.

Referential meaning in communication connects with our earlier discussion of early indicators of ASC, where declarative pointing is particularly prominent and, as noted earlier, related to vocabulary and grammar growth in typical development (Iverson & Goldin-Meadow,2005). In addition to these findings, reduced initiation of declarative pointing in toddlers correlates with overall symptom severity scores in ASC (Ibañez et al., 2013). Moreover, as already noted, declarative referential meaning is absent in as much as 30% of the spectrum (Slušná et al., 2017) and therefore is a critical aspect of the heterogeneity presents across the spectrum. Reference, as a link between formal language and communication, thus warrants a dedicated exploration, as there has been little research to date which assesses reference across the autism spectrum in a comprehensive manner.

This thesis does not attempt to cover the complexity of reference in autism across all of its dimensions. However, within this thesis I will profile this notion linguistically, across both production and comprehension, and across the entire verbal part of the autism spectrum. In Chapter 2, I explore language production among children with ASC-ID, specifically assessing referential language use as part of a fine-grained profiling of language in a narrative story-telling task. This will allow us to study in detail how children with ASC construe their language and refer to characters and objects. We will also assess the types of errors that are produced, with a specific regard to referential failures. In Chapter 3, I present two studies that assess comprehension of referential NPs among children with ASC both with and without intellectual disability and in comparison to TD as well as children with intellectual disability, without ASC (ID). The scope of the studies developed in Chapter 3 explores the comprehension of referential NPs that differ in

terms of their different grammatical configurations, such as whether they are definite or indefinite, lexical or pronominal. As explained in Chapter 3, such types of NPs systematically differ in their grammatical specifications along with their semantics and use, allowing the study of how grammar and the referential function of language connect in ASC. Although the studies presented in Chapters 2 and 3 are designed as being independent (though linked), Chapter 4 makes an attempt to discuss their results in the context of both of them together.

Chapter 2: Fine-grained linguistic analysis of storytelling in ASC without intellectual disability

2.1 Introduction

When we tell a story we are engaging in a linguistically rich form of social interaction. Unlike conversation, the speaker narrating a story bears primary responsibility for discourse construction and guiding the listener to track characters and events in a coordinated way through the language he or she uses. Narrating a story integrates language, cognitive capacities and social interaction and is a competence that matures throughout development (Norbury, Gemmell and Paul, 2014).

In recognition of this multifaceted cognitive nature of storytelling, narrative elicitation forms part of the standardized autism diagnostic assessment, the Autism Diagnostic Observation Schedule (ADOS). The narrative task in the ADOS evaluates the individual's sense-making abilities and how the child refers to social interactions or emotions that may be represented in the story, as deficits in social communication are one of the defining features of ASC (Diagnostic and Statistical Manual (DSM-V), American Psychological Association, 2013). Though assessing language is not a primary goal, the ADOS generates rich language samples that, through analysis, may provide greater understanding of language profiles in ASC.

As overviewed in Chapter 1, language impairment is widely variable in ASC though the nature and extent of structural language anomalies is still under active investigation. Where there is an absence of structural language deficits as determined by standardized language assessments, deficits in language production are generally attributed to difficulties integrating language within the speech context (i.e. pragmatic deficits). However, whether the origin of such difficulties are deficits in social cognition (e.g. Colle et al., 2008), or whether they are due to residual differences in language dys/functioning is an open question. Narrative elicitation tasks are an effective and

accessible means of identifying more subtle linguistic abnormalities that may go undetected in standardized language assessments or unstructured conversation tasks (Botting, 2002; Manolitsi & Botting, 2011), and can illuminate their cognitive basis. Narratives specifically allow us to assess various aspects of language in an integrated way, including (1) cataloguing the repertoire of word choices and grammatical devices used to construct the discourse, (2) assessing error patterns within the lexicon and grammar, and (3) tracking the way in which utterances coordinate with each other – both in terms of how characters and objects are referred to throughout the discourse (e.g. through the selection of indefinite or definite noun phrases, NPs) as well as in terms of the use of tense and aspect in the verbal domain to knit together richly construed and sharable novel realities. Finally, narratives allow us to (4) track disfluencies such as repetitions and self-corrections, which may be subtle indicators of greater effort in language processing (Arnold & Griffen, 2007). The aim of the present study was to utilize this clinical tool to more systematically identify the language profile(s) of ASC through a more fine-grained, linguistic lens across these four dimensions of linguistic organization.

2.1.1 Narratives indicate referential language as a point of weakness in ASC

A wealth of studies have assessed narrative competence in ASC through a variety of narrative task procedures, with the aim of better profiling language in the autism spectrum as well as identifying similarities or contrasts to other clinical groups, such as specific language impairment (SLI) or attention deficit hyperactivity disorder (ADHD) (e.g. Norbury and Bishop, 2003; Rumpf, Kamp-Becker, Becker and Kauchke, 2012). Narrative elicitation tasks have either involved: (1) a re-tell design in which the participant is asked to recall a familiar story or is shown a brief sequence of story panels and asked to retell the narrative after viewing the story from start to finish; or (2) a storytelling design in which the child is presented with a wordless picture book

to narrate. The type of task design has been shown to affect performance due to differing effects of memory, planning or facilitation effects of visual support (Pearce, 2003; Novogrodsky, 2013). The focus of our review will be the latter, as it is the design used in the ADOS and hence, our present study. Regarding narrative analysis, Norbury and Bishop (2003) in their seminal study distinguished looking at the macrostructure – the pattern of initiation and resolution of events – and also the micro structure – the formal aspects of grammar, including syntax, morphology, and anaphora. Their study assessed participants with autism without intellectual disability (ASC–ID, n= 12); specific language impairment (SLI, n=17), pragmatic language impairment[2] (PLI, n=21) and typical development (TD, n=18). According to diagnostic categorization, PLI is similar to ASC–ID in terms of social-communicative deficits though lacking the repetitive behaviors and restricted interests that are necessary for an autism diagnosis. One finding was that the rate of tense marking errors did not significantly differ between SLI and ASC groups, though only the SLI group was significantly different from TD controls. Both SLI and ASC–ID participants, however, produced syntactically less complex narratives in comparison to TD controls. The PLI group did not differ from either the TD controls or the ASC–ID and SLI groups in terms of tense marking or syntactic complexity, while all clinical groups showed reduction in the macrostructure. These findings suggest that despite discrete diagnostic categories, narrative competence and indeed language profiles show substantial overlap between groups and as such motivates a further exploration of underlying processes.

Within the past several years, various studies have utilized the narrative task of the ADOS to assess narrative competence in ASC, specifically using the storybook *Tuesday* (Weisner, 1991),

[2] Pragmatic language impairment is currently classified as Social (Pragmatic) Communication Disorder in the DSM-V and ICD-10 (World Health Organization, 2019) and distinguished from ASC.

which is the story book option intended for children with fluent phrase speech, three of which (Rumpf et al 2012, Banney et al. 2015, and Suh et al 2014) will be reviewed here. This is also the same storybook that is used in our present study. Rumpf et al. (2012) assessed the narratives of children with Asperger's syndrome[3] (AS, n=11) and found no distinctions in terms of syntactic and productive measures in comparison to TD controls (n=11) or children with ADHD (n=9). In contrast, Banney, Harper-Hill, and Arnot (2015) found that their ASC participants (n=11) produced narratives that were significantly reduced in syntactic complexity in comparison to TD controls (n=17) on narratives. These studies along with Norbury and Bishop (2003) suggest that ASC–ID may show greater language processing difficulties than participants with diagnoses that are often deemed to be 'less severe' in terms of absence of language delay (as in AS) or in the extent and range of ASC-like symptoms (as in PLI). A notable commonality between these studies were referential anomalies as a distinguishing feature in their respective ASC populations, in the form of a reduction of pronouns in the AS group in relation to TDs and children with ADHD[4] in the study by Rumpf et al. (2012), an increase of ambiguous pronouns in the study by Banney et al. (2015), as well as in an increase of both ambiguous nouns and pronouns in Norbury and Bishop (2003).

A recent meta-analysis by Baixauli et al. (2016) assessed results from 24 narrativity studies of children and adolescents with ASC–ID and found that all linguistic variables under consideration were reduced in the ASC groups in comparison to TD controls, including utterance length and lexical diversity, coherence and referential cohesion, and a reduction of mental state

[4] While the ASC group quantitatively produced less pronouns than the ADHD group (p=.048) this difference was not considered significant due to the Bonferroni correction which lowered threshold for significance to p=.017).

language. Greatest effect sizes were found in temporal and causal organization of propositions, which the authors labeled as problems of 'cohesion' (standard mean difference = 0.79; 95% CI = 0.49-1.09); and referential accuracy to characters throughout the narrative, which the authors called 'cohesive adequacy' (standard mean difference = 0.93; 95% CI= 0.61-1.26). Therefore, we find that narratives are sensitive in detecting a broad range of differences in language profiles while highlighting troughs of weakness – particularly referential anomalies – within the terrain of reduced language capacities.

One subgroup of the autism spectrum that was not found to produce referential anomalies in storytelling tasks is the 'optimal-outcome' group, assessed in a study by Suh et al. (2014). Individuals who are deemed optimal outcomes have a history of ASC diagnosis yet no longer reach thus diagnostic threshold based on current assessments. Suh et al. (2014) compared the narratives of these optimal outcomes participants (OO, n= 15) to children and adolescents with ASC–ID (n= 15) who met diagnostic criteria for ASC at the time of the study. They found that only very subtle markers distinguished OO participants from their TD peers, primarily in the rate of self-corrections and idiosyncratic speech. The ASC–ID group, similar to the ASC–ID participants in Banney et al. (2015), produced more ambiguous pronouns as well as showing greater disfluencies and greater difficulty making sense of the storyline. Together, these studies suggest that referential anomalies may dissipate alongside autism symptomatology, while greater prevalence of self-corrections than TD peers may suggest that self-monitoring of speech may help to reduce communicative breakdown.

In sum, previous literature indicates that elicitation of narratives is a sensitive tool to assess language anomalies across broad domains of language – from assessing grammatical devices to tracking disfluencies such as self-corrections or repetitions, which may indicate more effortful

language processing (Arnold and Griffen, 2007). The previous literature has specifically highlighted referential language as an area of interest for distinguishing participant groups with ASC from peers without ASC. The studies to date that have assessed narratives through the *Tuesday* picture-book have shown that pronouns are an area of weakness among symptomatic ASC children, either through an overproduction of ambiguous pronouns (Banney et al., 2015) or a reduction of pronouns in favor of full definite NPs (Rumpf et al. (2012) relative to TD peers. In the study by Banney et al. (2015), reference to characters was assessed through a measure of 'cohesion,' which tracked the means of introduction and maintenance of novel characters (through indefinites NPs, definite NPs, nouns only, pronouns, or omission). This variable of cohesion did not come out as significantly different between the ASC and TD groups, which may be due to a lack of statistical power due to relatively small group sizes (ASC, n=11) yet also to the particular nature of the *Tuesday* book as a narrative prompt. As an illustrated prompt, *Tuesday* facilitates situational reference (to entities in the shared space with the experimenter) rather than the textual reference (i.e. a linguistic discourse representation), therefore allowing for a more flexible introduction of new referents through definite noun phrases, a phenomena characteristic of child language until approximately until approximately age 9 (Schneider & Hayward et al., 2010). While these studies have explored syntactic complexity in a broad sense, coding for grammatical and referential errors has yet to receive a more fine-grained exploration within the context of this task. Norbury & Bishop (2003) tracked tense-marking errors in their ASC population, finding that they produced these errors more than TD peers and similar to children with SLI. In the studies by Banney et al. (2015) or Rumpf et al. (2012), grammatical errors were either not under exploration as in the case of the latter or rather coarse-grained as in the former, such as assessing whether an error occurred at the lexical or clausal level. Furthermore, previous research has focused on the

nominal domain leaving the verbal domain relatively unexplored. A more fine-grained analysis of both the variety and proportions of different grammatical construction used at the nominal and clausal level, as well as a more detailed classification system of error types across these domains, with larger group sizes, may help shedding further light on the nature of the language difficulties that children with ASC face, and their cognitive basis.

2.1.2 Aims of the present study

The aim of the present study was to address these open questions in children with ASC without ID using the narrative production task of the diagnostic assessment, the ADOS, with a focus on three domains of cognitive control over language processes: 1) How entities are referred to throughout the discourse (e.g. the selection of indefinite or definite-anaphoric and definite-non-anaphoric reference to entities in the nominal domain) while remaining neutral to any potential errors; 2) How scenes are elaborated within the verbal domain either through a richer episodic structure or more descriptive stative structure; 3) Grammatical complexity as assessed by clausal embedding, namely relative, complement and adjunct clauses; 4) Error patterns in terms of lexical choice, referentiality, as well as any other grammatical error; 5) Disfluencies such as repetitions and self-corrections. We also assessed the degree in which the children were able to grasp the story structure and comment on key characters and events. Our motivation to assess these narratives in children with ASC without ID (ASC–ID) was to be able to explore more deeply points 1-3 while still being able to draw comparisons to the previous literature, which has focused thus far on children without intellectual disability.

Analyzing the narratives generated during the ADOS in particular allows us to connect more directly with several previous studies such as Rumpf et al. (2012), Suh et al. (2014), and

Banney et al. (2015), which have also utilized this standardized task. Our study shares commonalities with previous studies, yet also strikes differences in creating more fine-grained error categories that fall along linguistic distinctions. Previous work focused more on the presentation type of errors as being an omission, a word level error, or a multiword level error (e.g. Banney et al. 2015). Our coding also aims to provide a more complete picture of grammatical profiling by also comparing use of verb type and fine-grained embedded clause types, as embedded clauses have been identified in previous studies such as Rumpf et al. (2012) to be the most frequent complex clause type in comparison to other clause types such as passive constructions, etc. Our study also provides a language sample in Spanish which provides further cross-linguistic data and which is under-represented in studies on language in autism.

2.1.3 Predictions of Study 1

Regarding our predictions, as referential language has been demonstrated to be a particular trough of weakness in ASC, we hypothesized that definite anaphoric NPs would be particularly affected in the ASC participants in relation to TDs. We specifically predicted, for the children with ASC relative to TD controls:

- A reduction of production in anaphora and in particular third-person clitic pronouns.
- More referential errors, particularly instances of failed reference.
- A reduction in mental state or emotional content and grammatical complexity as measured by embedding, in line with findings in previous studies (e.g. Siller et al., 2014).
- Greater disfluencies, which may indicate greater cognitive load related to language planning (in line with Arnold & Griffen, 2007).
- Fewer key elements of the story structure than in typically developing peers.

We also included an exploratory verbal domain variable, as little attention has been made to date with regard to verb structure.

- In particular, we predicted fewer non-stative (episodic) verbs in relation to TD peers, as children with ASC may favor describing the state of affairs rather than developing a plot.

To explore the cognitive basis of these anomalies, we also ran correlations with verbal IQ, as assessed by the Peabody Picture Vocabulary Test (PPVT-III) as well as with IQ measures (IQ, working memory and processing speed) in the ASC case.

2.2 Methodology

2.2.1 Participants

18 children (mean age: 9;11, range: 7;4-12;6 years old) diagnosed with ASC participated in this study. All children were native speakers of Spanish and Catalan, recruited from juvenile mental health clinics in the Baix-Llobregat county of Barcelona, a region that tends to be Spanish-dominant. Participants were recruited based on an autism spectrum diagnosis and being able to speak in at least two-word utterances. To ensure that the participant fulfilled an ASC profile, a further inclusion criterion consisted of scoring positive for ASC in either the ADOS or the Autism Diagnosis Interview- Revised (ADI-R). The ASC participants were further passed the Wechsler Intelligence Scale for Children (WISC-IV) to insure that their non-verbal IQ was above the threshold for intellectual disability (IQ ≥70). The ASC participants were individually matched with 18 typically developing controls based on verbal IQ, as measured by the PPVT-III, which is a measure of receptive vocabulary. Controls were identified as being a match if they were within one chronological year of age and five verbal IQ points. Independent t-tests showed that there were no significant differences between groups on verbal IQ ($t(34) = -.112$; $p = .911$) and age ($t(34)= 316$; $p = .754$) (see Table 1).

Table 1. *Participant Information*

	ASC (n=18)	TD (n=18)	t	*p*
Verbal IQ	99.33 55-120	99.94 55-122	-.112	.911
Chronological Age	9;11 7;4-12;6	9;9 6;8-12;10	316	.754

2.2.2 Narrative procedure

All participants were tested individually in a quiet room and video-recorded so that their narrative production could be later transcribed and annotated. The participants were presented with the picture book, *Tuesday* (Weisner, 1991). This book is comprised of 30 pages with colorful images without text, though words occasionally appear on adjacent pages to add time frames such as '4:38 am'. Following ADOS administration procedure, the examiner began narrating the first page and then allowed the child to continue the narration until the final pages when the experimenter would take over again.

Tuesday presents a story of how, on a Tuesday night, a group of frogs levitated out of their swamp on lily pads and flew over a nearby town at night, startling unsuspecting animals and people while zipping past unnoticed by others. When day broke, the frogs lost their ability to fly and fell from their lily pads, hopping back disgruntledly to the swamp. The following day, police and reporters tried to solve the mystery of the many wet lily pads strewn all over the town.

As there was some variability between ADOS examiners with regard to when they indicated the participant to begin and end, we normalized the narratives by beginning annotations from when the frogs began flying towards the town and ended when they hopped back to their swamp.

2.2.3 Coding Scheme

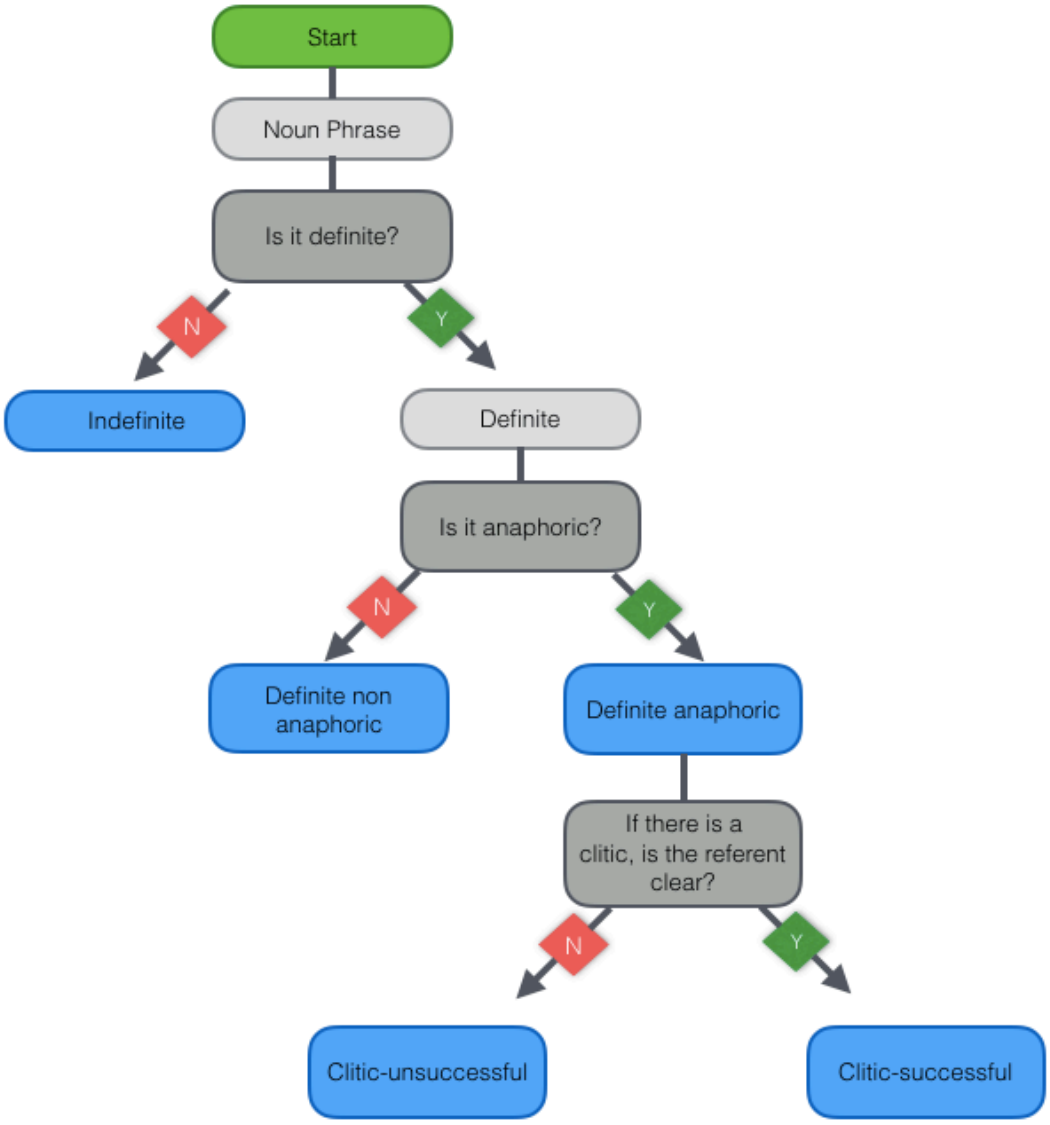
Grammatical Profiling

The narratives were transcribed and annotated using CLAN. All narratives were divided into utterances, which were defined as self-standing units of discourse providing new information. For example, if a child said 'they're flying, they're flying' this would be considered a single utterance while 'they're flying, they're flying all over the town,' would be classified as two utterances. In the interest of profiling how the ASC participants refer to entities in the story and maintain reference, we categorized each NP according to NP type, as outlined in the flowchart in Figure 1. First, NPs were first classified as indefinite (e.g. *some frogs*) or definite (e.g. *the frogs*; *he*). Definite NPs were further classified as anaphoric (e.g. *they*; *the frogs she had seen*) or not (e.g. *this/a policeman*). Clitic pronouns were further tagged as +clitic in addition to being classified as definite-anaphoric, with a further judgment of whether they were 'successful' or 'unsuccessful', depending on whether the clitic was involved in a referential error or not. Errors were coded in order to be able to contrast the ratio of referential errors that were pronominal in nature. As such, clitics were tagged both as 'definite-anaphoric' and as either 'clitic-successful' or 'clitic-unsuccessful'. Regarding the verbal (VP) domain, finite verbs were classified as to whether they were stative or not (here called episodic). We also annotated utterances for syntactic complexity, based on tagging clausal embedding, distinguishing syntactically between relative clauses, complement clauses, and adjunct clauses. General productive measures such as rate of words per utterance as well as type token ratio were also computed.

As the length of narratives varied across participants, frequencies of all variables were normalized by the total number of words, with the exception of embedded clauses, which were normalized by the total number of utterances. We also compared the proportional use of each

grammatical device type relative to its domain, creating nominal, verbal, and clausal domain ratios. For example, the definite-anaphoric NP ratio was calculated by dividing the total number of definite-anaphoric NPs by the total number of NPs.

Figure 1: Noun phrase coding flowchart



Errors

In addition to tracking the way in which entities were referred to, we also coded errors within the nominal (presented in Tables 2 and 3) and verbal domains (Table 4). In these tables, nominal domain errors are defined while names of verb domain errors entail their definitions. We also created composite score variables for group comparisons, specifically: total errors, total lexical errors, total grammatical errors, total verb errors, total nominal errors, and finally total nominal reference errors. The total error coding plan with composite and subordinate variables is represented in Figure 2.

Table 2. *Nominal Domain Errors: Lexical errors*

	Definition	Example (extracted from transcripts)	English gloss/ comments
Wrong Lexical Item (non-relational)	(1) Imprecise content word, which may either be incorrect such as such as saying *parachute* rather than *cape*, or overly general such as saying *clothes* rather than *cape* (in the context in which the frogs wear a cape like that of superman). Or (2) neologism (e.g. superdonde for a frog in a cape, like a superhero)	*y luego los sapos cogen **una alfombra***	And then the frogs take a **carpet** Comment: the frogs actually get caught in a bed sheet.
Wrong Lexical Item (relational)	Imprecise word choice, particularly in relation to functional words such as prepositions, conjunctions, etc.	*Ladrarlos **a** perseguirlos*	Bark them **to** chase them (**a* rather than *para*) Comment: In this case, the preposition *a* is incorrect as it does not impart intention. The correct preposition in this case would be *para. We also note a secondary grammatical error here, which is addressed in Table 3 'verb error'.*

Table 3. *Nominal Domain Errors: Grammatical errors*

Variable	Definition	*Example (from transcripts)*	English gloss and comments
Vague Reference	Reference to entities that is too vague in nature to allow the listener to capture the scene well.	*Aqui pasa algo*	Something is happening here
Failed Specific Reference	Misuse of pronouns or person inflection, or whenever it is unclear who or what the intended referent is (e.g. saying 'the other one' when there are several potential referents to disambiguate or introducing a new referent with an implicit pronoun ([she-$_{PRO}$] is sleeping). Note: Pointing can salvage errors of unclear referent. Such that if the child utters 'the little frog said the same thing to the other one*' while pointing to another frog, this utterance will not be marked as an error.	*'que viene un perro y se asusta'*	A dog comes and gets scared Comment: In the book, it is the frog who gets scared.
Definite/Indefinite mismatch	Use of a definite or indefinite article in an inappropriate context. Note: Introducing characters with a definite article is not generally infelicitous with pictorial support (Martin, 1978).	*'se quedaron horas y horas viendo la televisión. la rana la rana levitadora iba tranquilamente por el jardín'*	[the frogs] stayed hours and hours watching television. *The frog the flying frog went slowly through the garden.[5]
Noun agree error	Error in noun agreement (e.g. gender or plural).	Y por la mañanas cayeron al agua.	And in the$_{singular}$ mornings$_{plural}$ they fell in the water
Verb error[6]	Error in verb agreement (e.g. person or number) or licensing errors such as making a non reflexive verb reflexive)	*Ladrarlos [ladrar a ellos]*	*Bark them [bark at them]

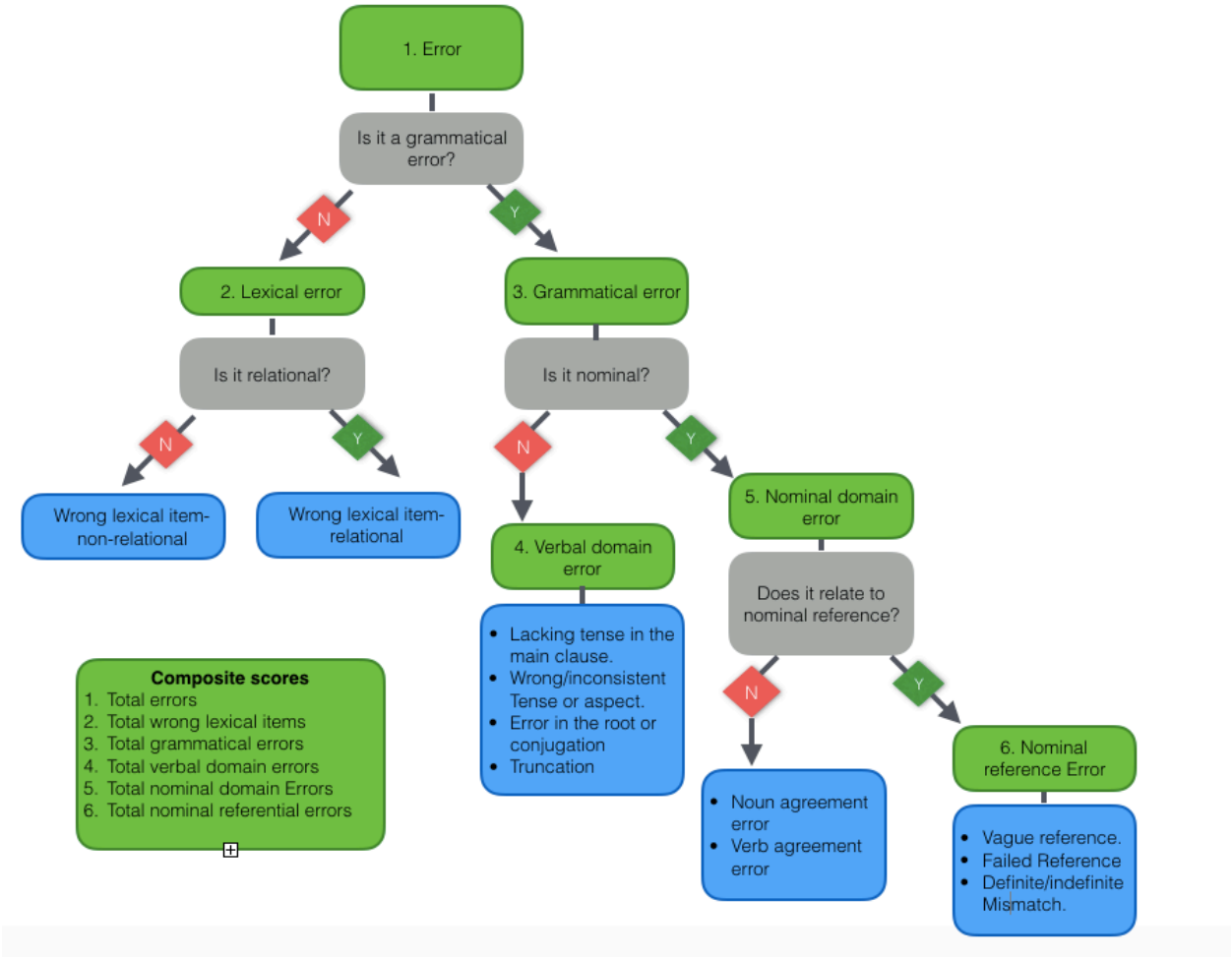
[5] In this case, the definite article is infelicitous as the child is referring to one frog from the previous established set of flying frogs.

[6] We coded verb agree errors as part of nominal errors due to the fact that in pro-drop languages, such as Spanish, verb inflection is the morphological person marker and as such relates to the nominal domain.

Table 4. *Verb Domain Errors*

Verbal Domain Errors	Example	English Gloss
Lacking tense in the main clause	*Un señor cenando y mira los sapos pasar.*	A man eating and looks at the frogs passing.
Wrong or inconsistent tense or aspect	*Cuando vuelan los sapos en las hojas encontraron un pequeño pueblito*	When the frogs fly they found a tiny little town.
Error in the verb root or conjugation	**ponieron (pusieron)*	*They putted (they put)
Truncation	*Se ha con la manta*	The had with the blanket

Figure 2: Error coding flow chart



Mental state content

All mental state or emotion words were annotated and marked as being either 'good' if used correctly or plausible or 'wrong' if there is a clear mix-match between the emotion or mental state described by the participant and what is depicted in the illustration. For example, on one page a man is eating a sandwich when he notices the frogs flying outside his window – the expression on his face is surprised or perhaps confused. If the child said that the man was scared or angry, this utterance would be tagged with the code 'mental state content- wrong.'

Fluency measures

Regarding fluency measures, we annotated instances of repair, such as reformulations and self-corrections, as well as repetitions. Repetitions were classified as single word repetitions, phrase repetitions, or 'other' which included both partial word repetitions and multi-word repetitions that were not discrete phrases. Self-corrections were also classified as being either 'good' or 'bad', in order to capture the phenomena of participants occasionally correcting themselves, while either creating a novel error or maintaining one. Reformulations differed from self-corrections in that in the former there was no identifiable error prior to the reformulation.

Story macro-structure

Following Norbury & Bishop (2003) we established a set of propositions that are necessary to adequately tell the story, and rated the narratives based on the occurrence and completeness of these propositions on a 0-2 scale for a variable we called Story Points. Zero points were awarded when the proposition was absent, 1 point if not fully mentioned and 2 points if fully mentioned. A total of 10 propositions were generated, resulting in a total of 20 possible points.

Reliability

Video recordings were transcribed by a first transcriber, Miriam Garcia Subirats, then subsequently edited by a second transcriber, Inas Assaf, in consultation with myself. Ten and a half percent of the transcripts were selected randomly via a random number generator and were independently transcribed by a third transcriber, Antonia Tovar. Prior to calculation of the inter-transcriber reliability, the main author met with the third transcriber and established a consensus of disagreements after re-listening to the transcriptions together. We used the reliability calculation procedure outlined in Banney et al. (2015). We took the total number of point-by-point agreements and divided it by the sum of the total agreements and disagreements. The inter-transcriber reliability came to 90.3%. We suggest this reliability to be quite robust due to the fine-grained nature of the transcription, which e.g. included partial word repetitions.

Annotation reliably was carried out in a similar method. The annotations were first carried out by the first transcriber, Miriam Garcia Subirats and then edited by the main author. 10.5% percent of the transcriptions were annotated by a third annotator, Joana Rosselló Ximenes. The third annotator and the main author met and formed a consensus before calculating the inter-annotator reliability. The inter-annotator reliability was 90.1%.

Analysis plan

Statistical analyses were carried out using SPSS version 24. Where assumptions of normal distribution were met, independent t-tests will be run to assess group differences. Where there were significant outliers or data were skewed, as assessed by the Shapiro-Wilk test, we carried out non-parametrical Mann-Whitney U-tests and compared medians. Following Banney et al. (2015) and Rumpf (2012), a more stringent threshold for significance was adopted ($p<.01$, two-tailed) in order

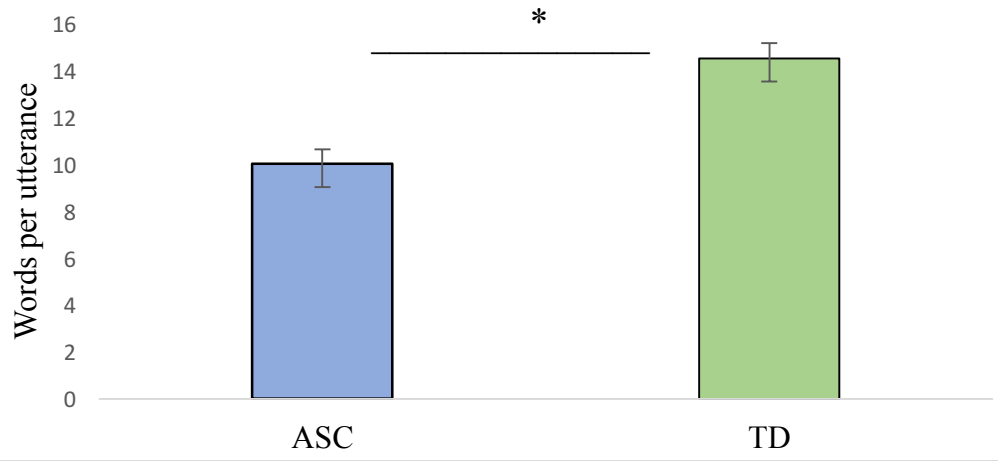
to limit the possibility of the null hypothesis being incorrectly rejected in light of the various comparisons being made.

2.3 Results

2.3.1 General productive measures

The ASC group generated utterances that were significantly shorter in length than their typically developing peers (ASC: mean= 10.04, range =4.76-14; TD: mean=14.53, range =10.87-21.3; t(34) = -5.067, p = .0001) (see Fig.3). Both groups, however, produced narratives with similar levels of lexical diversity as measured by the type-token ratio (ASC: mean= .51, range= .42-.71; TD: mean = .48, range= .37-.59; t(34)=1.294; p=.172).

Figure 3. Rate of words per utterance

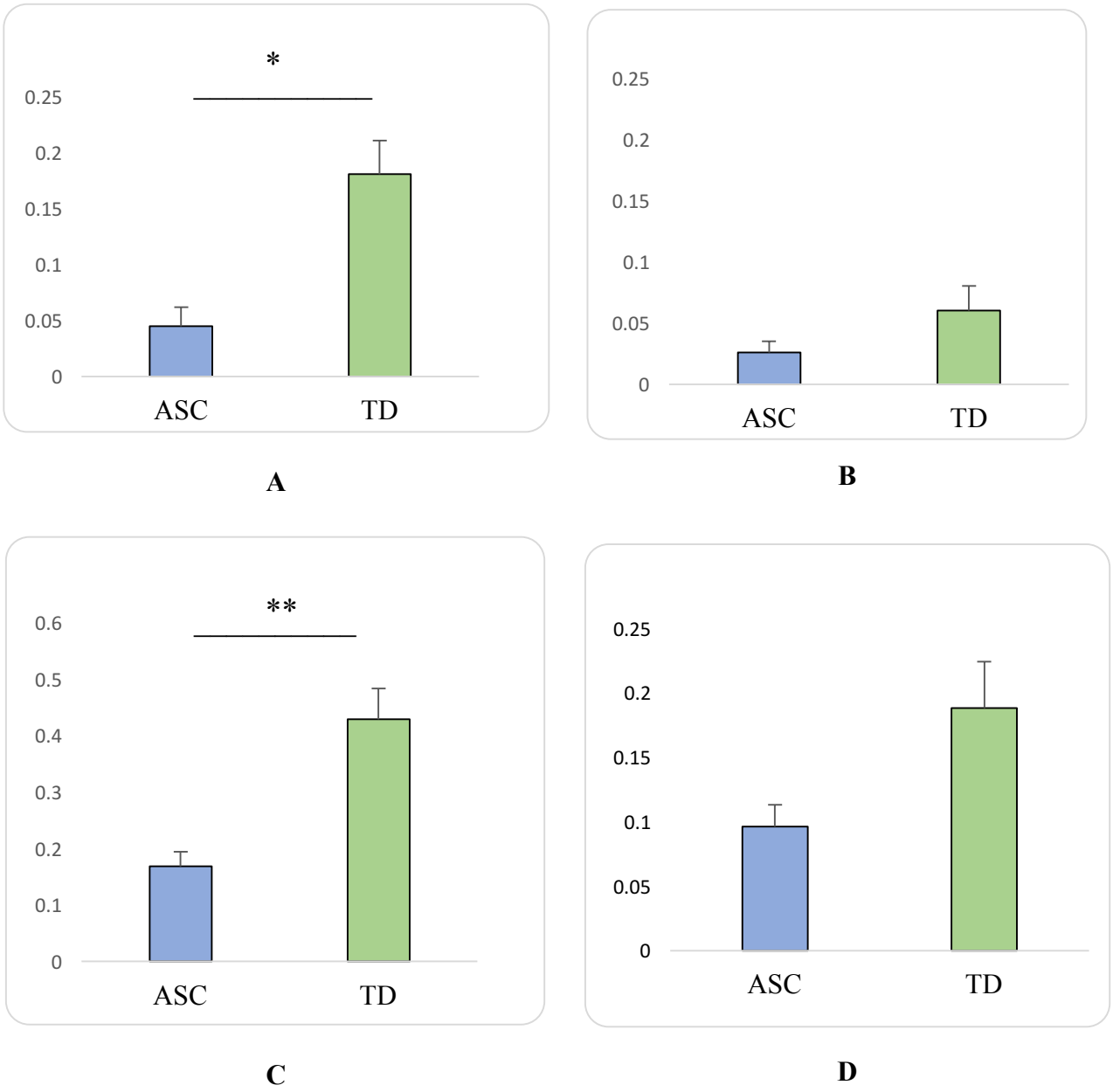


* indicates $p \geq .0001$

2.3.2 Grammatical profiling

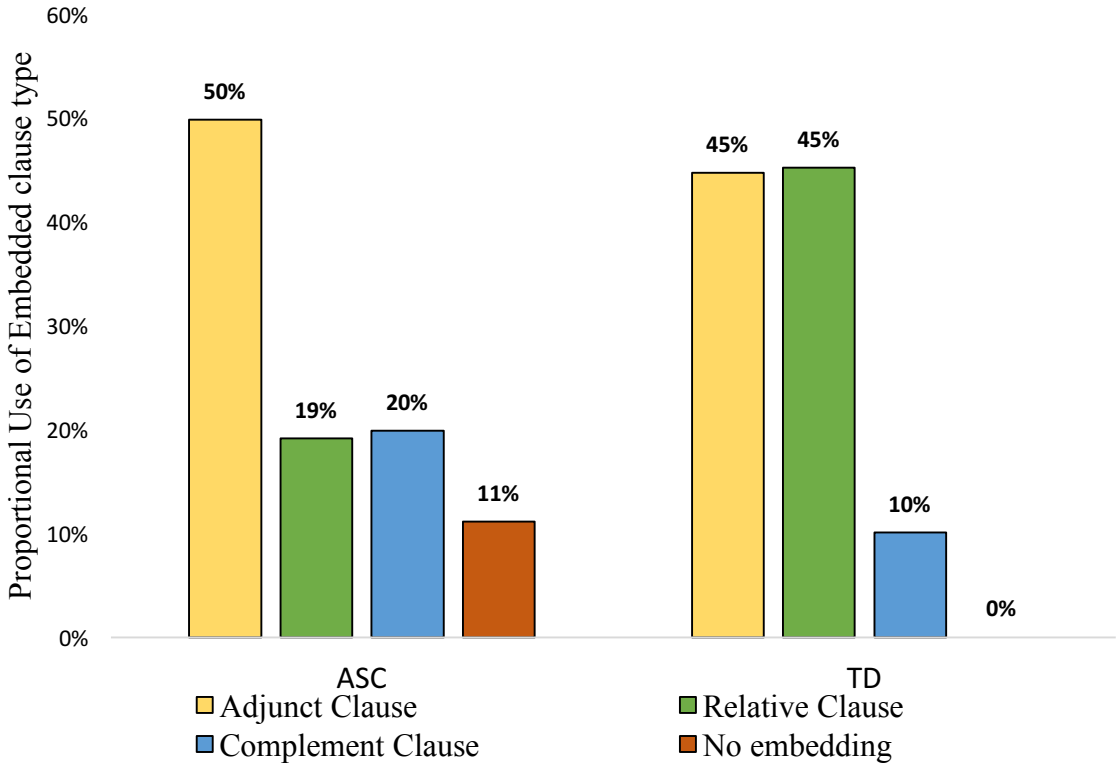
In addition to producing fewer words per utterance, the narratives produced by the ASC participants contained significantly fewer embedded clauses; 36.6 percent of utterances in the TD group had an embedded clause in comparison to just 15.2 percent of utterances in the ASC group. In particular, the ASC group produced significantly fewer relative clauses yet similar levels of adjunct clauses and complement clauses (Figure 4), complement clauses being the least frequent embedded clause type across both groups. Setting aside overall frequency, the ratio of relative clauses – i.e. the proportional use of relative clauses in relation to the total instances of embedding – was lower among the ASC group than that of TD controls, however this difference did not reach our threshold for significance (p=.012). See Figure 4 and 5 for visualizations of embedded clause data and Table 7 for medians, U, z-, and p-values.

Figure 4: Rate of embedded clauses per utterance



A= rate of relative clauses per utterance; **B**= rate of complement clauses per utterance; **C**= rate of total embedding per utterance; **D**= rate of adjunct clauses per utterance. * indicates $p \geq .01$; **indicates $p \geq .001$.

Figure 5: Proportional use of embedded clause types



In contrast to the clausal domain, there were no significant differences between groups in the frequency of NP types or VP types, both in relation to total frequency (rate of each NP of VP type in relation to total number of words) or proportional use (ratio of each NP or VP type in relation to total number of NPs or VPs, respectively). Furthermore, there were no significant differences in frequency of clitic production, either in total or according to the 'successful' or 'unsuccessful' sub-classification. Results from the comparisons of grammatical profiling measures

are presented in Table 5 (parametric analyses) and in Table 6 (non-parametric analyses). Results from the comparison of grammatical device ratios are presented in Table 7.

Table 5. Frequencies of NP types

Rate of NP types	ASC mean (range)	TD Mean (range)	t	*p*
Indefinite	.062 (.036-.122)	.057 (.017-.098)	.671	.506
Definite-anaphoric	.136 (.088-.179)	.142 (.091-.102)	-.644	.524
Definite non-anaphoric	.049 (.005-.098)	.05 (.022-.08)	-.061	.952

Table 6. Frequencies of clitics, verbal and clausal construction types

	ASC Median (SD)	TD Median (SD)	*U*	*z*	*p*
Clitics					
Clitics- successful	.009 (.008)	.015(.01)	197.5	1.126	.265
Clitics- unsuccessful	0 (.004))	0 (.004)	145	-.828	.606
Total clitics	.011(.01)	.017(.01)	196	1.077	.293
Verbal Domain					
Stative verbs	.038 (.017)	.039 (.018)	168	.19	.864
Episodic verbs	.108 (.023)	.124 (.026)	182.5	.649	.521
Clausal Domain					
Relative Clauses	0 (.072)	.191 (.128)	266.5	3.387	.001*
Complement clauses	0 (.036)	0 (.085)	191	1.025	.372
Adjunct Clauses	.118 (.073)	.177 (.154)	224.5	1.990	.047
Total Embedding	.152 (.111)	.366 (.227)	282.5	3.814	.0001**

Mann-Whitney U-tests were applied to compare differences in distributions.

Table 7. Ratio of grammatical devices

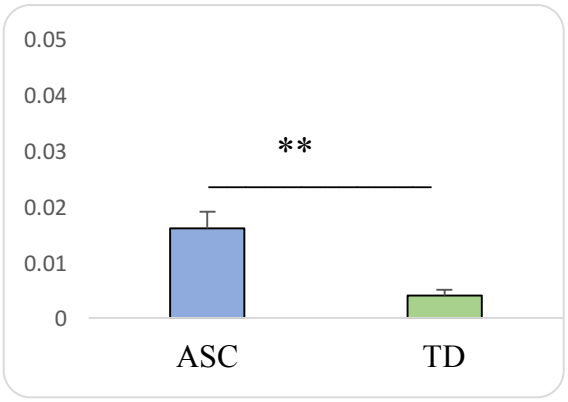
	ASC Median (SD)	TD Median (SD)	U	z	*p*
Ratio of NP types					
Indefinite	.231 (.081)	.218 (.097)	146.5	-.490	.628
Definite-anaphoric	.545 (.063)	.568 (.069)	181.5	.617	.542
Definite non-anaphoric	.19 (.096)	.192 (.066)	174	.380	.719
Ratio of verbal types					
Stative verb ratio	.245 (.034)	.259 (.114)	157	-.158	.888
Episodic verb ratio	.755 (.084)	.74 (.114)	167	.158	.888
Ratio of embedded clause types					
Relative clause ratio	0 (.24)	.408 (.311)	241	2.565	.012
Complement clause ratio	0 (.331)	0 (.149)	154	-2.83	.815
Adjunct Clause ratio	.633 (.367)	.367 (.319)	145	-.543	.606

2.3.3 Errors

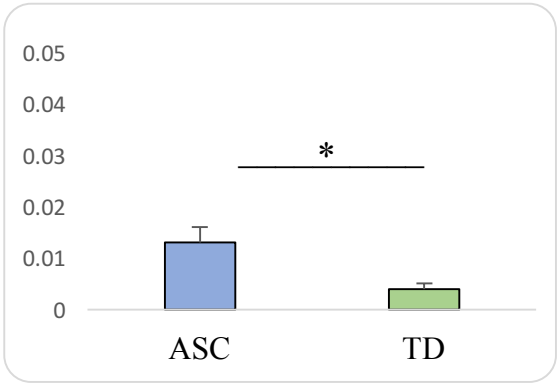
Composite error scores

ASC participants produced lexical and grammatical errors at a significantly higher rate than TD controls (see Table 8). We will first present the composite scores of error types, followed by specific lexical and grammatical error types (recall the flowchart presented in Figure 2). The ASC group produced both significantly more lexical errors and grammatical errors than TD controls. Within grammatical errors, the ASC group produced significantly more total nominal errors as well as, specifically, nominal reference errors, while there was no significant difference between groups regarding verb errors. Regarding lexical errors, the ASC participants produced significantly more lexical errors than TD participants, with a rate of approximately 1.2 percent of words being used incorrectly in the ASC case in comparison to .5% among TDs (see Figure 6).

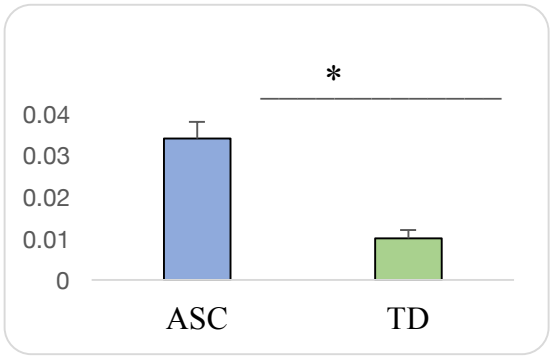
Figure 6. Frequency of error composite scores



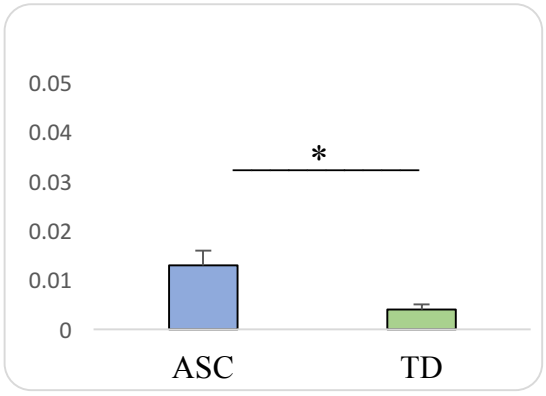
A



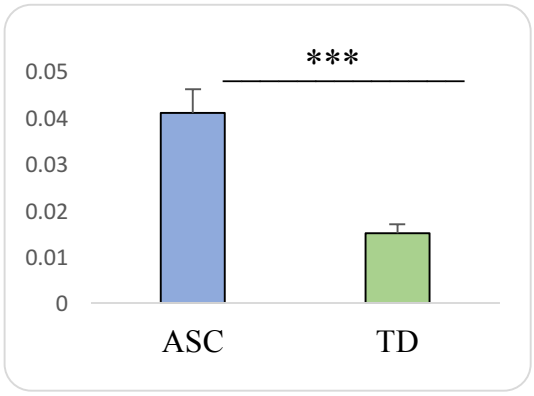
B



C



D



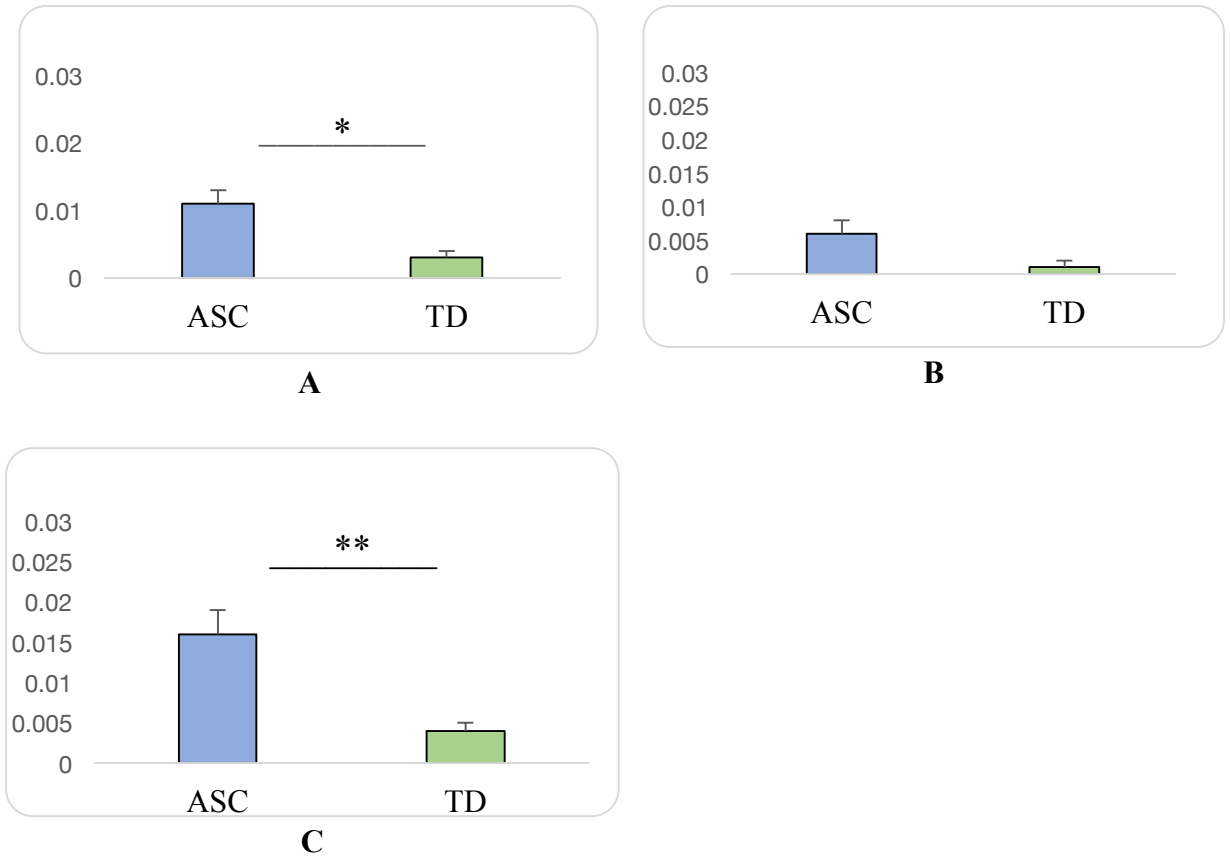
E

A= wrong lexical item (WLI); **B**=total referential errors; **C**=total nominal errors (grammatical) included; **D**= total verb errors; **E**=total errors; *indicates $p \geq .01$; ** indicates $p \geq .001$; *** indicates $p \geq .0001$

Lexical errors

When looking at specific types of lexical errors, the ASC participants produced significantly more non-relational lexical errors in comparison to TDs. While the ASC group also quantitatively produced relational lexical errors at a higher frequency this difference was not significant (U = 59; z =-3.86; p =.055). The frequency for relational lexical item errors was low across both groups, the median rate being .3% of words for the ASC group and 0% for the TD groups (see Figure 7).

Figure 7. Frequency of lexical errors

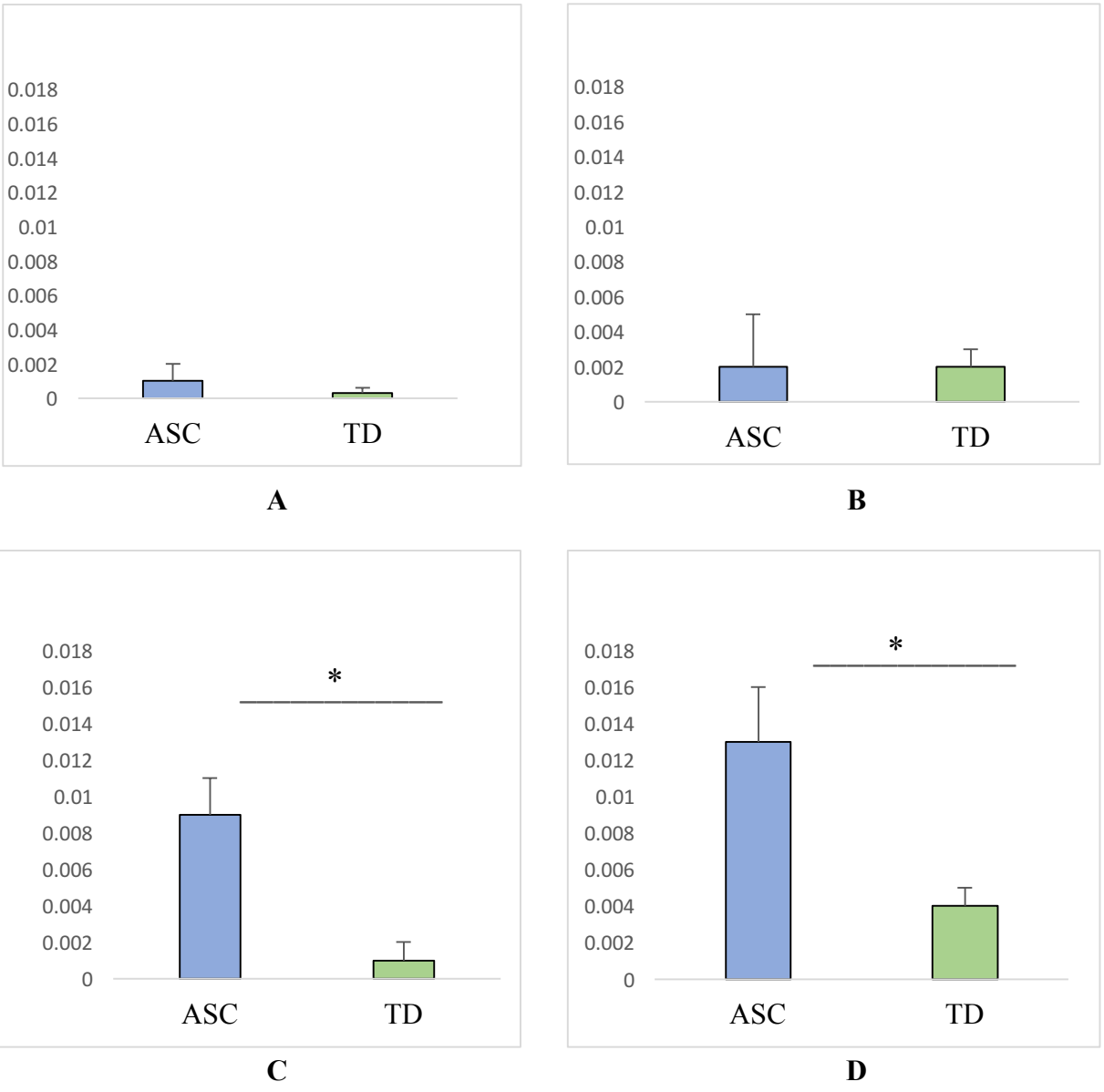


A= non relational wrong lexical items; **B**= relational wrong lexical items; **C**= total wrong lexical items; *indicates $p \geq .01$; ** indicates $p \geq .001$

Grammatical errors

Regarding grammatical errors, significant differences were only identified in the variable *failed specific reference*, which captured referential errors in which it is unclear who or what the intended referent is. As the data presented non-normal distribution, medians were compared across groups. The ASC group presented a median error failed specific reference rate of .7% of total words while the median rate for TD was zero failed specific reference errors. There were no significant differences between frequencies of other nominal reference errors such as definite indefinite mismatch or vague reference (see Figure 8), nor among non-referential nominal errors such as noun agreement or verb agreement errors (see Table 8).

Figure 8. Rate of referential NP errors (normalized by number of words)



A= Definite-indefinite mismatch; **B**= Vague reference, **C**=Failed specific reference; **D** = Total referential errors; * indicates $p \geq .01$

Table 8. Error scores normalized by number of words. Medians and ranges are presented.

	ASC (*n*=18) Median (range)	TD (*n*=18) Median (range)	U	*z*	*p*
Total errors	.038 (.013-.089)	.015 (0-.032)	40	-3.860	.0001****
Total wrong lexical item	.012 (0 - .04)	.005 (0-.012)	59	-3.284	.001***
WLI relational	.003(0-.018)	0 (0-.007)	101	-2.244	.055
WLI non-relational	.011 (0-.031)	.002 (0-.007)	73	-2.884	.004**
Total grammatical errors	.022 (0-.053)	.009 (0-.032)	70	-2.914	.003**
Total verbal domain errors	.005 (0-.027)	.002 (0-.019)	142.5	-.646	.542
Lacking tense in the main clause	0 (0-.01)	0 (0-.006)	161.5	-.040	.988
Wrong or inconsistent tense or aspect	0 (0-.026)	0 (0-.017)	167	.184	.888
Error in the verb root or conjugation	0 (0-.013)	0 (.007)	179	.828	.606
Truncation	0 (0-.008)	0 (0-0)	126	-2.087	.265
Total NP errors	.017 (0- .047)	.005 (0- .02)	58	-3.298	.001***
Noun agreement	0 (0-.017)	0 (0-.007)	117	-1.745	.161
Verb agreement	0 (0-.009)	0 (0-.007)	162	0	1
Total referential errors	.011 (0-.037)	.004 (0-.018)	75	-2.792	.005**
Vague Reference	0 (0-.018)	0 (0-.012)	180	.698	.584
Failed Specific Reference	.007 (0-.037)	0 (0-.007)	71	-3.117	.003**
Definite/Indefinite mismatch	0 (0-.017)	0 (0-.006)	144.5	-1.015	.584

2.3.4 Mental state and emotion content

There were no significant differences between groups on frequency of mental state or emotional content (see Table 9). We note that errors in mental state content, such as attributing an emotion that is incongruent with the story's depiction only occurred, though quite infrequently, among the ASC participants.

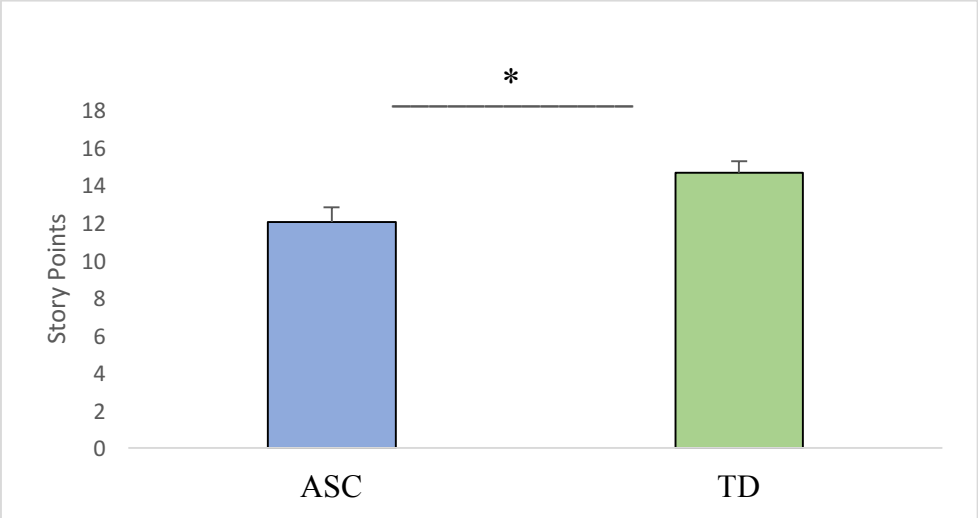
Table 9. Mental state and emotion content

	ASC Median (range)	TD Median (range)	U	z	*p*
Total mental state /emotion content	.013 (0-.054)	.012 (0-.028)	140.5	-.682	.406
Mental state /emotion content-Good	.012 (0-.054)	.012 (0-.028)	152	-.301	.767
Mental state /emotion content-Bad	0 (0-.002)	0 (0-0)	135	-1.782	.406

2.3.5 Macrostructure

We compared the Story points between groups, measuring how complete the narratives were in terms of including relevant events in the narrative. The ASC group produced significantly fewer story points than the TD group, (ASC: median = 13, range = 3 – 16, TD: median= 15, range = 4-18, U= 248.5; z=2.755, *p*=.005). See Figure 9.

Figure 9. Story points



**indicates p ≥ .001*

2.3.6 Fluency measures

Repair

We compared the frequency of repair the participants engaged in, either in the form of reformulations or self-corrections. Comparing medians, the TD participants produced quantitatively more reformulations or self-corrections that were successful i.e. *self-correction-good,* which was counter to our predictions; however, these differences in frequency were not statistically significant (see Table 10a). The ASC participants produced more self-corrections-bad, in which the self-correction either did not address the underlying error well or introduced a new error; however, as with the other repair variables, this difference was not statistically significant.

Repetitions

We compared the frequency of repetitions across groups. The ASC group produced slightly more phrase and total repetitions while the TD group produced slightly more frequent 'other' type repetitions, which was comprised of both partial word repetitions and repetitions that were greater than a phrase in length. However, none of these group differences reached the threshold of statistical significance (see Tables 10a &10b).

Table 10a. Repair and Fluency measures: non-parametric analysis

	ASC Median (range)	TD Median (range)	U	z	*p*
Reformulations	.004 (0-.033)	.008 (0-.021)	189	.875	.406
Total Self-corrections	.006 (0-.054)	.009 (0-.019)	187.5	.815	.424
Self-correction- Good	.006 (0-.036)	.008 (0-.019)	189.5	.379	.389
Self-correction-Bad	0 (0-.018)	0 (0-.006)	142	-.975	.542
Total Repair	.013 (0-.055)	.018 (0-.033)	180.5	.586	.563
Repetitions other	.009 (0-.044)	.004 (0-.029)	114.5	-1.531	.134
Repetitions phrases	.005 (0-.018)	0 (0-.007)	104	-1.921	.068
Total Repetitions	.029(0-0-.102)	.021 (.006-.065)	134	-.886	.389

Table 10b. Repair and Fluency measures: parametric analysis

	ASC Median (range)	TD Median (range)	t	*p*
Repetitions: words	.016 (0-.044)	.017 (0-.041)	-1.81	.857
Total Fluency	.05 (0-.124)	.043 (.02-.078)	.799	.432

2.3.7 Correlation Analysis

Correlation analyses were carried out between verbal IQ and key variables that were significant points of difference between groups, specifically, failed specific reference, wrong lexical items, embedded clause frequency, and story points. However, no linear relationships were identified. Furthermore, for the ASC group cognitive measures were available as part of the WISC, including working memory and processing speed. None of the cognitive variables related to these key linguistic variables.

2.3.8 Results summary

We identified several differences across groups that were robust in nature:

(1) Significantly shorter utterances in the ASC group in relation to the TD group.

(2) Significantly fewer story elements included in the narratives produced by ASC participants in relation to the TD group.

(3) A significantly greater rate of referential errors among the ASC participants in comparison to TDs, particularly that of failed specific reference.

(4) A significantly greater rate of lexical item errors in the ASC group in relation to the TD group, particularly in non-relational wrong lexical item choice.

(5) A significant reduction in the rate of embedded clauses in the ASC group in relation to the TD group, particularly in relative clauses but not adjunct clauses or complement clauses.

In contrast, the ASC participants did not differ significantly with regard to either the frequency of NP and VP types. Nor did they differ in terms of the frequency of verbal errors, repair or repetitions, or rate of use of internal state language across groups. No significant correlations were identified.

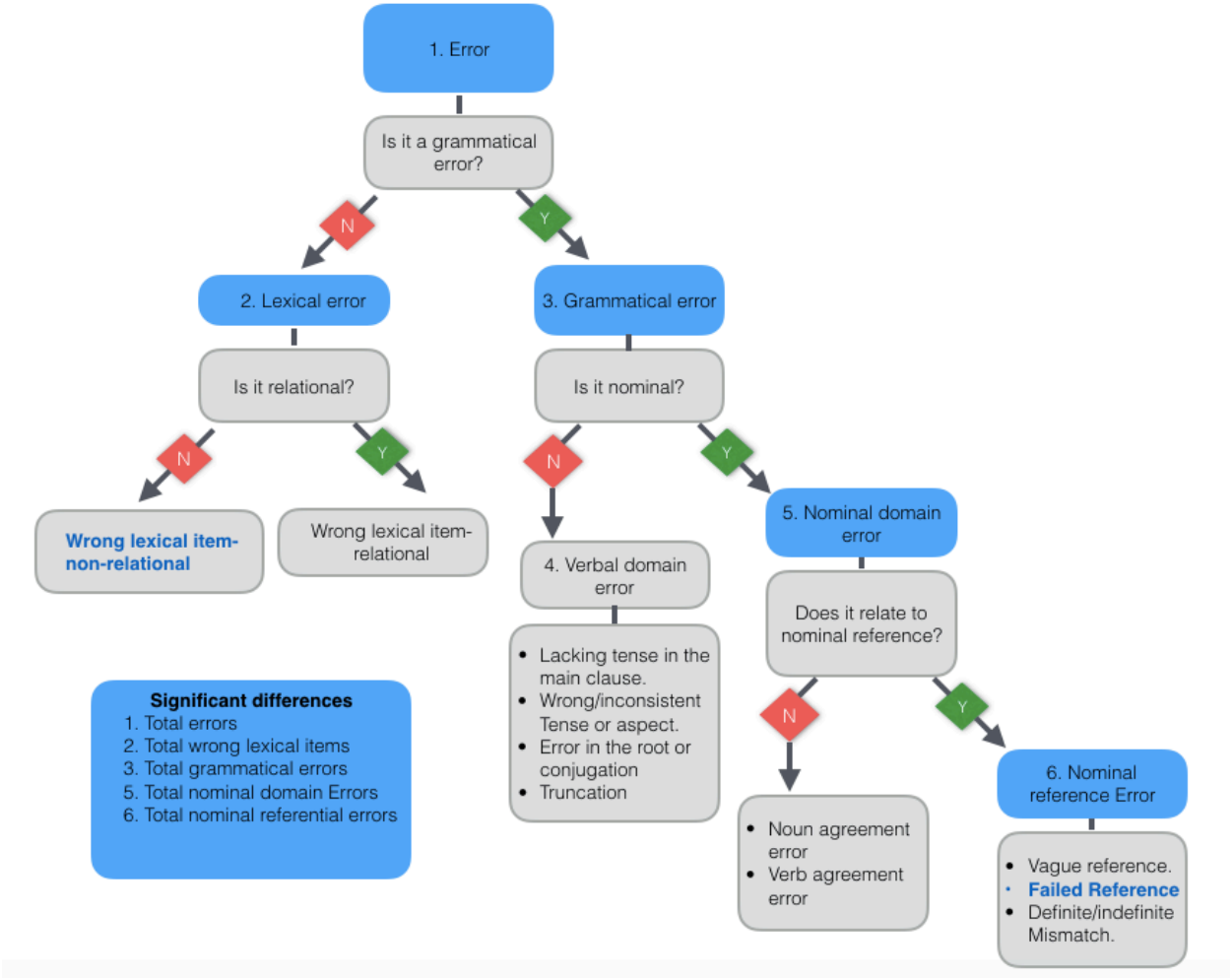
2.4 Chapter 2 Discussion

The aim of this study was to explore productive language capacities of children with ASC, within the domain of narrative production, in comparison to TD children, matched on verbal IQ. Our particular aim was to explore the use of referential expressions against a backdrop of broader linguistic profiling. In this discussion, we will first begin an overview of the differences identified between participant groups, followed by an in-depth descriptive discussion of the referential errors. Secondly, we will discuss group differences with regard to embedded clauses, lexical errors and other significant results. Finally, we reflect on how our results square with our initial predictions as well as the ways in which our participant groups performed similarly.

The present study identified several differences in the narratives that were produced by our ASC and TD participant groups. Our ASC participants tended to narrate the picture book using shorter utterances and with a reduced variety of word choice (TTR). The utterances produced by the ASC group were also structurally less complex with a lower overall rate of embedded clauses and, in particular, fewer relative clauses. Throughout the narratives, the ASC group also tended to omit central story elements (story points) more often, resulting in narratives that were less qualitatively complete in comparison to their TD peers, in line with previous findings by Norbury and Bishop (2003). This may relate to either (or both) a greater difficulty among the ASC group to make sense of the story and attend to all of the story elements or a performative difficulty such as knowing to be an informative storyteller or even difficulty managing the language needed to convey these points. The groups also differed significantly in terms of the frequency of errors produced. When assessing the types of errors that distinguished groups, we see that the ASC group produced greater total lexical errors, particularly 'non-relational' lexical errors (e.g. saying *blanket* rather than *sheet* or *turtle* rather than *frog*) as well as total grammatical errors, (non lexical)

nominal errors, and referential errors, and specifically that of specific reference relative to TD controls. If we return to our error flow chart, Figure 2 (page 27), we see significant differences in all the nodes of the grammatical error branch involved with the failed specific reference variable. See Figure 2 (here reproduced as Figure 10) below for a highlighted version of how significant differences appear in the context of our schema.

Figure 10. Error coding flowchart (Blue indicates significant differences between groups).



Our guiding hypothesis was that anomalies in referential language uses, particularly in relation to failed specific reference, would be more prominent in the ASC group than in the TD controls. We found support for this prediction as the participants with ASC presented errors of this type at a higher rate than the control group. We will now turn to exploring these errors in greater depth.

When exploring the presentation of these errors, different patterns of failed specific reference emerge. While instances of reference failure are readily identifiable, the *why* of the referential error may be underdetermined. For example, if a pronominal NP does not agree in gender or number to the presumed antecedent there is a referential failure. This may be due to problems in agreement per se or in the individual's ability to track co-referentiality and adequately introduce new referents. In our coding procedure we coded errors of this type as instances of failed specific reference given that the anaphoric link is unsubstantiated. Recall, however, that instances of agreement which do not include referential breakdown were coded as noun agreement error, such as when an adjective + noun or article +noun did not agree in gender or number. Instances of the difficulty introducing a new character, resulting in referential failure were found in both groups, see examples 1-2 by ASC participants and example 3-4 by TD participants. Brief descriptions of the referential failure are provided following each example.

Sample failed specific reference errors

(1) ASC

…se fueron saltando al lago y se tiraron de cabeza y están allí, como pensando, **está enfadado**.

Went$_{3p\text{-}pl}$ jumping back to the lake and jumped$_{3p\text{-}pl}$ head-first and are$_{3p\text{-}pl}$ there, like thinking, **is$_{3p\text{-}sing.}$ angry.**

'They went jumping back to the lake and they jumped head-first and they are there, like thinking, (*he) is angry.'

Comment on error: here the failed specific reference error emerges from a number mismatch between the singular, anaphoric inflectional NP (number marked through verb inflection, 'está enfadado'/'[he] is angry') which is incongruent with the plural antecedent ('fueron'/ '[they] went). In the book, there is indeed a single frog which appears angry and so the listener may infer who the child is referring to. Yet this example suggests that the child is not managing well the anaphoric links she or he is creating.

(2) ASC

Aquí **vuela** tranquilo y aquí se choca con un perro aquí se xxx[7] con un perro y **se van** y se **los** come.

Here calmly **flys$_{3p\text{-}sing}$** and here runs into a dog here xxx with a dog and **go$_{3p\text{-}pl}$** and **clitic:3p-pl.** eats$_{3p\text{-}sing.}$

[7] 'xxx' indicates an unintelligible word or strings of words in the transcription.

'Here (*he) calmly flies and runs into a dog here xxx with a dog and they go and he eats them.'

Comment on error: Similar to (1), the error here emerges due to an incongruence with a (presumed) anaphoric referent in which the anaphoric NP (se van, '(they) go') is marked as plural while the antecedent is singular. In the book, the illustration shows a single frog flying when a dog begins to chase it. Then more frogs show up and scare off the dog. As such again we see that the child appears to be failing to navigate the co-referential links being established by her or his narration and failing to introduce a new referent, or in this case reduce the scope from the set to an individual of the set.

(3) TD

Ellas estaban viendo la tele […] después **se fue** y apareció un perro.

They were watching TV […] then **left**$_{\textbf{3p-singluar}}$ and a dog appeared.

'*They* were watching TV […] then [*he*] left and a dog appeared.'

Comment on error: This error is similar to the errors 1-2 listed above by the participants with ASC in that there is a failed anaphoric link between 'they were watching' and '($_{pro}$ he) left'. In order to felicitously change from reference of the set of frogs to a particular, the child would have needed to use a demonstrative like 'éste, (*this one*)' or say 'uno se fue, (*one left*)', depending on whether the child is referring exophorically with the use of a demonstrative or endophorically with an indefinite NP.

(4) TD

Después el gato **le** ve y una rana con el mando.

After the cat clitic:3p-sing. sees and a frog with the remote.

'After the cat sees him and a frog with the remote.'

Comment on error: Here there is a lack of number agreement between the single clitic pronoun *le* and the plural antecedent, which from the context of the story we interpret to be the frogs. In this case, the intended referent of the clitic truly is ambiguous as there is no salient individual frog in the scene. At this moment all of the frogs enter a room, where there is a sleeping elderly woman in front of her TV set. As the woman is asleep she is totally unaware of the frogs watching TV and changing the channels. The woman's cat, on the other hand, watches the scene apprehensively in the corner.

These examples illustrate that we may identify similar errors of failed specific reference between groups, though these errors were more frequent in the ASC group. In particular, these errors arise when the child appears to be referring to a specific individual though this specificity is not adequately modulated in the referential forms used. It is relevant to note that the TD examples identified for failed specific reference above were produced by one of our younger participants, who was also the participant with the lowest verbal IQ. Indeed, example (4) came from our only TD participant who was an outlier in terms of verbal IQ (VIQ =55). There were no correlations between age or verbal IQ with failed specific reference in the TD group, however, this may be due to the fact that errors of this type were generally so infrequent.

As noted above, there is a second pattern of referential errors that could be identified regarding the ASC participants. In these cases, the participant's utterance in its literal interpretation

doesn't match a viable reality in consideration of the visual support provided. Again it is important to recognize that we remain naïve to the participant's referential intent or linguistic representation of the utterance in question and that our sense-making stems from interpretation and thus must be carried out with caution. Keeping that in mind, let us consider the following examples (5) and (6) from two ASC participants, failed specific reference errors marked in bold:

(5) ASC:

Entraban$_i$ por la ventana y **les$_i$ quitaban** los pañuelos.

They$_i$ entered through the window and (3rdPplural-CL$_i$) they$_a$ took the handkerchiefs.

'They$_i$ entered through the window and **they(-others) took** the handkerchiefs (from them$_i$)'

Comment on error: At this scene in the book, the frogs first are flying with capes tied around their necks, made from the sheets that they had flown into previously. The same page presents them flying into a house without their caps. In this error, the child is saying that they entered the house through the window and that *they* – although it is unclear who given the visual support – took the capes off them. This reflects the default interpretation provided the structure of the sentence. However, provided the visual support the more viable interpretation would be that the child intended 'se quitaron los pañuelos', a clitic –V-NP possessive construction where the clitic refers to the possessor of the entity denoted by the NP, a possessor that in this particular example should be co-referential with the subject of the coordinate clause.

(6) ASC

Cogieron el mando sin que se **den** cuenta y estaban viendo la tele.

They$_i$ took the remote control without **noticing$_{i-3p-pl}$** and were watching TV.

'They took the remote control without (*her) noticing and were watching TV.'

Comment on this error: In this example, there are three issues with the use of 'den cuenta'. First the new character, the older woman, is being introduced with an inflected pronoun. Second, there is a number mismatch as the verb *to notice* should be singular rather than plural if it is to refer to the woman. Finally, we note that there is also a tense/aspect error in Spanish as the verb 'den cuenta' must be in the past.

In examples 5 and 6, we see referential errors emerge that may be a result of language processing difficulties. In (5) we see the misuse of the dative clitic *les* in place of the reflexive pronoun *se*, which gives us the interpretation of a novel entity in the scene. This error suggests impoverished understanding of grammatical resources, such as the licensing of pronouns or of reflexive verbs. In (6) we see a number agreement error, specifically an unexpected plural given that the character who is 'unaware' is the single, older woman. What may be happening is an agreement attraction of the proximal NP which is plural. These kind of proximal concord errors are documented in early language acquisition as well as in written language errors in TD adults. Writing slows down our speed of language processing and makes individuals more susceptible to these proximal agreement errors (Fayol, Largy, & Lemaire, 1994). This phenomenon is also shown to be increased in written tasks when cognitive load is increased such as adding concurrent tapping tasks to the written exercises (Fayol, Largy, & Lemaire, 1994). If language processing is costlier for our ASC participants, this may lead to difficulty maintaining a linguistic representation of the discourse or relate to grammatical errors, or a combination of both. Referential functions may be highly susceptible to breakdown due to their relation to complex syntactic relationships. If the underlying structure of these relationships is less stable, it would follow that referential errors emerge. This notion of weaker language processing also aligns with our finding of general

reduction in grammatical complexity in the ASC group. It is important however to not over interpret these suggested explanations as they are tentative and speculative and require further qualitative analyses.

One robust finding that we did not anticipate was the high rate of lexical errors, particularly what we called 'non-relational' lexical errors which were virtually non-existent in the TD group. This, at first glance, is particularly surprising provided the matching of participant groups on receptive vocabulary level. The average verbal IQ as measured by receptive vocabulary of our ASC participants was 99.3 and so well within typical performance for their age[8]. The finding also contrasts to widely stated claims in the literature on children with autism without ID that vocabulary was relatively preserved, compared to grammar (e.g. Eigsti et al., 2007). We noted that the wrong lexical choices (WLI) were generally related to the target words in terms of appearance or function, such as calling the sheets that were hanging on the clothesline either a 'handkerchief' or a 'blanket' for example, yet not quite matching the visual stimuli. As we know that these children are highly comparable to the TD on receptive vocabulary, the lexical choice anomalies that we observed appear to relate to the constraints of lexical retrieval within the context of the task that is linguistically demanding.

Considering these results together suggests that greater difficulties in language processing may give rise to the pattern of errors that we identified. While cognitive capacities such as theory of mind are often drawn upon to provide a potential explanation for referential errors (e.g. Colle et al., 2008; Suh et al., 2014), it is unclear how mentalizing deficits would account for the kinds of group differences above, which include lexical errors, as identifying accurately a picture in the

[8] A verbal IQ score of 100 indicates being exactly in the 50% percentile for one's chronological age.

story such as ‘a sheet’ or ‘a frog’ would not depend on the mental state of the listener of the story. Indeed, the overall error pattern suggests deficits within the child’s own language processing rather than a difficulty in adjusting information to the knowledge of the listener. We also note that as a group our participants did not differ with regard to attribution of character’s mental states or emotions. This is not to preclude that our participants may present mentalizing deficits in other test contexts or daily life. Rather we intent to highlight that several of the kinds of errors, both referential and lexical, are independent of such considerations as taking the other person’s perspective into account. Cognitive capacities such as phonological or working memory may be relevant to explore further as this has been argued to relate to the ability to maintain discourse representation and maintain anaphoric links in ASC (Schaeffer et al., 2018). However, a deficit in working memory alone would also not incorporate the lexical errors that we found.

Reflecting on the ways in which our predictions were not met, we had anticipated greater disfluencies and repair in the ASC group. In fact, although there were no significant differences, we found that the TD participants actually engaged in more self-corrections than the ASC participants in absolute terms. The greater rate of errors in the ASC group, on the other hand, could indicate that speech monitoring was less present in our ASC group. We also had explored the verbal domain, in particular, whether scenes were described episodically or if there was a tendency towards using stative verbs more frequently in the ASC group. This was a relatively exploratory variable and we indeed found that the two groups performed similarly. We also did not identify any group differences with regard to the verbal domain errors, indeed none of the variable approximated significance.

One way in which our predictions related to referential language were not met was related to the rate of use of anaphoric reference, whether in full NP or in clitic forms. While as a group

the ASC group used fewer clitics and more incorrect clitics, group differences were minimal and not significant. Reflecting on this non-significant result, we identified a potential limitation of our study. We are not able to distinguish whether definite NPs (either full NPs or clitics) are being represented endophorically (in the sense of 'true' anaphoric reference) or exophorically, (referring to an entity in the shared visual domain). Depending on the degree of salience, definite full NP or clitics may indeed be exophoric. While we had purposefully made this distinction in our classification of definite-non anaphoric reference (such as when a novel referent is introduced with a definite article), it is still possible that a child referring to a maintained referent (which we as the listener would interpret as anaphoric) need not necessarily be represented as such by the child. Therefore, what we had coded as definite anaphora may be under-determined with regard to the child's representation and therefore overstated. Indeed, provided the general dampening of linguistic skills in the ASC group, it would be parsimonious to expect that they may be relying on greater exophora rather than building and maintaining an internal representation with endophora. Given this confound, results from narratives with visual support may need to be contrasted with language production elicited though other means such as story recall without visual support or free speech, in which the endophoric/exophoric distinction would be more easily determined and allow us to better interpret this null result.

In sum, we had anticipated that the participants with ASC would present greater referential anomalies than their TD peers. Greater scrutiny of the nature of referential errors in the ASC group as well as their broader pattern of lexical selection errors and reduction of grammatical complexity, suggests that the underlying cause of referential errors may lie in a weaker language processing system. This conclusion however is not definitive and based on qualitative assessment of error patterns so this interpretation must be taken with caution. Nevertheless, the present results suggest

that by better exploring language processing deficits, we may reach a better understanding of how reference breaks down in ASC. Previous studies have identified that use referential language is a point of weakness in ASC, yet have not ventured to relate this weakness to broader language functioning. Our study suggests that language capacities and referential competence may be more deeply intertwined in ASC.

Chapter 3. Comprehension of referential expressions across the autism spectrum*: two studies on the comprehension of referential NPs types across the autism spectrum (ASC+/- ID)*

3.1. Introduction

Language is a primary component of human communication and social interaction. As such, it is unsurprising that many symptoms and behaviors related to communicative deficits in Autism Spectrum Conditions (ASC) are identified through atypical language use. For example, difficulty maintaining conversation and building on the comments of interlocutors, reliance on stereotyped language during interactions, or even misuse of personal pronouns, are kinds of verbal behaviors that are assessed in standardized diagnostic evaluations such as the Autism Diagnostic Observation Scale (ADOS). The role, if any, that atypical language processing may have on communicative deficits, however, has been less explored. Across the autism spectrum, correctly using language in context is a common difficulty. In particular, data from narrative production studies suggest that difficulties managing the referential use of language is a prominent characteristic of language production in ASC (e.g. Baixauli et al., 2016). Similarly, we have seen in Chapter 2 that referential errors, along with lexical errors and reduced sentence complexity, were factors that distinguished children with ASC–ID from their verbal-IQ matched peers. Less is known however about the comprehension of the specific grammatical constructions involved in linguistic reference. While content words like *cat* denote the general concept CAT, referential noun phrases (NPs) relate content words to the communicative context in ways that specifically depend on determiner choice, by which deictic, definite, and indefinite forms of reference can be distinguished. A growing body of literature has explored aspects of referential language use and comprehension in ASC, yet not

systematically. A better understanding of potential weaknesses in different types of referential NPs may help further our understanding of the linguistic basis of communicative symptoms in ASC.

3.1.1 Use of referential NPs in ASC–ID

In reciprocal conversation, the discourse is construed collaboratively by conversational partners. Part of the cohesive dynamics of a conversation are the referential ties which relate novel information with shared, 'common-ground' information. Early work on language in ASC by Fine, Bartolucci, Szatmari and Ginsberg (1994) explored how difficulties using referential devices may contribute to the breakdown in conversational reciprocity in autism and Asperger's syndrome (AS). Through an analysis of spontaneous speech gathered through a semi-structured interview, they found that children with ASC–ID used fewer anaphoric references than the children in an AS group or TD controls. The AS group tended to make more unclear references, particularly of the sort in which the antecedent is lacking (e.g. saying 'another river' without first establishing a given river). The ASC–ID group (with language delay) also showed less *endophoric* reference and more *exophoric* reference than the AS group, meaning that they referred more to things present in the physical context around them rather than referring to entities anchored in the discourse. Using storytelling tasks to elicit narratives, various studies have compared differences in grammatical and cohesive devices used in comparison to TD peers. These studies show both an overproduction of ambiguous pronouns (e.g. Banney et al., 2015) and a general reduction of pronouns (e.g. Rumpf et al., 2012) among children with ASC–ID or AS in comparison to TD controls.

Two studies have also explored the cognitive mechanisms which may contribute to deficits in referential language use. Kuijper et al. (2015) found that pronoun use correlated to working memory in ASC ── individuals with poorer working memory capacities tended to use more full

NPs (e.g. *the boy*) rather than pronouns (e.g. *him*). Schaeffer et al. (2018), using an elicitation design, found that children with ASC–ID occasionally overused indefinite articles in contexts in which a definite should be used. This related to poor phonological memory, which was similar in another group of participants with specific language impairment (SLI). Their interpretation suggested that weak phonological memory may cause the speaker to lose track of their discourse representation and so the ability to make definite/indefinite judgments is undermined. Durrleman et al. (2016) identified deficits in accusative clitic production through an elicitation task. Unlike strong pronouns such as *she* in subject positions, clitics such as *la* in French are weak pronouns and argued to involve a more complex syntactic representation. This study explored both third-person clitics and first-person clitics in French speaking children with ASC and SLI in comparison to two TD control groups, matched respectively on chronological age and language level (as assessed by a standardized grammar test). Third-person clitics differ from first-person clitics in that they are anaphoric; and in French they involve obligatory morphological marking for gender while first-person clitics do not. On the other hand, they are not evaluated positively for Person, on the common assumption that the three grammatical Persons can be analyzed through a binary matrix where third Person comes out as -1P and -2P, i.e. the 'non-Person'. While deficits in third-person clitics related to working memory in both the ASC and SLI groups in this study, and additionally to phonological memory in the ASC group, there was no such correlation with low performance on first-person clitics in the ASC group. Together these studies suggest a weakness in using referential expressions, particularly in the use of anaphora both in the form of pronouns (e.g. *it*) and full definite NPs (e.g. *the ball*) may relate to a weakened ability to maintain a representation of the linguistic discourse. However, as seen from the first-person clitic data,

deficits may affect all pronouns, though for different reasons, and therefore a reduced working memory or phonological memory capacity may not be an exhaustive explanation.

3.1.2. Comprehension of specific types of referential NPs in ASC–ID

Several studies have assessed comprehension of referential NPs in children with ASC including their ability to comprehend indefinite NPs (e.g. *a fish*) in comparison to definite-anaphoric NPs (e.g. *the fish*), definite non-anaphoric NPs that are not anaphoric (e.g. *the fish with rainbow scales*), as well as sensitivity to the syntactic binding of pronouns (e.g. that in a sentence like *Julia's mother is washing her*, *her* refers to Julia in spite of preceding another gender-congruent entity). This literature allows us to explore some of the production data reviewed above in greater detail.

Modyanova (2009) compared the comprehension of indefinite NPs (e.g. *turtle touches a fish*) from definite-anaphoric NPs (*Turtle touches the fish*) among children with ASC–ID, Asperger's syndrome and TD controls. She found that children with ASC–ID, though not Asperger's syndrome, had substantial difficulty understanding definite-anaphoric NPs in comparison to their TD peers, while indefinite NPs were unimpaired. Group differences reduced when the ASC group was matched to TD children on grammatical language level, suggesting that difficulties in the ASC group trend with more general syntactic abilities, which were stronger in the AS group. Anaphora has similarly been identified as problematic regarding comprehension of pronouns by Perovic et al. (2013), particularly in the case of reflexive pronouns bound by complex NPs such as *Bart's dad is washing himself*. In their study, they found that personal pronouns also posed problems, though performance was similar to slightly younger TD children. Reflexive pronouns, by contrast, presented a substantial difficulty that is not part of a typical yet delayed trajectory of language development. In this case, the ASC group appeared to have a weak

understanding of specific syntactic relationships such as binding, which links reflexive pronouns to their antecedents subject to general syntactic constraints. On the other hand, Terzi et al (2014) found reflexive pronouns to be unaffected among Greek children with ASC–ID, while clitic pronoun comprehension was significantly poorer in comparison to TD controls. The participant sample assessed by Terzi et al (2014) was more homogeneous in terms of IQ level, which was within the normal range or above (IQ ≥ 85), while the Perovic et al. (2013) sample had a wide range of IQ scores and mean scores below the threshold for intellectual disability (mean IQ = 67). The type of NP antecedent of the reflexive pronoun also differed between studies. Thus the English structure in the Perovic et al (2013) study had a complex NP (e.g. *Bart's dad is washing him/himself*) while the Terzi et al (2014) had simple NP antecedents (*George is washing him/himself*). This suggest that differences in comprehension of reflexive pronouns may relate either to general psychometric differences of the participants and/or to differences in the complexity of the antecedent. In contrast, pronouns and clitics appear to present more generalized difficulty, even among individuals with autism with cognitive functioning well within the normal range. Together these studies suggest that weakness in syntax may in part drive comprehension deficits on anaphoric NPs in ASC.

Regarding non-anaphoric reference, Bavin et al. (2016) explored how noun-modifying information encoded in NP structure is processed in real time among children with ASC. The referential expressions explored were of the form: adjective + noun + prepositional phrase, such as *the blue square with dots*. The task consisted of the participant looking at a screen where two blue squares were presented that were distinguishable by the post-nominal prepositional phrase. In this case, the definite determiner marks specificity rather than anaphoricity. Performance was assessed through accuracy and reaction time. Results showed that children with ASC–ID were

equally accurate as their TD peers, though slower at updating their selection as they heard the following, disambiguating prepositional phrase. The authors found that performance did not relate to language or IQ in the ASC group, but rather to autism symptoms. They speculated that in real-time scenarios, having a lower processing speed of complex NP subjects may hinder comprehension. This again raises the question whether the deficits in comprehending reflexives identified in the IQ-heterogeneous group of participants with ASC in Perovic et al. (2013), might relate not to a deficit in reflexives per se, but rather to processing the complex subjects to which the reflexive pronouns were bound.

In sum, the Modyanova et al. (2013) and Bavin et al. (2016) studies suggest that comprehension of indefinite and definite non-anaphoric NPs is relatively unimpaired in ASC at least among the section of the autism spectrum that has been under exploration, namely autism without intellectual disability. However, individuals with ASC–ID with greater levels of autism symptoms tended to process definite, non-anaphoric NPs more slowly than TD peers or peers with ASC–ID with lower levels of autism symptoms. Definite anaphoric constructions, on the other hand, were more greatly impaired, which may relate to weakness in syntactic knowledge.

3.1.3 Motivation for the present studies

Thus far, we have outlined current evidence on the production and comprehension of various referential NPs among individuals with ASC–ID. To date, these NP types have generally been looked at separately, adopting different task designs as well as outcome variables (e.g. reaction time, success rate etc.), which makes drawing direct comparisons between these types difficult. Looking at comprehension of NP types in a more systematic way can facilitate these comparisons

and allows considering differences in NP type in terms of grammatical configuration and complexity which may have an effect on performance.

At the beginning of this chapter, we considered how the NP integrates content words into into the discourse context. We may think of the NP as a 'referential template' with a content word 'interior' and a functional 'edge' (i.e. [edge[interior]], which yields various referential implications. In this study, we made use of the fact that NP types may be hierarchically ordered in terms of referential complexity both in relation to structure as well as semantic entailments (Hinzen and Sheehan, 2015; Martin & Hinzen, 2014). For example, if your flat mate tells you, (1a) 'I pet the cat', you would infer she had pet a specific cat (that you had heard of before). It also logically follows that that (1b) 'she pet *a cat'*. Moreover, both statements in turn involve the lexical concept CAT. As such we see that there is a relationship of semantic entailment of indefinites in definite NPs, as also brought out in the earliest and most prominent logical-semantic formalizations of the definite determiner (Russell, 1905). At a structural level, (1a) and (1b) are structurally similar both with regard to the content word *cat* in the respective interiors of the NPs, and the involvement of a determiner in the edge. This element determines the referential functioning of the NP in relation to context, in particular whether a specific, already familiar instance of the kind CAT is referenced, or not. We also find structural differences in these types of NPs with respect to pronouns, see example 2:

2a. I pet [him [-] again today.

2b. I pet [the [cat]] again today.

2c. I pet [a[cat]] again today.[9]

In (2a) and (2b) are the same in terms of specificity and anaphoricity in reference to a specific cat, and both entail (2c). What differentiates (2a) from (2b) and (2c) is the absence of an NP interior. In this case, we no longer have the descriptive content to guide referential anchoring but rather have to rely on features such as gender or number to indicate structural relationships. Together, we can depict the system of NPs as in Figure 1:

[9] Beyond the types of NPs that will be considered in the study, the referential hierarchy goes further in that there is an additional indefinite NP form in which the edge is empty, e.g. 'I pet cats', which corresponds to generics.

Figure 1. Referential hierarchy

		DP type	Edge	Interior	Reference type (and entailments)	Referent
–	Interior descriptive content	Clitic/pronoun	I pet	[him [—]]	A (B,C)	
+	Interior descriptive content	Definite (anaphoric)	I pet	[the [cat]]	A (B,C)	
		Definite (non anaphoric)	I pet	[the [cat [with orange stripes]]]	B(C)	
		Indefinite	I pet	[a [cat]]	C	

Specificity depends on determiner choice, by which deictic, definite, and indefinite forms of reference and the grammatical context. Higher anaphoric forms may also, in addition to differences in structural complexity, entail differences in processing load. In particular, definite NPs must be able to link back to earlier in the discourse when the referent was last introduced, which may be relatively local or spanning larger chunks of discourse. Pronouns or clitics have the same dependence on maintaining representation of the discourse, while also lacking descriptive content. Under this light, we may consider how patterns of impairment described in the literature may meaningfully relate to the NP system, in that greater difficulties in comprehension have been found in anaphoric NPs. These studies will assess the system of different referential NPs together and under the same test conditions, which allows us to address two research questions: (1) will participants with ASC have greater difficulty understanding anaphoric reference (such as definite anaphoric NPs or clitic/pronouns) relative to controls? (2) whether the pattern described in the literature so far is characteristic of autism generally, as the research to date has focused primarily on the subset of autism which does not have concurrent intellectual disability (ASC–ID). By including verbal individuals with intellectual disability (ASC+ID), we will not only be able to better understand the pattern of comprehension weaknesses among NP types, but also whether the patterns of weakness identified in the literature in the ASC–ID group hold for the ASC+ID group as well, or whether difficulties permeate through to all of the NP types. Neither production nor comprehension of different NP types has been specifically explored in ASC+ID to date.

In order to assess these questions, we will present two studies which explore comprehension of referential NPs across the autism spectrum. Study 2 compares comprehension of NP types which have been reviewed here (indefinite NPs, definite non-anaphoric NPs, definite

anaphoric NPs and inflectional pronouns/clitics), while Study 3 takes a deeper look at anaphoric reference through an alternative design that hinders an exophoric representation of definite anaphoric NPs (referred to as Target Condition, in the predictions listed below). The specific predictions of these studies are outlined below:

3.1.4 Predictions of Study 2 (Book task):

- We predicted greater difficulty in the anaphoric NP types (definite anaphoric NPs and Pro/clitics in the ASC–ID in comparison to TD controls.
- We predicted a significantly poorer performance in the ASC groups (+/- ID) when comparing comprehension of the maximally complex referential NP type (pro/clitic) in comparison to non anaphoric NP types that are similarly specific (Definite non-anaphoric NPs). We selected the definite non-anaphoric variable to compare to the *pro* as within the definite NPs, they occupy lowest the and highest referential ranks, respectively when controlling for specificity. This is to say, that indefinite NPs had more than one exemplar of its kind in the scene while the rest of the NP types there was only one unique target referent. By controlling for specificity, we are better able to compare higher and lower referential ranks.

We remained neutral to predictions of comparisons between the ASC+ID group and the ID control group, as there is little data to date on referential tracking within these groups. Through this comparison, we were interested in seeing whether the pattern of deficits that may be found in the ASC+ID pattern with that of the ID group or whether we may identify ASC specific weaknesses that therefore not be attributable to a generalized lower IQ level.

3.1.5 Predictions Study 3 (Box task):

- We predicted that the ASC–ID participants would have significantly fewer correct answers to the test question in the Target Condition in comparison to the typically developing group.
- We also predicted that the ASC+ID group would have significantly fewer correct answers to the test question in the Target Condition than the ASC–ID group.
- We predicted a lack of significant group differences in the Filler Condition across any group pairs (ASC–ID and TD, ASC–ID and ASC+ID, and ASC+ID and ID).

We remained neutral to the relative performance of the ID group due to the relatively few studies on anaphoric reference in either population. The comparison between groups with ID allows us to explore whether participants with ASC+ID perform significantly different that children with ID without autism which may give insight to ASC-specific profiles. However, if they perform similarly, lower performance may relate to general lower cognitive functioning capacities in both groups.

3.2 Methodology

3.2.1 Participants of both studies

A total of 48 children and adolescents with ASC participated in the study as well as 43 controls. The ASC and control participants were grouped according to presence or absence of intellectual disability (IQ ≤70), therefore creating four distinct groups: ASC without intellectual disability (ASC–ID, n=34), ASC with intellectual disability (ASC+ID, n= 14), typically developing children (TD, n =34) and children and adolescents with intellectual disability without an autism diagnosis (ID, n= 9). All participants were Spanish-Catalan bilingual and recruited from the Barcelona and

Tarragona Provinces. Table 1 presents the chronological age and standardized test scores for all participants.

Table 1: Participant information

	ASC–ID (n=34)	TD (n=34)	ASC+ID (n=14)	ID (n=9)
CA	9.81	9.03	13.143	13.07
range	6.91-12.91	7-11.5	6.85-20.67	11.92-15.58
VMA	9.35	9.62	6.34	7.27
range	5.42-16.75	6.58-13.75	3.75-12.08	5.08-10.08
IQ	98.63	110.73*	51.79	47.63
range	70-146	91-146	43-67	41-56
Working memory	99.84	104.4	59.23	58
range	65-133	85-130	50-82	50-72
Processing speed	94	102	62.85	59.38
range	70-119	53-126	50-82	54-67

CA= Chronological Age; VMA = Verbal Mental Age (PPVT-III).

Inclusion and exclusion criteria

All participants in the ASC groups (ASC+/-ID) were diagnosed by clinical psychologists according to the diagnostic guidelines established in the ICD-10. Inclusion for ASC participants consisted of reaching diagnostic thresholds for autism symptoms as assessed by the ADOS (Autism Diagnostic Observational Schedule) and/or the ADI-R (Autism Diagnostic Interview-Revised). All participants in the TD group were not known to have any developmental disability and were not under the care of a speech pathologist. All participants in the ID group were recruited from special education schools for children and adolescents with intellectual disabilities and ID was confirmed though passing the Wechsler Intelligence Scale for Children –IV (WISC-IV. The

ID participants were further screened for autism symptoms by the Autism Spectrum Quotient (AQ) and four participants were excluded for showing substantial autism-related traits as per the AQ.

Matching procedure

The ASC–ID were matched to TD participants on verbal mental age as assessed by the PPVT-III. Upon comparing IQ scores which were available for a subsection of the TD participants, Man-Whitney U tests revealed that there the TD group had a significantly higher total IQ (U=349.5, z= 2.500; *p*=.012, however the groups did not differ significantly in working memory (U=245.5; z=1.131; *p*=.258) and approached a significant difference in processing speed (U=294; z=1.896, *p* = .058). Due to limitations in recruitment, we were not able to closely match the ASC+ID participants to the ID participants. However, there was no significant difference in verbal mental age between groups. The ID and ASC+ID groups also did not differ in IQ (U= 42; z=-1.029; *p*=. 330) or measures of working memory (U=49.5; z=-.457; *p*=.664) or processing speed (U=45.5; z=-.491 *p*=.645).

Table 2. Verbal Mental Age (VMA) matching; p-values indicated

	ASC-HF (n=34)	ASC-LF (n=14)
TD (n=34)	.719	---
ID (n=9)	---	.336

Ethical procedure

Families of all participants were informed of the aims and procedures of the study and provided written consent to participate. The research plan was assessed and approved by the Vall D'Hebron University Hospital Ethical Board in Barcelona, Spain.

3.3 Study 2: The book task

Materials and procedure

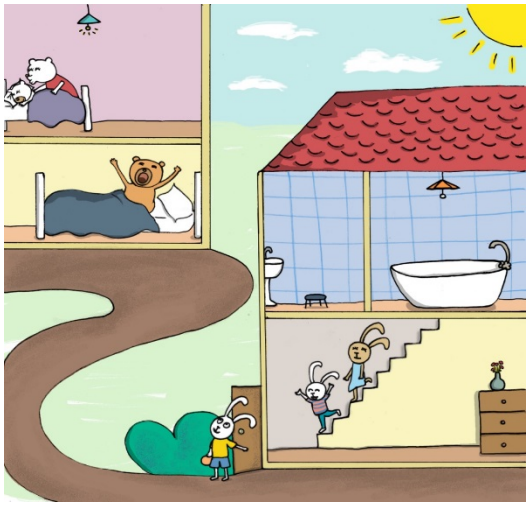
The present study consists of a reference tracking task that assesses the comprehension of referential phrases in a systematic way. A wordless picture book was designed and illustrated for the purposes of this task. The picture book is comprised of six scenes, each involving six characters (a family of three bears and a family of three rabbits) who engage in different actions in each scene. The experimenter explains to the participant that there are two families who spend the day together. The child is then instructed that he or she will hear a sentence and that they should then point to the corresponding object or character on the page and given a trial sentence. The sentences indicated by the examiner differ in referential scope and correspond to six NP types (indefinite NPs, definite non anaphoric NPs, definite anaphoric NPs and inflectional pronouns and clitics, called here 'pro/clitic'). For example, the participant hears an utterance, such as 'show me *a rabbit,*' and indicates one of the rabbits on page one. Then, on the following page, the experimenter says, 'where is the rabbit here?'. A correct response would be selecting the rabbit previously on page one. There are six items per type, twenty-four items total (see Table 3 for examples of NP types).

Table 3. NP types and examples

Type	Example	
	Spanish	English gloss
Indefinite	*Enséñame un conejito*	*Show me a rabbit*
Definite Non-Anaphoric	*Enséñame la osita con el vestido azul*	*Show me the little bear in the blue dress.*
Definite- Anaphoric	*¿Dónde está el conejito aquí?*	*Where is the rabbit here?*
Pro/clitics	*¿Donde está ahora? (el regalo)*	*where is it now? (the present)*

The characters have common and distinguishing features between each other. For example, the mother rabbit and little bear are both wearing blue dresses, therefore for the Definite Non-Anaphoric item 'show me the little bear in the blue dress' there is a foil option (the mother rabbit in the blue dress). See Figure 2.

Figure 2. *Page 1 and 2 of story book.*



*Enséñame **un** conejito*
*'Show me **a** little rabbit'*

*Dónde está **el** conejito aquí*
*Where is **the** rabbit here?*

The visual presentation of the book facilitates working memory load as two pages were made visible to the participant at all times. The sentences types were presented in a pseudo-random order. The session was video-recorded and all responses were marked in a score sheet. The task took approximately 5-10 minutes per child. Immediate self-corrections were not penalized. Correct responses were tallied for each NP type, created a maximum score of 6 points per NP type. The participants were given neutral positive feedback.

Analysis plan

The results were analyzed using SPSS, version 24. Non-parametric bi-lateral Mann-Whitney U-tests were applied in between-group comparisons for all NP type variables, given the skewedness of the data and a relatively small group sizes for the groups with intellectual disability. We also

conducted within-group comparisons between two definite types of NPs, pro/clitic and definite non-anaphoric NPs. Correlational analyses were run between the language and cognitive variables (verbal mental age, working memory and processing speed) and task success, provided significant group differences were identified. Other subtests of the WISC were not explored as there was variability in the versions of the WISC that were passed between some of the ASC and TD participants (WISC-IV and WISC-V, respectively), which led to differences in the way that perceptual reasoning was assessed. Further correlational analyses were not attempted for other NP types so as not to accrue a greater probability of incorrectly rejecting the null hypothesis due to the confound of multiple comparisons. Results are provided in box-plots graphs to illustrate the distribution of correct responses across groups.

3.4 Book task results

3.4.1 Group comparisons

Table 4 presents the comparison of scores for the paired group comparisons (ASC–ID vs TD; ASC+ID vs ID; and ASC–ID vs ASC+ID). Results show that comprehension of all NP types was relatively high across all groups, and at or near ceiling in the groups without ID. The ASC+ID group had significantly poorer comprehension relative to the ASC–ID group, and had significantly poorer comprehension of pro/clitics relative to the ID group. Summaries of paired group comparisons and corresponding figures will follow Table 4.

Table 4. Non-parametric pair-wise group comparisons of NP types

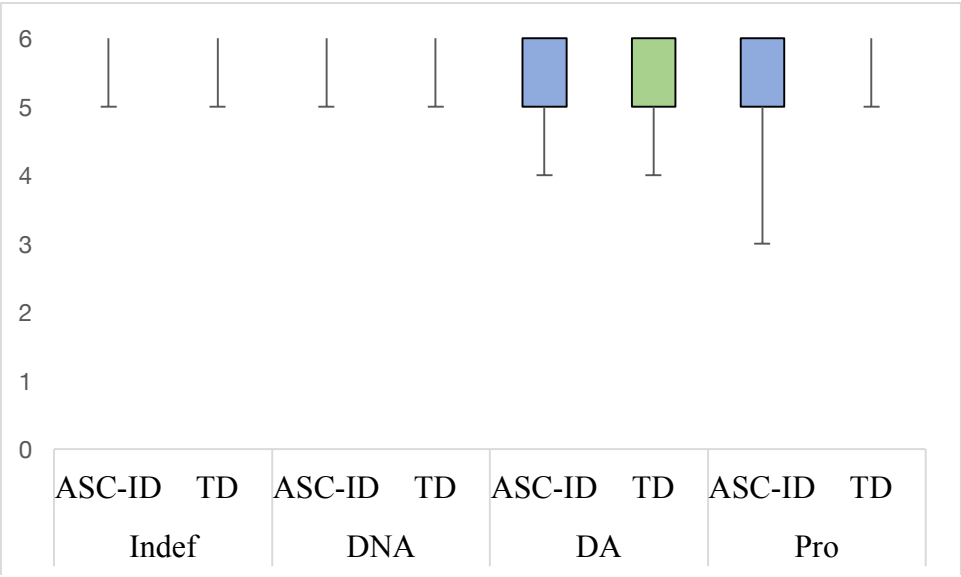
NP type	Groups	Mean rank	Median (range)	Group comparison	U	Z	p
Indefinites	TD	35.01	6 (6-6)	ASC-ID vs TD	595.5	.021	.984
	ID		5 (4-6)	ASC+ID vs ID	63.5	.034	1.0
	ASC-ID	34.99	6 (5-6)	ASC-ID vs ASC+ID	124	-3.987	.0001***
	ASC+ID	11.96	5.5 (3-6)	–	–	–	–
Definite non anaphoric	TD	33.61	6 (5-6)	ASC-ID vs TD	546.5	-.859	.390
	ID		6 (5-6)	ASC+ID vs ID	93.5	2.110	.053
	ASC-ID	36.43	6 (5-6)	ASC-ID vs ASC+ID	110	-3.642	.0001***
	ASC+ID		5.17(4-6)	–	–	–	–
Definite anaphoric	TD	36.71	6 (4-6)	ASC-ID vs TD	655	.832	.405
	ID		5 (4-6)	ASC+ID vs ID	80.5	1.157	.277
	ASC-ID	33.24	6 (4-6)	ASC-ID vs ASC+ID	142	-2.353	.019*
	ASC+ID		5 (4-6)	–	–	–	–
Pro/clitics	TD	38.40	6 (5-6)	ASC-ID vs TD	714	1.804	.071
	ID		6 (3-6)	ASC+ID vs ID	103	2.575	.011*
	ASC-ID	31.50	6 (3-6)	ASC-ID vs ASC+ID	57	-4.372	.0001***
	ASC+ID		3.5 (3-6)	–	–	–	–

* indicates p≤.05; ** indicates p≤.01; *** indicates p≤.0001

ASC–ID vs TD comparison

The ASC–ID group and TD controls showed high performance on all NP types – the median score being the maximum score (6) on all NP types for both groups. We predicted increased difficulty in the referential structures that were less dependent on lexical content – especially pro/clitics. We did not identify any significant difference between groups in pro/clitics, however we note that this was the only NP type that approached significance p≤.1). Performance on clitic/pro was indeed high among participants with ASC–ID, however we note descriptively a wider range of variability than their verbal mental age matched TD peers, as seen in Figure 3.

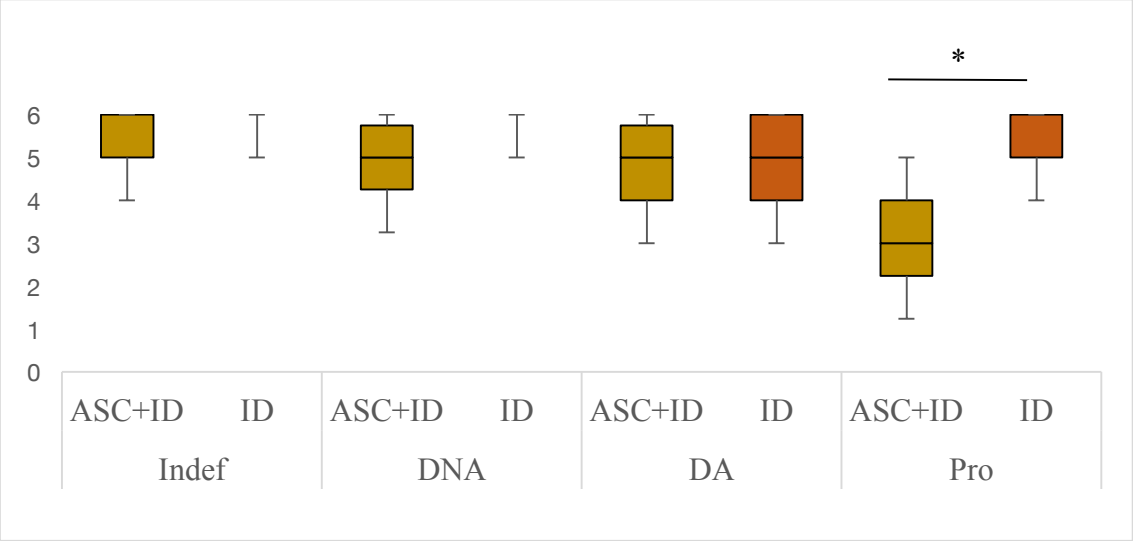
Figure 3. Comparison of ASC–ID and TD groups on NP types using Mann-Whitney U tests.



ASC+ID vs ID comparison

Regarding the two groups with intellectual disability (ASC+ID & ID), the results show that the participants with ASC+ ID had significantly poorer comprehension of pro/clitics than those with ID (p=.011). This pattern of the pro/clitic NP type being the variable that best distinguished the paired groups was similar to that of the ASC–ID and TD comparison. We also note that although other NP types were not significantly different, the ASC+ID group tended to have a more variable performance which approached significance in the case of definite non-anaphoric full NPs (p=.053). Group differences are shown visually in Figure 4.

Figure 4. Comparison of ASC+ID and ID groups on NP types

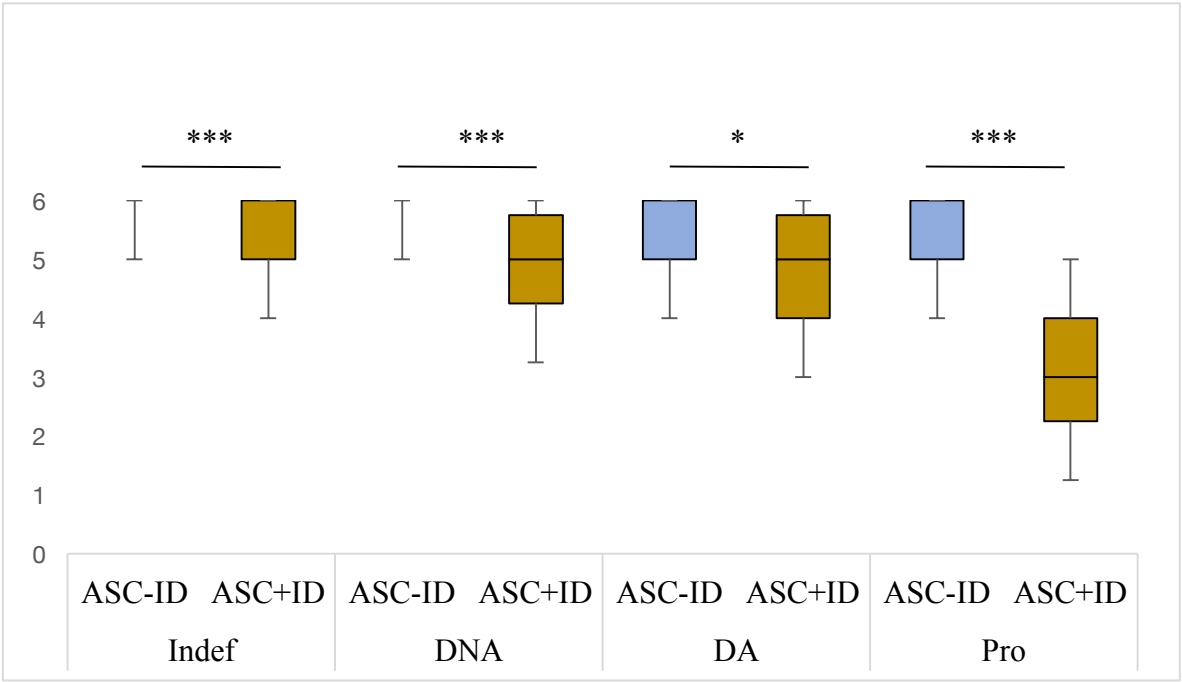


* *Indicates p*=.011

ASC–ID vs ID comparison

Regarding the comparison between both ASC groups, with and without intellectual disability, the results show that the ASC+ID group underperformed significantly in all NP types relative to the ASC–ID group. In general, there was a greater range of variability among the ASC+ID group (Figure 5).

Figure 5. Group comparison between ASC–ID and ASC+ID on NP types



* *Indicates p*≥.05; ***indicates p≥.001

3.4.2 Within-group comparisons

Table 5 presents within-group comparisons between definite non-anaphoric and pro/clitic variables. Significant differences appeared between these variables in both ASC groups (+/- ID). There were no significant differences between these variables in the TD or the ID groups.

Table 5. Wilcoxon signed rank for within group comparison of definite non-anaphoric and pro/clitic

	Def.non-anaphoric Median (range)	Pro/clitic Median (range)	z	*p*
ASC–ID	6 (5-6)	6 (3-6)	-2.207	.027*
ASC+ID	5 (3-6)	4 (0-6)	-2.829	.005**
TD	6 (5-6)	6 (5-6)	.333	.739
ID	6 (5-6)	6 (3-6)	-1.518	.129

3.4.3 Correlational analyses

The group comparison and within-group results indicated that the pro/clitic NP type was a source of weakness within the two ASC groups and was a discriminating variable between the ASC+ID and ID groups. Correlational analyses between neuropsychological variables (full IQ, verbal mental age, working memory and processing speed) and linguistic variables were thus conducted with pro/clitics. No statistically significant correlations were identified for any of the groups between pro/clitic and verbal mental age, full IQ, working memory or processing speed.

Looking at the ASC groups together, correlations were found across the board: success on pro/clitic was highly correlated with verbal mental age, IQ, working memory, and processing speed (with r and *p* values of at least $r \geq .6$; $p \leq .000s1$). Taking IQ as a continuous variable, the TD and ID groups were also pooled. No significant correlations emerged between the pooled TD and

ID control groups and the neuropsychological variables (with r and *p* values of maximally r≤.342; *p*≥.109) under evaluation.

3.4.4 Results summary

The results showed overall high performance for participants without intellectual disability. The ASC–ID group showed greatest difficulties in definite anaphora and pro/clitics in absolute terms, however there were no significant differences in comparison to TD controls. Regarding within-group analyses, the ASC–ID group showed significantly poorer comprehension of pro/clitics in comparison to definite non-anaphoric NPs, a pattern which was not found among TD participants.

Regarding ASC+ID, we found a similar pattern to that of ASC–ID in that pro/clitics had the lowest performance, although performance in ASC+ID was significantly lower in relation to the ASC–ID group in all NP variables. Interestingly, the ASC+ID group showed similar performance to the ID group on all NP types except pro/clitics, where the ASC+ID group had significantly lower scores. Regarding within-group analyses, the ASC+ID group, like the ASC–ID group, had significantly poorer comprehension of pro/clitics in relation to definite non-anaphoric NPs. In contrast, there were no significant differences between pro/clitics and definite non-anaphoric in the ID group. No significant correlations were found between cognitive mechanisms and performance on pro/clitics within each group. However, we found correlations across the board (with verbal mental age, IQ, working memory, processing speed) in the ASC group when all ASC participants were pooled together. No correlations were found for pooled control groups between the pro/clitic NP variable and independent variables.

3.5 Discussion of Study 2 (Book task)

The results of this study show a high performance on comprehension of NPs across the ASC–ID and TD groups, suggesting that these NP types do not pose a great difficulty for comprehension within the context of this task. However, despite a generally high level of performance, we did identify patterns of performance specific to the ASC groups: individuals with ASC tended to have greater difficulty identifying the intended referent of pronouns or clitics relative to definite non-anaphoric NP types as well as relative to their paired control groups. To interpret these results, it is important to keep in mind the context in which comprehension of NP types was assessed. As the task materials consisted of a picture book, resolution of anaphoric reference may have been bolstered by the visual nature of the task as the antecedent is often visually maintained while the definite-anaphoric NP or inflectional pronoun or clitic was targeted (i.e. the antecedent is often first anchored on the left-hand panel and the anaphoric referent is elicited on the right). Nonetheless, referential anchoring requires knowledge that an anaphoric NP refers to the most recently engaged referent of its type (e.g. if we are looking at 'the bear in the red dress' and then the child is asked 'what happened to her here?' on the next page, that *her* necessarily refers to *the bear*). However, how the anaphoric reference is resolved may be different than in typical verbal interactions in which implicit pronouns and clitics generally refer endophorically rather than exophorically. As Fine et al. (1994) noted, individuals with autism tended to engage more in exophoric rather than endophoric reference. We also note that anaphoric reference was supported through the immediacy of the anaphoric elicitation (in cases of anaphoric reference, either pro/clitic or definite-anaphora, the target NP was requested immediately after the initial presentation of the referent). Nonetheless, it is telling that despite these supports, a relative weakness regarding implicit pronoun and clitics is identifiable among the ASC groups, which was

not identified among the TD or ID controls. Furthermore, we see that although the ASC+ID group performed consistently lower across all NP types, this dampening of performance may not be attributable to the psychometric differences between groups alone, as the ID group, with similar levels of intellectual disability, demonstrated significantly better comprehension on implicit pronouns/clitics. This, along with the trend-level difference between ASC–ID and TD controls in the pro/clitic variable, and the within-group differences between the pro/clitic variable and the definite non-anaphoric variable, suggests that resolving anaphoric reference – at least in the context of implicit pronouns or clitics – is a weakness that may relate to an ASC profile.

It is relevant to note that we identified tend-level or significant group differences for the implicit pronoun/clitic NP type and not the definite-anaphoric NP type. This fits in part with our predictions in that we predicated that pro/clitic would be the most impaired, however we also predicted that the definite-anaphoric variable would be significantly lower in the ASC–ID group in comparison TD groups, which we did not find. Looking at the performance distribution between groups, we see that descriptively the ASC–ID group performed similarly as TDs on both definite-anaphoric NPs and pro/clitics in terms of median score, although with a wider variability in the pro/clitic position. While the TD group performed similarly to the ASC–ID group on the definite-anaphoric NP type, it was the NP type with the greatest variability in that group. Indeed, we see that for this TD group, the definite-anaphoric NP type had a lower performance than the pro/clitic NP type. Looking at this performance pattern, there may be a task effect due to the order of presentation of sentences. Due to the nature of using a picture book in which to establish antecedents relevant for setting up the anaphoric conditions, the order of presentation of NP types was rigid. The first NP type was always an indefinite (show me a rabbit) followed by the definite-anaphoric NP (where is the rabbit here). Upon inspection there was a greater error rate on this first

definite-anaphoric NP relative to subsequent items, suggesting that at least some of the poorer performance may be due to the participants becoming accustomed to the task only more gradually, and which may have been more noticeable in the TD case. This concern cannot be ruled out and hence reflects a limitation of the study design.

All in all, and despite these limitations, we may draw the interim conclusions that regarding NP types, anaphoric NPs that are lightest in lexical content (pro/clitics) tend to pose greater difficulty relative to non-anaphoric NP types for participants with ASC, and that increased difficulty among the ASC+ID group may be indicative of an autism-specific pattern rather than that of a general dampening down of performance due to concurrent intellectual disability. Validation of this results depends on further studies addressing the limitation above and increasing the size of the ID samples.

3.6 Study 3: The box task

Aims

This study tracks the comprehension of definite-anaphoric NPs. Unlike study 1 (the book task), Study 3 tracks comprehension without the benefit of having the antecedent visually available. Indeed, in this task we isolate the edge of the NP so that the content word in the interior of the NP is not able to facilitate performance. The merits of this design for assessing comprehension of definite NPs will be explored in greater detail after the presentation of the materials and procedure. Recall the predictions enumerated on page 68.

Materials

The task materials consisted of a medium-sized blue box that has an opening cut on one side with a felt flap attached so that the participant may place his or her hand inside the box and feel an object without seeing it. The task items consisted of five sets of paired objects and three unpaired objects[10]. In four of the sets (the Target Condition), the object pairs had the same kind-membership yet they were differed in terms of a property, for example one pair was a blue colored pencil and a pink colored pencil – both are pencils but differ in the property *color*. In one additional set (the Control Condition), the paired objects were identical in shape yet represented different kind-categories ('orange' and 'watermelon'). The three objects that did not have a pair were called 'filler' objects as they were used to alternate between Target and Control conditions. This ensured

[10]Initially a sixth set was included (an empty jar and a jar full of marbles), however it was later excluded due to low success rate relative to other object pairs. We considered this object to be unsuitable because, unlike the other object pairs in the Target Condition, the property that was being manipulated (empty/full) was not intrinsic to the object and many children believed that the jar was the same despite of the fact that it had been filled while their eyes were closed rather than being a different jar.

that the pattern of correct responses would be a mix of both 'yes' and 'no' responses. The materials are described in Table 5 including the test questions in each instance and correct answers. All participants had been previously passed a short color naming task to ensure that the participant has no difficulty discriminating colors.

Procedure

Environment:

The task took place in a quiet room with the experimenter sitting adjacent to the participant at a table. The session was video-recorded and responses were recorded in a score sheet by the experimenter. As in Studies 1 & 2, the language of this task was Spanish.

Showing phase:

The experimenter showed the blue box to the participant and opened it to show that it was empty. Next the experimenter showed an object, hereafter known as the 'Shown Object' (for example, a blue pencil) to the participant, and asked 'What's this?' At this time, the participant was able to look at and touch the 'Shown Object'.

Hidden phase:

Next, the participant was told to close his or her eyes. While the participants' eyes were closed, the Shown Object (e.g. the blue pencil) was put away and the 'Hidden Object' (in this example, a *pink* pencil) was placed in the box.

The participant was then told they could open their eyes and place their hand inside the box to feel the Hidden Object. The child was asked the question, 'Is it the [Shown Object]?' In the above example, the participant was asked, 'Is it *the pencil*'? Note that the question asked always had the form of 'the + noun' without any adjective identifying properties.

Test phase:

In this final phase, the child was asked to take the hidden object out of the box and the experimenter asked, *What is it?*. The experimenter then asked the test question: *Is it the 'Shown Object'* (in this example, *Is it the pencil*)? The dependent variable compared across groups were the mean correct responses to the test question, *is it the [Shown Object]?* in the Target Condition. Correct responses are indicated in Table 6 below.

Table 6. Box task materials and test questions

	Shown Object	Hidden object	Set Difference	Test question	Correct response
Target items					
T1.	Blue colored pencil	Pink colored pencil	Color	Is it the pencil?	no
T2.	Plastic white ball with bell inside	white Styrofoam ball	Texture and sound	Is it the ball?	no
T3.	White feather	Blue feather	Color	Is it the feather?	no
T4.	Miniature clothespin	Normal-sized clothespin	Size	Is it the clothespin?	no
Control items					
C1.	Wooden orange slice	Wooden watermelon slice	Kind	Is it the orange?	no
Filler items					
F1.	Toy laser gun	Toy laser gun (NS)	N/A	Is it the pistol?	yes
F2.	Toy doll	Toy doll (NS)	N/A	Is it the doll?	yes
F3.	Toy dolphin	Toy dolphin (NS)	N/A	Is it the dolphin?	Yes

NS= no switch

As noted, in the four Target Condition trials, the child was shown an object in the showing phase and then a different object that belonged to the same kind category, yet importantly differed in terms of property, was hidden in the box. For some of the object pairs, the participants could

'feel' in the box that the object had been changed, such as when there was a difference in size or texture; in other case, as in the case of change of color, this was not possible. Alternating these perceptual variables kept the activity engaging to the participant. Asking the child the question, 'Is it the [Shown Object]?' while they were feeling the hidden object gave the child additional reinforcement to link the NP *the pencil* to the object presented in the Showing Phase. These target trials were interspersed with the three filler trials in which the object that was shown (e.g. a plastic toy dolphin) was the same that was placed inside the box during the hidden phase. This ensured that correct responses to the test question varied in being either yes or no depending on whether a Filler or target item was presented. Two orders were generated for order of presentation of trials, which was counterbalanced across participants.

In the Control Condition, the participant was shown a wooden orange slice in the Showing Phase and a wooden watermelon slice was hidden inside the blue box in the Hidden Phase. The procedure followed the same structure as described above. Thus, when the participant removed the Hidden Object, in this case a wooden watermelon slice, he or she was asked the test question, *Is it the orange?*. If the participant answered the test question wrong, replying 'yes' to the question *is it the orange?*, while having the watermelon slice before them, the participant was excluded from the analysis. This was to prevent the possible confounding factor, particularly in the case of participants with low language level, of persevering 'yes responses' in which they answer yes to all questions. Failure on the task in these instances may not be indicative of a problem with tracking anaphora but rather of lower level language or difficulties or attention.

Analysis plan

We conducted a series of paired group comparisons (ASC–ID and TD, ASC–ID and ASC+ID, and ASC–ID and ID) on the mean correct response on the test question in the Target condition. Due to non-normal distribution of data, bi-lateral non-parametric Mann-Whitney U tests were used to assess group differences. Where significant results were identified, effect sizes were calculated. We were also interested in assessing the role of the independent cognitive measures for performance. Specifically, we conducted correlation analyses within each group between task performance and working memory, which was taken from the IQ assessment. Also, where there were significant group differences on the task and there were significant differences between the groups on the independent variables (age, verbal mental age, processing speed), correlations were calculated between the total score of the test condition and said independent variable. The aim of these correlational analyses was to determine whether differences in group performance may relate to group identity or if, alternatively, other factors may be contributing to performance.

3.7 Results

3.7.1 Group comparisons

Table 7 presents the comparison of performance on the test question between the ASC–ID and TD groups. The median score for the TD group was at ceiling (4) while the ASC–ID group had a significantly lower median score of 2, p=. 001. Effect size for this analysis (d =.641) exceeded the convention for a medium effect size (d=.50) according to Cohen (1988). Both groups performed similarly at ceiling on the filler items. A visualization of all group differences is presented in Figure 6.

Table 7. Median and mean rank for correct response for ASC–ID and TD with U, z- and p-values (two-tailed).

		ASC–ID n=34	TD n=34	U	z	*p*
Test condition	Median [max. 4] Mean rank	2 (26.50)	4 (42.50)	850	3.477	.001*
Filler condition	Median [max 3] (Mean rank)	3 (33.97)	3 (35.03)	596	.488	.625

Table 7 presents the comparison of correct response to the test question in the Target and Filler Conditions between the ASC–ID and the ASC+ID groups. Three participants with ASC+ID were excluded from the analysis as they failed the Control Condition and presented a consistent 'yes' response to all questions asked. For the included participants, we found a statistically significant difference between these groups in the Test Condition. Effect size of this analysis (d=1.334) exceeds the convention for a very large effect size (d=1.2) (Sawilowsky, 2009). The ASC–ID group had a median score of 2 (out of 4) and the ASC+ID group had a mean score of 0 out of 4. Both groups, on the other hand, performed at ceiling in the Filler Condition.

Table 8. Median and mean rank for correct response for ASC–ID and ASC+ID with U, z- and p-values (two-tailed).

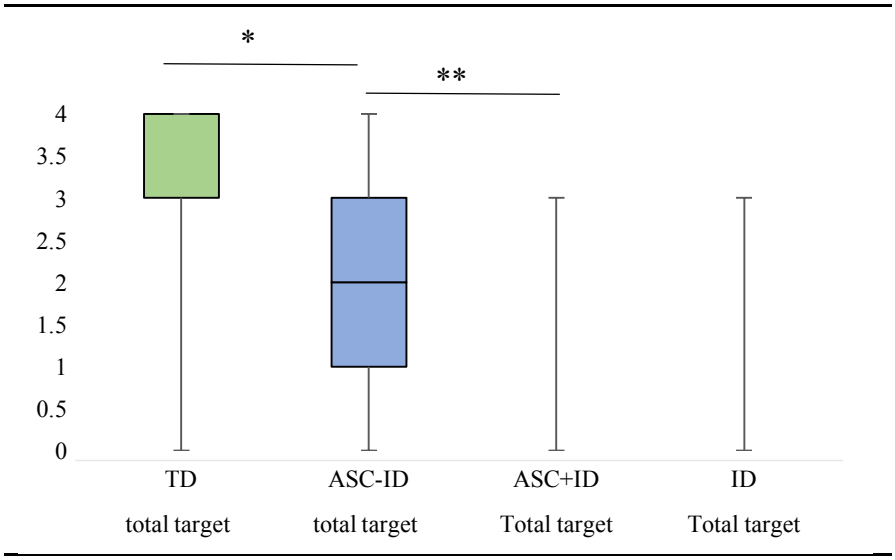
		ASC–ID n=34	ASC+ID n=9	U	z	*p*
Test condition	Median [max. 4] Mean rank	2 (27.66)	0 (11.71)	62.5	-3.689	.0001*
Filler condition	Median [max 3] (Mean rank)	3 (22.97)	3 (25)	222	1.052	.293

Table 9 presents the comparison of correct responses to the test question in the Target and Filler Conditions between the ASC+ID and the ID groups. We note that all ID participants passed the Control Condition and were included in the analysis. Both groups showed a floor effect, with median scores of 0. Similar to all other groups, the ID group also performed at ceiling on the Filler condition.

Table 9. Median and mean rank for correct response for ASC+ID and ID with U, z- and p-values (two-tailed).

		ASC+ID n=9	ID n=9	U	z	p
Test condition	Median [max. 4] Mean rank	0 (10.33)	0 (11.89)	62	.35	.602
Filler condition	Median [max 3] (Mean rank)	3 (11.5)	3 (10.33)	48	-1.155	.702

Figure 6. Correct response in the Target Condition by group



s* indicates $p \geq .001$; ** indicates $p \geq .0001$

3.7.2 Correlational Analyses

We first assessed potential correlations between correct responses to the test question in the Target Condition and measures of working memory as assessed by the IQ assessments (either the WISC-IV or WISC-V). Upon visual inspection, only the ASC–ID group presented an approximate linear relationship between working memory and Target Condition score. A Pearson's correlation indicated, however, that the relation did not reach significance, (r=.271, *p*=.171). There was no relation between working memory and success in the Test Condition within any other group. Although there was substantial variability in the ASC+ID and the ID groups in terms of their working memory levels (ASC+ID: mean = 60, range = 50-82; ID: mean = 58, range 50-72), performance on the Target Condition was consistently at floor. Similarly, there was no linear relationship between task performance and working memory in the subset of the TD group (*n*=15) for whom working memory scores were available.

As the ASC–ID and the TD groups were individually matched on verbal mental age, we may infer that verbal mental age is unlikely to relate to the group differences identified. Regarding the significant differences in task performance between the ASC groups +/- ID, as well as significant differences in the independent variables (verbal mental age, IQ, working memory and processing speed), we assessed through visual inspection whether a linear or monotonic relationship was present within both groups in order to proceed with a correlation analysis. We found that upon visual inspection there was no approximate linear or monotonic relationship within either group relating to task performance and any of the other independent variables, namely IQ and processing speed.

3.7.3 Results summary

We identified significant group differences between the ASC–ID and TD groups and the ASC–ID and ASC+ID groups on correct response to the test question in the Target Condition (Figure 5). Children in the TD group showed a near-ceiling performance (though with some outliers) while children in the ASC–ID group presented a widely variable performance. Both the ASC–ID and ID groups showed a floor performance. Correlation analyses did not identify any relation between the independent variables and task success. Working memory, though not significant, was the only variable which presented an approximate, positive linear relationship with task success among the ASC–ID group.

3.8 Discussion of Study 3 (Box task)

In this study, we identified significant group differences between ASC–ID and TD participants, suggesting that individuals with ASC do indeed present greater difficulty comprehending definite-anaphoric NPs in the scenario established in this study design. The design of this experiment allowed us to isolate the comprehension of the edge of the NP, specifically its anaphoric interpretation. Although the test question in the Filler Condition is equally definite and anaphoric (e.g. *is this the dolphin*?), a correct response is not dependent upon an anaphoric interpretation, i.e. if the participant is only concerned with the content word 'dolphin', the response will still be correct. We also note that when the participants made a mistake in the Target Condition, they often did so with a response that commented on the change that had occurred. For example, in the blue feather/white feather set, when asked the test question, 'Is it the feather?', children who failed tended to provide a response such as, 'yes, but white'. As such we see that the participants, even when failing, did not appear to fail due to a lack of remembering what had been shown to them in

the showing phase as they were commenting on the relevant change in property. Another possibility is that the participants who failed had difficulty tracking the common ground, as it is difficult to disentangle the ability to track common ground from anaphoric reference in comprehension. However, again the types of comments that accompanied wrong responses suggest that the child may be aware of the object that had been in the common ground due to the participants' initiative as to comment on the distinguishing property. We are not however able to completely reject these potential objections as an independent control for recollection was not included.

Considering which general cognitive measures may be driving poorer performance in this task, we did not identify any significant correlations between independent variables and task success. Although there was an approximate linear relationship between task success and working memory scores in the ASC–ID case, the relation was not statistically significant. We also found that in the ASC+ID case, although there were three individuals with working memory scores in the borderline or normal range (two participants with scores of 75 and one with a score of 82), all of these participants performed at floor on this task. In general, correlations were not possible to be assessed within the two groups with ID due to the homogeneity of floor performance. Similarly, there was no linear patterning within the ASC groups pooled together between successful performance and IQ, suggesting that IQ cannot be a predictive factor. The task appeared too difficult for individuals with intellectual disability, however due to the homogeneity of performance it is difficult to disentangle what is driving poor performance. Regarding the participants without intellectual disability, previous studies such as Kuijper et al. (2015) had identified that, in production, working memory relates to the capacity to keep track of referents. Our results do not speak against Kuijper et al.'s (2015) finding, in particular as we did find a

positive yet not significant patterning of stronger working memory skills with task success. However, we found that strong working memory skills is not sufficient to account for performance.

Chapter 4: General Discussion

This thesis has assessed the comprehension of referential NPs among verbal individuals with ASC both with and without intellectual disability as well as narrative language production with particular attention to referential language among children with ASC–ID. Chapter 2 identified various differences between the ASC–ID and TD groups with regard to narrative productive language. Through a narrativity analysis, we were able to assess how the children with ASC–ID structure their language and refer to objects and characters throughout an unfamiliar story and assess how difficulties in the narrative may arise. We found that the narratives produced by the ASC–ID group were construed through shorter utterances and with fewer relative clauses, as part of a pattern of overall reduction in embedded clauses relative to TD controls. We also identified significantly more frequent failures of specific reference within the ASC group, often leading the narratives to become confusing to the listener. The types of referential errors that occurred varied between participants, though overall, suggest that deficits in structural language may in part drive some of the referential breakdown and confusion. For example, agreement errors that cause mismatches between the literal interpretation of the utterance and the scene depicted on the page can lead to confusion on the side of the listener.

The pattern of referential and lexical errors exhibited by the ASC–ID group suggests that narrative deficits may not be attributable, at least in full, to a deficit in taking the listener's perspective into account. However, it is important to note that this does not preclude that our participants may also struggle with theory of mind as this capacity was not independently assessed. Durrleman et al. (2018) found that comprehension of relative clauses correlated with false belief performance among children and adolescents with ASC, and that relative clause comprehension

was significantly poorer than that of TD controls matched on non-verbal IQ. Although we have not assessed comprehension of relative clauses here (though such research is currently ongoing in a separate study), production rate of such clauses was significantly lower than in TD controls in our study, suggesting at least a productive weakness with regard to relative clauses. We did not find differences in frequency or accuracy of use of mental states, nor in attribution of emotion. However, it was notable that only the ASC group had incidences, though rare, of misattribution of emotions.

Studies 2 & 3 from Chapter 3, the book and the box tasks, assessed the comprehension of referential NPs among four participant groups, ASC +/–ID as well as two control groups one with typical development and another with ID without an ASC diagnosis. The two studies presented in Chapter 3 indicate a weakness in understanding certain complex referential expressions among individuals with ASC, particularly NP types in which the content word of the NP is either absent, as in the case of inflectional pronouns or clitics in the book task, or when it is not key, as in the full lexical NPs in the test condition of the box task. Study 2, the book task, assessed a range of NP types that have been explored separately in the literature. By comparing comprehension in the same task with the same participants and conditions, we were able to more systematically assess performance patterns. All in all, we found that comprehension on this task was high, especially for participants who do not have intellectual disability (ASC–ID and TD). The ASC+ID group, on the other hand, showed significantly poorer comprehension in all NP types relative to the ASC–ID group. One of our predictions had been that for the ASC groups there would be a relative weakness in relation to the most grammatical and least lexical NP types, namely inflectional pronouns and clitics (pro/clitic) relative to NPs of lower referential complexity. We specifically compared the pro/clitic to definite non-anaphoric NPs, since this controlled for the element of specificity: there

was only ever a unique referent to fulfill a satisfactory response (e.g. show me *the bear in the blue dress* vs where is *she*?. We identified the same performance pattern among both ASC groups in that pro/clitic NPs were understood significantly worse than definite non-anaphoric NPs. Both the TD and the ID control groups did not present this pattern, showing no significant difference between the pro/clitic and definite-non anaphoric NP variables. For the ASC–ID and TD group comparison, we found that descriptively the ASC–ID group showed a more variable comprehension of pro/clitics than the TD controls though this not reach significance. In the ASC+ID group, we found a similar pattern of this NP type presenting the poorest performance, which also was significantly worse than that of the ASC–ID group. We also identified that the ASC+ID also had a significantly poorer comprehension of pro/clitics in comparison to the ID group who had similar levels of intellectual disability. This suggests that the weakness we find in the ASC groups with regard to pro/clitics may not be attributable to the general cognitive profile of intellectual disability alone. Rather, inflectional pronouns and clitics may be a particular trough of weakness for participants with ASC.

When assessing correlations between successful comprehension of pro/clitic and the cognitive measures that were available to us, we found that in the combined ASC group there were correlations across the board relating to working memory, processing speed, IQ and verbal mental age. Among the TD and ID groups we found no such correlations. This may suggest that for our ASC participants, successful comprehension may rely upon greater cognitive resources to process inflectional pronouns and clitics adequately. Individuals with greater cognitive resources at their disposal may therefore be able to compensate for weakness in this complex NP type. Therefore, the ASC–ID group who, by definition, presents greater cognitive resources, were able to more frequently arrive at high success, though variability patterned with cognitive variability across all

ASC participants. The participants with ASC+ID, in turn, may not have been able to achieve compensation and hence presented significantly poorer success rate.

In Study 3, the box task, we found a homogenously poor performance in response to the test question of the Target Condition among participants with ASC+ID (and ID controls), while the ASC–ID group showed significantly poorer performance and wider variability in comparison to TD controls. In comparing performance between these studies, the book task may have been an easier task due to the visual supports available. In Study 3, the referent was not available visually when it was referred to by the experimenter and indeed another exemplar of its kind was present, hence truly isolating the definite article as the relevant element for consideration in order to perform successfully on the task. This rendered an exophoric interpretation unavailable, as the NP necessarily referred to something that was no longer present in the physical context, yet had been within the recent discourse. We cannot exclude that the participants may have simply forgotten what the *shown object* was. However, we note descriptively that the participants, in addition to answering *yes* or *no* to the target question, also tended to comment on the changed property, suggesting that the participants were aware of the change. Indeed, the task overall was surprising and interesting to the participants across groups. We therefore suggest that the ASC and ID participants, although capable to follow the chain of events, failed to correctly comprehend the test question and the anaphoric nature of the definite article, which necessarily related to the *shown object*. While there was an approximate linear relationship between working memory and task success in the ASC–ID group, a Spearman's correlation indicated that it was not statistically significant. Other studies such as Schaeffer (2018) and Kuijper et al. (2015) did find a relation between working memory and anaphoric reference, with the interpretation that the former may help the individual to maintain a representation of the discourse which is necessary for anaphoric

resolution. In relation to our results, while we identified a linear patterning with task success and stronger working memory, we did not identify a tight or significant relationship. Indeed, individuals with ASC+ID who had working memory scores within the normal range performed equally poorly as others in their participant group with poorer working memory scores.

It is striking that, in both comprehension studies, the elements that best distinguished the ASC groups were those in which there was greater reliance on the grammatical or functional element of the NP (i.e. inflectional pronoun or clitic, or definite article) to give the adequate response. This lends qualified support to our hypothesis that ASC linguistic phenotype is sensitive to the higher levels of the referential hierarchy. There may have been an extra layer of difficulty in Study 3, in that the participants needed to respond that it was *not* the object, despite having an object of the same kind present. However, comparing the box task results to the book task we suspect that this factor cannot alone explain the pattern seen, as the book task also suggests a relative weakness in referentially complex NPs.

In relation to the other NP types assessed in the book task, we saw that differences between the ASC+ID group and ID group approached significance in the definite non-anaphoric condition. It would be interesting to address how relative difficulties in processing non-anaphoric yet complex NPs could affect comprehension. For example, recall that there is an open question in the literature with regard to the comprehension of reflexive pronouns across ASC profiles. Terzi el al. (2014) had found a TD-like comprehension while the participants assessed by Perovic's (2013) presented significant difficulty among a group with moderate intellectual disability. The antecendent NPs to which the reflexive pronoun referred in the in the study by Perovic et al (2013) were also more complex (e.g. *Bart's father is washing him*) compared to those assed in Terzi et al (2014) (e.g. *George is washing him*). Given the relative weakness we identified in definite non

anaphoric NPs in our ASC+ID group, it may be worthwhile to assess further how complex NPs may affect comprehension in ASC+ID mainly when it is grammar that informs the correct interpretation. This may be relevant for example to daily living skills like dealing with currency such as having 'three two-euro coins'.

In terms of total performance, the ASC+ID participants presented a more pervasive impairment of referential expressions in comparison to their peers with ASC–ID and a similar performance to that of to ID controls with the exception of the pro/clitic NP variable which was a signal of distinctive group differences. We also note that the individuals who needed to be excluded from the box task due to failure in passing the Control Condition were members of the ASC+ID group only. These individuals also showed substantial difficulties responding adequately to questions or maintaining conversation throughout the sessions. ADOS scores were not available for all ASC participants, however it would be interesting to explore whether some of the poor scoring of communicative ability in this test could be attributable to increased deficits in language processing. We further note that the task design in the book task likely supported performance of individuals with intellectual disability, given the stark difference between their performance on the book and box tasks. Crucial factors that may be in play include the visual support and immediacy in which the anaphoric NPs were elicited, by which increasing the saliency of the most recently commented on character and so potentially supporting anaphoric resolution.

Overall, the two comprehension studies document referential impairment in ASC+ID that would not be fully explicable by IQ or other specific non-linguistic cognitive factors. However greater cognitive resources may make individuals with ASC more resilient to compensate for weakness in comprehension of complex referential expressions. Residues of such impairment in ASC–ID group are noticeable in the higher regions of the referential hierarchy, where lexical

content does not help in identifying the referent but grammar is required. These studies suggest a pattern of difficulty that spans the spectrum yet is more profound in the ASC+ID case. With regard to Study 1, the narrativity study, we find substantial differences in the language profiles of children with ASC-ID, with particular regard to referential errors, though not restricted to them. The presentation of referential failures along with lexical errors, total grammatical errors and reduction in grammatical complexity, suggested that while referential language may indeed be a particular point of weakness in ASC, these difficulties occur within the context of reduced language skills relative to controls. Through the research developed in this thesis, we hope to focus attention on differences that may present in the language profiles relative to controls, which may relate in part to the heterogeneity found in ASC in that individuals with ASC+ID do indeed have more pervasive language impairment than with ASC–ID. Nevertheless, increased language difficulty may not be reducible to the presence of intellectual disability. In particular, difficulties in comprehending inflectional pronouns and clitics are more pronounced in individuals with ASC+ID, but we cannot attribute this deficit to ID alone, given the relatively high performance of the ID controls. By assessing language across the spectrum, we may thus gain a deeper understanding of how variability presents in ASC.

A further study that this work is calling for would be to assess how deficits in reference relate to communication breakdown as measured non-linguistically in clinical observation and how they factor into diagnosis or severity of communication-related symptoms. Furthermore, future work should include larger sample sizes particularly for the groups with intellectual disability as the conclusions from this thesis are also limited by the small sample sizes available for participant groups with ID. Additionally, a future direction would be to assess in detail measures of real-time

language processing across the autism spectrum to assess how deficits in language may occur in real time and affect the communicative competence of the individual.

References

Adams, C., Green, J., Gilchrist, A., & Cox, A. (2002). Conversational behaviour of children with Asperger syndrome and conduct disorder. *Journal of Child Psychology and Psychiatry*, *43*(5), 679-690.

Allen Preissler, M., & Carey, S. (2004). Do both pictures and words function as symbols for 18-and 24-month-old children?. *Journal of Cognition and Development*, *5*(2), 185-212.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of memory and language*, *56*(4), 521-536.

Baio, J. (2014). Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010.

Baixauli, I., Colomer, C., Roselló, B., & Miranda, A. (2016). Narratives of children with high-functioning autism spectrum disorder: A meta-analysis. *Research in Developmental Disabilities*, *59*, 234-254.

Baltaxe, C. A., & D'Angiola, N. (1992). Cohesion in the discourse interaction of autistic, specifically language-impaired, and normal children. *Journal of Autism and Developmental Disorders*, *22*(1), 1-21.

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. *International journal of speech-language pathology*, *17*(2), 159-171.

Bavin, E. L., Prendergast, L. A., Kidd, E., Baker, E., & Dissanayake, C. (2016). Online processing of sentences containing noun modification in young children with high-functioning autism. *International journal of language & communication disorders*, *51*(2), 137-147.

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child language teaching and therapy*, *18*(1), 1-21.

Boucher, J., Mayes, A., & Bigham, S. (2012). Memory in autistic spectrum disorder. *Psychological bulletin*, *138*(3), 458.

Colle, L., Baron-Cohen, S., Wheelwright, S., & van der Lely, H. K. (2008). Narrative discourse in adults with high-functioning autism or Asperger syndrome. *Journal of Autism and Developmental Disorders*, *38*(1), 28-40

Colonnesi, C., Stams, G. J. J., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review*, *30*(4), 352-366.

Durrleman, S., & Delage, H. (2016). Autism spectrum disorder and specific language impairment: overlaps in syntactic profiles. *Language Acquisition*, *23*(4), 361-386.

Durrleman, S., Hinzen, W., & Franck, J. (2018). False belief and relative clauses in Autism Spectrum Disorders. Journal of communication disorders, 74, 35-44.

Durrleman, S., Hippolyte, L., Zufferey, S., Iglesias, K., & Hadjikhani, N. (2015). Complex syntax in autism spectrum disorders: a study of relative clauses. *International Journal of Language & Communication Disorders*, *50*(2), 260-267.

Eigsti, I. M., Bennetto, L., & Dadlani, M. B. (2007). Beyond pragmatics: Morphosyntactic development in autism. *Journal of autism and developmental disorders*, *37*(6), 1007-1023.

Eigsti, I. M., Stevens, M. C., Schultz, R. T., Barton, M., Kelley, E., Naigles, L., ... & Fein, D. A. (2016). Language comprehension and brain function in individuals with an optimal outcome from autism. *NeuroImage: Clinical*, *10*, 182-191.

Elsabbagh, M., & Johnson, M. H. (2007). Infancy and autism: Progress, prospects, and challenges. *Progress in brain research*, *164*, 355-383.

Fayol, M., Largy, P., & Lemaire, P. (1994). Cognitive overload and orthographic errors: When cognitive overload enhances subject–verb agreement errors. A study in French written language. *The Quarterly Journal of Experimental Psychology*, *47*(2), 437-464.

Fine, J., Bartolucci, G., Szatmari, P., & Ginsberg, G. (1994). Cohesive discourse in pervasive developmental disorders. *Journal of autism and developmental disorders*, *24*(3), 315-329.

Gardener, H., Spiegelman, D., & Buka, S. L. (2011). Perinatal and neonatal risk factors for autism: a comprehensive meta-analysis. Pediatrics, 128(2), 344-355.

Garrido, D., Petrova, D., Watson, L. R., Garcia-Retamero, R., & Carballo, G. (2017). Language and motor skills in siblings of children with autism spectrum disorder: A meta-analytic review. *Autism Research*, *10*(11), 1737-1750.

Hartley, C., & Allen, M. L. (2015). Symbolic understanding of pictures in low-functioning children with autism: the effects of iconicity and naming. Journal of autism and developmental disorders, 45(1), 15-30.

Hinzen, W., & Sheehan, M. (2013). The philosophy of universal grammar. OUP Oxford.

Ibañez, L. V., Grantz, C. J., & Messinger, D. S. (2013). The development of referential communication and autism symptomatology in high-risk infants. Infancy, 18(5), 687-707.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological science*, *16*(5), 367-371.

Kasari, C., Kaiser, A., Goods, K., Nietfeld, J., Mathy, P., Landa, R., ... & Almirall, D. (2014). Communication interventions for minimally verbal children with autism: A sequential multiple assignment randomized trial. Journal of the American Academy of Child & Adolescent Psychiatry, 53(6), 635-646.

Kuijper, S. J., Hartman, C. A., & Hendriks, P. (2015). Who is he? Children with ASD and ADHD take the listener into account in their production of ambiguous pronouns. *PloS one*, *10*(7), e0132408.

Levy, S. E., Giarelli, E., Lee, L. C., Schieve, L. A., Kirby, R. S., Cunniff, C., ... & Rice, C. E. (2010). Autism spectrum disorder and co-occurring developmental, psychiatric, and medical conditions among children in multiple populations of the United States. *Journal of Developmental & Behavioral Pediatrics*, *31*(4), 267-275.

Losh, M., Martin, G. E., Lee, M., Klusek, J., Sideris, J., Barron, S., & Wassink, T. (2017). Developmental markers of genetic liability to autism in parents: A longitudinal, multigenerational study. *Journal of autism and developmental disorders*, *47*(3), 834-845.

Maljaars, J., Noens, I., Jansen, R., Scholte, E., & van Berckelaer-Onnes, I. (2011). Intentional communication in nonverbal and verbal low-functioning children with autism. *Journal of Communication Disorders*, *44*(6), 601-614.

Maljaars, J., Noens, I., Scholte, E., & van Berckelaer-Onnes, I. (2012). Language in low-functioning children with autistic disorder: Differences between receptive and expressive skills and concurrent predictors of language. *Journal of Autism and Developmental Disorders*, *42*(10), 2181-2191.

Manolitsi, M., & Botting, N. (2011). Language abilities in children with autism and language impairment: Using narrative as an additional source of clinical information. *Child Language Teaching and Therapy*, *27*(1), 39-55.

Martin, J.R. Learning how to tell. Unpublished doctoral dissertation, University of Essex, 1978. (Available from University Microfilms International, 18 Bedford Row, London, WCIR 4EJ, #78-70,018.)

Martin, T., & Hinzen, W. (2014). The grammar of the essential indexical. *Lingua*, *148*, 95-117.

Mizuno, A., Liu, Y., Williams, D. L., Keller, T. A., Minshew, N. J., & Just, M. A. (2011). The neural basis of deictic shifting in linguistic perspective-taking in high-functioning autism. *Brain*, *134*(8), 2422-2435.

Modyanova, N. N. (2009). Semantic and pragmatic language development in typical acquisition, autism spectrum disorders, and Williams syndrome with reference to developmental neurogenetics of the latter (Doctoral dissertation, Massachusetts Institute of Technology).

Nayar, K., Voyles, A. C., Kiorpes, L., & Di Martino, A. (2017). Global and local visual processing in autism: An objective assessment approach. *Autism Research*, *10*(8), 1392-1404.

Norbury, C. F., & Bishop, D. V. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders*, *38*(3), 287-313.

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.

Noterdaeme, M., Wriedt, E., & Höhne, C. (2010). Asperger's syndrome and high-functioning autism: Language, motor and cognitive profiles. *European child & adolescent psychiatry*, *19*(6), 475-481.

Novogrodsky, R. (2013). Subject pronoun use by children with autism spectrum disorders (ASD). *Clinical linguistics & phonetics*, *27*(2), 85-93.

Pearce, W. M. (2003). Does the choice of stimulus affect the complexity of children's oral narratives?. *Advances in Speech Language Pathology*, *5*(2), 95-103.

Perovic, A., Modyanova, N., & Wexler, K. (2013). Comprehension of reflexive and personal pronouns in children with autism: A syntactic or pragmatic deficit?. *Applied Psycholinguistics*, *34*(4), 813-835.

Pierce, K., Gazestani, V. H., Bacon, E., Barnes, C. C., Cha, D., Nalabolu, S., ... & Courchesne, E. (2019). Evaluation of the Diagnostic Stability of the Early Autism Spectrum Disorder Phenotype in the General Population Starting at 12 Months. *JAMA pediatrics*.

Preissler, M. A. (2008). Associative learning of pictures and words by low-functioning children with autism. Autism, 12(3), 231-248.

Rumpf, A. L., Kamp-Becker, I., Becker, K., & Kauschke, C. (2012). Narrative competence and internal state language of children with Asperger Syndrome and ADHD. *Research in Developmental Disabilities*, *33*(5), 1395-1407.

Russell, B. (1905). On denoting. *Mind*, *14*(56), 479-493.

Schaeffer, J. (2018). Linguistic and cognitive abilities in children with specific language impairment as compared to children with high-functioning autism. *Language Acquisition*, *25*(1), 5-23.

Schneider, P., & Hayward, D. (2010). Who does what to whom: Introduction of referents in children's storytelling from pictures. *Language, Speech, and Hearing Services in Schools*.

Siller, M., Swanson, M. R., Serlin, G., & Teachworth, A. G. (2014). Internal state language in the storybook narratives of children with and without autism spectrum disorder: Investigating relations to theory of mind abilities. Research in Autism Spectrum Disorders, 8(5), 589-596.

Silverini, S., Prévost, P., & Tuller, L. (2017). Computational complexity and error typology on sentence repetition in children with ASD: is there a phenotypical SLI profile? Paper presented at the Generative Approaches to Language Acquisition (GALA-13), Palma de Mallorca.

Slusna, D., Rosselló, J., Salvadó, B., Rodriguez, A., Hinzen, W. (2018). Use of communicative gestures in school-age children and adolescents with nonverbal or minimally verbal autism. *Proceedings of the INSAR 2018 Annual Meeting*, Rotterdam, 2018.

Suh, J., Eigsti, I. M., Naigles, L., Barton, M., Kelley, E., & Fein, D. (2014). Narrative performance of optimal outcome children and adolescents with a history of an autism spectrum disorder (ASD). *Journal of Autism and Developmental Disorders*, *44*(7), 1681-1694.

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.

Terzi, A., Marinis, T., Kotsopoulou, A., & Francis, K. (2014). Grammatical abilities of Greek-speaking children with autism. *Language Acquisition*, *21*(1), 4-44.

Viding, E., & Blakemore, S. J. (2007). Endophenotype approach to developmental psychopathology: implications for autism research. *Behavior genetics*, *37*(1), 51-60.

Volden, J. (2004). Conversational repair in speakers with autism spectrum disorder. International Journal of Language & Communication Disorders, 39(2), 171-189.

Volden, J. (2017). Autism spectrum disorder. In *Research in clinical pragmatics* (pp. 59-83). Springer, Cham.

Volden, J., Magill-Evans, J., Goulden, K., & Clarke, M. (2007). Varying language register according to listener needs in speakers with autism spectrum disorder. *Journal of autism and developmental disorders*, *37*(6), 1139-1154.

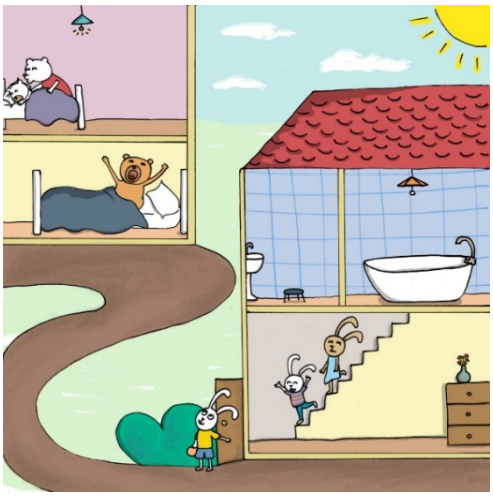
Wiesner, D. (1991). Tuesday. New York: Clarion Books.

Wiklund, M., & Laakso, M. (2019). Ungrammatical utterances and disfluent speech as causes of comprehension problems in interactions of preadolescents with high functioning autism. *Clinical linguistics & phonetics*, 1-23.

World Health Organization. (1992). The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines. Geneva: World Health Organization.

Annex 1: Images involved in the book task in Study 2 in Chapter 3

Page 1&2



Page 3&4

Page 5&6