Making Sense of the Word and the World:

Figurative Language Comprehension in Individuals with Autism Spectrum Disorder

Tamar Kalandadze

Making Sense of the Word and the World: Figurative Language Comprehension in

Individuals with Autism Spectrum Disorder

Making Sense of the Word and the World:

Figurative Language Comprehension in Autism Spectrum Disorder

'All the world's a stage,
And all the men and women merely players;
They have their exits and their entrances'

"As you like it"; Shakespeare, 1975, 2,7,10-12

This extract is about performance on the stage, that is, the world. This doctoral thesis contributes to our understanding of how individuals with Autism Spectrum Disorder (ASD) participate on the world stage and provides insight into how people with ASD could be supported to improve their everyday communication.

Acknowledgments

I am very grateful to have had the excellent opportunity to carry out this research and to the many people who have supported me throughout the PhD-journey. This thesis, quite literally, would not have been possible without them.

First and foremost I extend warm thanks to all children and families who participated in my research.This research would not have been possible without them. I would also like to thank the schools, Autism Assiciation (Autismeforeningen) and the Pedagogical-psychological services for assisting me in the recruitment process.

A special thank my main supervisor, Kari-Anne, who has generously shared her knowledge and experience and who taught me to be a critical thinker. I appriciate all the time and effort you have invested in this thesis, Kari-Anne! Thank you for beleiving that this day would come!

I would also acknowledge Terje Nærland, Mila Vulchanova and Ingrid Lossius Falkum who have helped me with specific parts of my thesis at different stages of the research. Thank you!

My warm and special thanks to the co-authors of the papers: Cecilia Brynskov who has also been my 'informal' supervisor; Thank you for your professional support as well as for our friendship! Valentina Bambini, with whom I share love for metaphors. Valentina, thank you for your support and encouragements, and also for inviting Kari-Anne and me to beautiful Pavia-an unforgettable experience and memory from my PhD-period; Johan Braeken, whose assistance with carrying out the analyses was invaluable. Thank you, Johan!; Courtenay Norbury, whose work on autism and language I admire; Thank you, Courtenay! Special thanks to Morton Ann Gernsbacher and Janne von Koss Torkildsen for their help with the metaphor task created for one of the studies.

I would also like to thank to my colleagues at the Department of Special Needs Education at the University of Oslo. Huge thanks to Kathrine and Marika for answering my questions and offering help.

A very special thanks to all the wonderful PhD-candidates at the department of Special Needs Education. I especially want to thank Jannicke, Anette, Hanne, Anne, Linn, Jana, Arne, Ellen and Marit. You have been particularly important for me during these years, thank you!

Abstract

Pragmatic language ability, including comprehension of figurative language, has traditionally been considered universally impaired in individuals with Autism Spectrum Disorder (ASD) compared to individuals with typical development (TD). However, recent studies reporting no significant group differences, have fuelled debate about the extent of the figurative language difficulties in ASD, as well as factors underlying these difficulties. This PhD-thesis aimed to investigate: a) figurative language comprehension in ASD; b) factors related to performance on figurative language tasks.

In Paper I, a meta-analysis including 41 studies investigated performance of individuals with ASD and individuals with TD on figurative language tasks. Overall, participants with ASD were outperformed by participants with TD (Hedge's g =-0.57). Group matching strategies and differences in figurative language type accounted for significant between-study variation, whilst cross-linguistic differences and age did not. Paper II reports an empirical study in which 28 participants with ASD (10-16 years) performed poorer than controls with TD (N = 31) on a multiple-choice metaphor comprehension task (Δ = -5.63; SE = 2.31; p = .02; Cohen's d =-.66). However, performance could be explained by aspects of language form and content (Δ= -2.55; SE = 2.34; p = .281; d = -.32). Paper III presents a meta-analysis and a systematic review of 14 studies investigating metaphor comprehension in ASD, with a focus on task properties. Individuals with ASD on average scored significantly lower than individuals with TD (Hedge's g =-0.63). Variation in linguistic characteristics of the tasks were found. Verbal explanation response format was associated with the largest group differences.

The findings of this thesis indicate that individuals with ASD, on average, show inferior performance on different measures of figurative language comprehension compared to individuals with TD. However, impaired performance is associated with language form and content, group matching strategies, differences in figurative language type and task properties, rather than diagnostic status *per se*.

Content

Papers I-III

Paper I

Kalandadze, T., Norbury, C., Nærland, T., & Næss, K.-A.B. (2016). Figurative Language Comprehension in Individuals with Autism Spectrum Disorder: A Meta-analytic Review. *Autism,* 22 (2), 99–117. doi:10.1177/1362361316668652

Status: Published

Paper II

Kalandadze, T., Bracken, J., Falkum, I.L., Brynskov, C, & Næss, K.-A.B. (submitted). Metaphor Comprehension in Autism Spectrum Disorder: Language Form and Content matter.

Status: Submitted to *Autism*, 07.04.2019

Paper III

Kalandadze, T., Bambini, V., & Næss K.-A.B. (submitted). 'A Systematic Review and Meta-analysis of Studies on Metaphor Comprehension in Individuals with Autism Spectrum Disorder: Do Task Properties Matter?'

Status: Revised after peer-review and resubmitted to *Applied Psycholinguistics*, 22.03.2019

Note: The papers follow the extended abstract in this thesis.

1 Introduction

1.1   Background

*I was always much slower to understand any new topic and always found
that I was slower to finish my work."*

"I found it difficult to follow the lessons."

*I was switched off. If the support was there I could have accessed the
curriculum".*

These quotes are from the paper by Goodall (2018, p. 9), entitled "I felt closed in and like I
couldn't breathe: a qualitative study exploring the mainstream educational experience of
autistic young people". The quotes describe the reality for many individuals with Autism
Spectrum Disorder (ASD) at schools. Unfortunately, the support students with ASD need
might not always be provided to help overcome difficulties related to understanding and
learning.

Individuals with ASD are increasingly educated in the mainstream schools. Many benefits of
inclusive education such as increased social skills and more advanced educational goals have
been reported (Eldar, Talmor, & Wolf-Zukerman, 2010; Lindsay, Prouix, Thomson, & Scott,
2013). However, students with ASD often struggle academically and socially (Goodall,
2018).

Generally, one of the main obstacles to social and academic success is difficulties related to
language ability. As a powerful tool for academic and social success, language is considered
one of the main factors contributing to optimal outcome in individuals with ASD (Howlin,
2005). Unfortunately, difficulties in different language domains, among them figurative
language (i.e., metaphor, irony; see Chapter 2 for broader definition of figurative language),
an aspect of pragmatics (understanding and effectively using language), has been often
reported in individuals with ASD (e.g., Happé, 1993; MacKay & Shaw, 2004).
Pragmatics is an essential component of social communication, and the impairment in social
communication is a core feature of ASD (Norbury & Sparks, 2014). Therefore, good
command of figurative language is important for successful social communication.

Figurative language invades everyday language and communication and difficulties in its understanding may considerably hamper social communication and learning. Therefore, a better understanding of the extent of difficulties in figurative language and of the variables related to these difficulties will guide interventions aiming at improving language and social communication skills of individuals with ASD.

1.1 Aim and research questions

The overarching aim of this thesis has been to improve our understanding about figurative language comprehension in ASD and to inform future research and practice about how we can predict and facilitate figurative language comprehension in these individuals. The aim was achieved through (1) comprehensively and systematically reviewing and/or meta-analyzing earlier research; (2) empirically examining the extent to which individuals with ASD differ from individuals with TD on figurative language (metaphor) comprehension, and the potential link between difficulties in metaphor comprehension and impaired language form and content in ASD.

The thesis sought to answer two research questions:

Research question 1: What is the extent of the group difference in figurative language comprehension in ASD compared to individuals with TD?

Research question 2: Which variables are related to figurative language comprehension in ASD?

2 Theoretical background

2.1 What is ASD?

'The scientific and medical community's understanding of what autism is has changed dramatically throughout the last century' (Kenny et al., 2016, p.1)

2.1.1 Current conceptualization and the diagnostic criteria

ASD is a neurodevelopmental disorder characterized by persistent difficulties in two domains: social communication and interaction, and restricted, repetitive interests and behaviours (American Psychiatric Association, 2013). ASD has a biological basis, but a reliable biological marker has yet to be found. Therefore, ASD is diagnosed based on a set of observable behaviors (Boucher, 2017; Fletcher-Watson & Happé, 2019).

The terms that have been used in earlier versions before the Diagnostic and Statistical Manual of Mental Disorders, (DSM-5;APA, 2013) to refer to ASD has been 'childhood schizophrenia/psychosis', 'early childhood/infantile autism', and 'pervasive developmental disorders, including Asperger syndrome, autistic disorder and pervasive developmental disorder-unspecified/not otherwise specified (PDD-NOS)' (Boucher, 2017).

The dramatic changes in guidelines and criteria for the diagnosis have been made in the DSM-5 in 2013, reflecting the changes in the understanding of what ASD is. Specifically, the diagnosis of autistic disorder (childhood autism, PDD-NOS, and Asperger's disorder/syndrome, Rett's disorder, and childhood disintegrative disorder) were collapsed into two diagnoses: autism spectrum disorder (ASD) and social (pragmatic) communication disorder. The latter is intended to apply to individuals who have the socio-emotional-communicative (SEC) impairments and differences typical of ASD, but not the restricted and repetitive behaviors (Boucher, 2017).

Since in Norway the International Classification of Diseases (ICD; World Health Organization, WHO) is used, participants in the empirical study (paper II) were categorized according to the 10th version of ICD (WHO, 1992) in which the above-mentioned sub-categories were included. DSM is however used as the framework for any diagnosis-related

discussions in this thesis because the manual tends to be more influential in research and on psychological theory of ASD (Fletcher-Watson & Happé, 2019).

In the DSM-5, the diagnostic criteria are supplemented by descriptors or specifiers, which concern the severity of the core impairments and the presence of any additional conditions or special circumstances. One aspect that is particularly important for this thesis is delayed or impaired language which in DSM-5, differently from the DSM-IV, is listed as a possible specifier and no longer as a possible manifestation of impairment in communication (Boucher, 2017). Delayed or impaired language is accordingly no longer a requirement for an ASD diagnosis. Note, however, that difficulties in language are often found in individuals with ASD, although heterogeneity is pervasive (Naigles & Chin, 2015).

The DSM-5 explicitly states that ASD can be accompanied by other conditions such as genetic disorders (e.g., fragile X syndrome), psychiatric conditions (e.g., attention-deficit hyperactivity disorder (ADHD) and other physical, mental, or neurodevelopmental conditions. These co-existing conditions can have a substantial, often negative, effect on the development and wellbeing of the individual with ASD and their family (Boucher, 2017). The high frequency of the co-existing conditions presents a challenge for research in ASD, because the co-occurrence of other conditions may change the ways individuals with ASD perform on certain (language) tasks and teasing apart the impact of ASD versus other conditions is complicated.

2.1.2 Prevalence

The update in the diagnostic criteria from the DSM-IV to the DSM-5 is the consequence of the changes in conceptual understanding of ASD from being conceptualized and defined as an uncommon childhood disorder to a more frequent lifelong condition (Bishop, 2010; Lord et al., 2018). Indeed, internationally, the number of individuals diagnosed with ASD has been increasing since the 1980s (Boucher, 2017). While a review from 2012 estimated the global prevalence of ASD at about 1% (Elsabbagg et al., 2012), a more recent review suggested the prevalence to be 1-5 % in developed countries (Lyall et al., 2017). In Norway, a prevalence of 51 per 10,000 was reported in two counties (Isaksen, Diseth, Schjølberg, & Skjeldal, 2012).

Based on the scientific evidence, the reasons for this raised number of diagnosed cases of ASD seems to be at least partially related to the broadened diagnostic criteria, greater public

awareness, and improved case finding (Boucher, 2017; Gernsbacher., & Frymiare, 2005). More specifically, from the 1980s, it was recognized that ASD can exist among people at every level of intelligence and therefore, individuals who would not have been included in the spectrum earlier are now diagnosed with ASD (Boucher, 2017; Gernsbacher et al., 2005).

The generally accepted knowledge that ASD is more common in males than in females by a ratio of approximately 4:1 (Fombonne, 1999) is not well understood (Boucher, 2017). Importantly, as opposed to the proposed 'extreme male brain' theory (Baron-Cohen, Knickmayer, & Belmote, 2005), it is now increasingly suggested that there may be many females, especially among more 'high-functioning' individuals on the spectrum, who are currently undiagnosed (Cage & Troxell-Whitman, 2019; Gould & Ashton-Smith, 2011).

Regardless of the reasons, the prevalence of ASD indicates a pressing need for research-based knowledge and understanding of this condition to offer appropriate support to individuals diagnosed with ASD (Fletcher-Watson & Happé, 2019). They will need support in most areas of life, such as schools, university and work (Boucher; 2017; Hurlbutt & Chalmers, 2004; Wei et al., 2013).

2.1.3   Social communication in ASD

Communication in this thesis refers to the ability to receive, send and comprehend verbal, non-verbal and graphic symbol systems (Baird & Norbury, 2015). Social communication is defined as the dyadic social exchanges and includes skills such as discourse processing, narrative and referential communication (Norbury, 2013). According to Adams (2005), 'social communication development is founded on the synergistic emergence of social interaction, social cognition, pragmatics (verbal and nonverbal aspects), and language processing (receptive and expressive)' (p.181). Thus, although pragmatics (see below for a broader definition) is often used synonymously with social communication, it is one of the aspects contributing to social communication (Adams, 2005; Baird & Norbury, 2015).

Difficulties in social communication and social interaction of people with ASD are part of the diagnostic criteria and one of the core or defining features of ASD (APA, 2013). Communicative impairments in children with ASD appear at an early age (Kim, Paul, Tager-

Flusberg, & Lord, 2014) and although difficulties may ameliorate with age for some individuals with ASD (Seltzer, Shattuck, Abbeduto, & Greenberg, 2004), for the majority, including those with intelligence within the normal range, significant difficulties in communication persist (Volkmar & Wolf, 2013).

Although a core feature of ASD, social communication difficulties are manifested in multiple aspects and multiple ways. For example, conversations or narrative production in individuals with ASD can be quite variable. Specifically, sometimes they may interact well, but at other times, they may show difficulties in initiating and maintaining new topics, talking coherently, and producing contextually and socially relevant utterances (Norbury & Sparks, 2013; Tager-Flusberg &Anderson 1991; Adams, 2002; Hale &Tager-Flusberg 2005; Nadig, Lee, Singh, Kyle, Ozonoff, 2010; Sng, Carter, & Stephenson, 2018).

2.1.4   Development and profile of language in ASD

"Language is so readily acquired and so universal in human affairs that it
is easy to forget what a complex phenomenon it is" (Bishop, 2014, p.1).

Language ability is fundamental for successful functioning in society (Bornstein, Hahn, Putnick, & Pearson, 2018; Bornstein, Hahn, & Suwalsky, 2013; Hoff, 2008). Language has been shown to be one of the most significant prognostic factors or predictors of long-term outcomes in social development, adaptive behaviour and educational achievement in individuals with ASD (Billstedt et al, 2007; Howlin et al., 2000; Kuhl, Coffet-Corina, Padden, Munson, Esters, & Dawson, 2013; Lord & Paul, 1997; Magiati, Tay, & Howlin, 2014; Mayo, Chlebowski, Fein, & Eigsti, 2013; Rogers & DiLalla, 1990; Stone & Yoder, 2001; Szatmari et al., 2003)**.**

Extreme variability in language ability, including both receptive and expressive aspects, within and across the individuals with ASD is one of the key characteristics of the condition (Eigsti, de Marchena, Schuc, & Kelley, 2011). Nearly 25-30% of individuals with ASD never acquire functional language, while others develop typical level of language. Between these are individuals with language ability below the expected language level according to their chronological and mental age (Loucas et al., 2008; Tager-Flusberg & Kasari, 2013). Notably,

the studies in this thesis have focused on a sub-group of individuals with ASD with developed verbal skills, due to the language domain examined (figurative language), which is considered as a higher-order or advanced language ability. Individuals in this sub-group are sometimes referred to as 'high-functioning', typically defined as those with full-scale IQ scores >70 or 78 (or in the typical range; Eigsti & Schuh, 2016; Minshew, Goldstein, Taylor, & Siegel, 1994).

Many children with ASD who develop verbal ability achieve language milestones later than children with TD (Howlin, 2003). For example, first words in ASD are often produced at an average age of 36-38 months, rather than the typical 12-18 months (Tager-Flusberg et al., 2009; Zubrick, Taylor, Rice., & Slegers, 2007). Language delay or regression (loss of a skill) in language development is one of the frequent reasons parents seek professional help (De Giacomo & Fombonne, 1998; Short & Schopler, 1988). However, some children with so-called 'high-functioning' ASD do not usually show any significant delays in the onset of language milestones (Kim et al., 2014). Trajectories of language acquisition in children with ASD is typical but delayed in some children, and deviant in others (Eigsti, Bennetto, & Dadlani, 2007; Tek, Mesite, Fein, & Naigles, 2014).

Language, including pragmatic domain, difficulties have been explained in the context of main cognitive theories of ASD. For instance, as a consequence of poor social cognition (Theory of Mind, TOM; referring to the ability to understand mental states of self and others (Baron-Cohen, 1985; Baron-Cohen, 2000). Alternatively, as a reduced drive for coherence (weak central coherence; Happé & Frith, 2006) or limitations in cognitive flexibility (executive dysfunction; see Hill, 2004 for a review). However, evidence suggests that these accounts are not sufficient to explain variation in language in ASD (Norbury, 2014). Therefore, focusing on these theories when explaining a specific pragmatic difficulty without examining connections between the specific pragmatic difficulty and other language domains and interventions based on these studies will lead to interventions failing to remediate language impairment in ASD.

More recently, distinguishing between the neurocognitive subgroups or phenotypes identified based on the structural aspects of language ability in verbal individuals with ASD has been suggested (Kjelgaard & Tager-Flusberg, 2001; Tager-Flusberg & Joseph 2003). Two language phenotypes are defined as autism language normal (ALN) referring to those with ASD and normal language abilities (phonological skills, vocabulary, syntax and morphology), and autism language impaired (ALI), including those with ASD and impairments in receptive

18

language, expressive language, grammatical knowledge and phonological processing. Overlap between impairments exhibited by ALI individuals to impairments shown by individuals without ASD and with language impairment has been suggested. However, this suggestion remains contentious (Williams et al., 2008; Tomblin, 2011). This approach has not been incorporated in this thesis. Rather, the empirical study is based on the strengths and deficits in language ability across individuals.

2.1.5   Language form and content versus pragmatics in ASD

Language is commonly divided in content (vocabulary and semantics: the meanings of words and how meanings are connected); form (syntax: the rules governing how words are combined to make sentences; morphology: the rule system that governs the structure of words and the construction of word forms from the basic elements of meaning; phonology: the study and use of individual sound units in a language and the rules by which they are combined and recombined to create larger language units), and use (pragmatics: rules that govern how to use words in numerous of contexts) (Bloom & Lahey, 1978). For this thesis vocabulary, semantics, syntax and pragmatics have been important as language domains reported to be important for figurative language comprehension. Therefore, I will not focus on the phonology and morphology.

Terminologies used to refer to the domains of language form and content also vary in the literature. In parts of this thesis (e.g., paper II) terms 'language form and content' were used consistently (Bloom & Lahey, 1978), conceptualized as receptive syntax, (receptive and expressive) vocabulary and abstract semantical reasoning. However, in paper I I used the term 'core language'. Using more than one term is common in the literature in the field where the terms such as core language, general language, structural language, basic language, and formal language are commonly used synonymously.

Pragmatic is defined in several different ways in the literature. Pragmatics in this thesis is conceptualized as an essential linguistic component of social communication (Baird & Norbury, 2015; Matthews et al., 2018). Pragmatics refers to the effective use and understanding of language as a tool for communication, specifically how language is used in the context of social interactions (Eigsti et al., 2011). Pragmatic language includes for example inferencing, jokes and figurative language (Baird & Norbury, 2015; Norbury, 2014).

Although this proposed division of language may seem somewhat artificial because these language domains are highly interactive (Norbury, 2014; see the recent review by Matthews, Binney, and Abbot-Smith, 2018 for a thorough discussion), dividing them is a common practice in clinical research. I have also incorporated this approach in this thesis as I wanted to examine the interface between figurative and structural aspects of language with the hypothesis that they are closely related. Better understanding of the connections between pragmatic difficulties (i.e., figurative language) and other language domains will provide clinicians with insights on how to support the communicative needs of individuals with ASD. Unfortunately, fewer studies have examined the relations between pragmatics and other language domains in ASD. This weakness is exacerbated with the limited number of studies within the domains other than pragmatics in ASD.

Interestingly, as has been observed by Brynskov, Krøjgaard, & Eigsti (2016), one of the main tendencies of language research in ASD has been based on the assumptions that pragmatic difficulties in ASD are categorical and absolute, implying that pragmatic language abilities are more or less absent in ASD. On the other hand, other areas of language (i.e., form and content) have been considered more or less intact. Not surprisingly, focusing mostly on pragmatics in ASD and language research in the 1980s and 1990s happened at the cost of focus on other language aspects (Tager-Flusberg, 2000) and represents a serious bias. This is unfortunate because individuals with ASD do show impairments or vulnerability within all aspects of language (Brynskov et al., 2016; Eigsti et al., 2007).

Consequently, research in structural aspects of language is scarce, and findings are conflicting. For example, some studies have revealed intact lexical skills, where vocabulary shows similar patterns of development with children with TD (Swensen, Kelley, Fein, & Naigles, 2007; Tager-Flusberg et al., 1990). However, it has been reported that young children with ASD may not rely on similar lexical learning mechanisms as children with TD, although they can acquire a large vocabulary (Gastgeb, Strauss, & Minshew, 2006; Kelley, Paul, Naigles, & Fein, 2006). For instance, both young and older children with ASD have exhibited underrepresentation of certain classes of words such as mental states and social-emotional terms (Eskes, Bryson, & McCormick, 1990; Tager-Flusberg, 1992), suggesting that vocabulary used by children with ASD can be deviant. Research also shows that individuals with ASD produce less prototypical words (e.g., saying names of unusual animals; Duun, Gomes, & Sebastian, 1996) and show less priming effect of semantically-related words (see a

review by Eigsti et al. 2011; Kamio et al., 2007). Individuals with ASD have also failed to show typical word learning mechanisms such as shape bias, which usually emerges when children map the referent of a noun onto the shape of an object (Tek et al. 2008). All these findings might indicate a different semantic processing in individuals with ASD.

Similarly, the few existing studies on syntactic ability in ASD suggest subtle impairments in syntax (Brynskov et al., 2016). Possible difficulties in syntax have been suggested even in the context of intact lexical knowledge (Eigsti, 2011). However, other studies have demonstrated that syntax might be intact in ASD (Naigles, Kelty, Jaffery, & Fein, 2011; Tager-Flusberg, 1990).

Although pragmatics has been identified as consistently impaired and a pervasive feature of ASD (Eigsti et al., 2011; Lord & Paul, 1997), studies increasingly find that individuals with ASD do not differ from individuals with TD on selected measures of pragmatic language (Norbury, 2005; Norbury 2004; Pijnacker, Hagoort, Buitelaar, Teunisse, & Geurts, 2009; Schaeken, Van Haeren, & Bambini, 2018). Evidently, language form and content reliably has been more recently suggested to predict performance on pragmatic language tasks (Brock, Norbury, Einav, & Nation, 2008; Pijnacker et al., 2009; Volden, Coolican, Garon, White, & Bryson, 2009), proposing that pragmatic abilities are not always impaired but vary across individuals with ASD. Although language form and content (i.e., grammar, vocabulary and semantics) are important for pragmatic functioning (Matthews et al., 2018) in TD, due to the limited number of studies in ASD the relationships between pragmatics and form and content domains of language in ASD is yet to be elucidated.

The tendency to focus on pragmatics while almost disregarding domains of language form and content in ASD research has also influenced figurative language research in ASD. Only recently studies have started to examine language form and content as possible explanation of the difficulties these individuals experience with figurative language. One of the goals of my PhD-thesis has been to test the extent to which this assumption of a dissociation between pragmatics and language form and content holds when it comes to figurative language. This brings us to the main focus of the thesis, that is, figurative language comprehension in general and specifically in ASD.

2.2   Figurative language

"Promise me you'll look after yourself ... stay out of trouble ...'
'I always do, Mrs Weasley,' said Harry. 'I like a quiet life, you know me."

Harry Potter and the Half-Blood Prince (J.K.Rowling, p.327).

Harry Potter here tries to be funny and uses sarcasm for this purpose. Sarcasm is one type or trope of figurative language, a subtype of irony (e.g. 'What nice weather for a football game' – to describe a rainy day with extreme weather). Other common tropes of figurative language are for example *metaphors* (e.g. "Sally is a chameleon"), and idioms (e.g., "It is raining cats and dogs").

The different tropes of figurative language vary in their communicative function and comprehension demands (Colston & Gibbs, 2002; Nippold, 2016; see Pexman, 2008 for a review of usage of verbal irony for a wide range of social purposes). For instance, sarcasm involves a positively worded statement, which is meant to be taken negatively (Nicholson et al., 2013), whereas metaphors are powerful tools used to achieve educational (Katz, 2017) or political goals (for the broader discussion about a metaphor see subsection 2.2.2.below).

Figurative language is used to convey thoughts, ideas and emotions in ways that distinguish it from literal language. Imagine a situation in a classroom where a teacher tells one student: *'You have done a really great job!'*, when in fact the student handed in almost a blank page instead of the five-page assignment that was expected to be delivered. Understood literally, teacher's utterance makes little or no sense in this context, rather it must be interpreted figuratively because the teacher has been sarcastic and truly meant the opposite of what she/he explicitly uttered. Right after that, the teacher might say to another student: 'Such wonderful thoughts in the text! *You are a breath of fresh air'!"*. In this different example of figurative language usage, that is a metaphor, the teacher intend to enhance the admiration of the brilliant pupil, through using a metaphor.

Although the different tropes vary both in terms of comprehension demands and communicative functions, these tropes have the common shared trait: the encoded words do not overlap with the intended meanings of the words and utterances (Gluksberg, 2001). This

implies that while speaking figuratively, speakers mean something other than what they literally say (Gibbs & Colston, 2012). Therefore, to understand figurative language, an individual must be able to grasp the speaker's intention by inferring implicit meaning relevant for a given context (Rapp & Wild, 2011) based on the encoded words uttered.

2.2.1   Figurative language pervades everyday communication

Theoretical accounts of different types of figurative language are sometimes mutually exclusive, however they all agree that figurative language is a ubiquitous part of everyday spoken and written language and communication, and essential for successful social interaction (Nippold, 2016). All types of figurative language often occur in multiple contexts such as conversations, lectures, media communication, literally fiction, textbooks, newspapers and magazines (Nippold, 2016; Wilson, 2011). Figurative language pervades also the language directed to children: for instance, approximately 6 % of the sentences of the reading programs for the primary schools in the United States contain an idiomatic expression (Nippold, 1991). Furthermore, figurative language is common in the computer-mediated communicative settings such as e-mails and blogs. In their study, Whalen, Pexman, & Gill (2009) found that in about 94% of all e-mails written by young adults there was at least one nonliteral statement. Similarly, some form of verbal irony was used in around 73% of all regular personal blog entries (Whalen, Pexman, Gill, & Nowson, 2013). Not surprisingly, competence to comprehend and produce 'all types of figurative language is an important part of becoming a culturally literate and linguistically facile individual' (Nippold, 2016, p.16).

2.2.2   Metaphor

This subsection is about metaphor that is a paradigmatic type of figurative language and has been of special interest in this thesis. In the literature metaphor has been defined in many different and sometimes incongruous ways. This thesis is based on the understanding of a metaphor as a linguistic pragmatic phenomenon, evolving as a result of linguistic communication (Sperber & Wilson, 2008; Wilson, 2011), as speakers use words to achieve particular effects on listeners and readers (Carston, 2017). Metaphor is widely considered an essential tool for explaining and understanding complex topics. This can be put to different uses, for example, metaphors are central in explaining scientific discoveries, whereas in

commercial advertising metaphors are considered a great persuasive tool (Katz, 2017). Metaphors often appear in literary fiction to express complex or abstract ideas or are used to persuade readers to experience a particular relationship among apparently unrelated ideas (Katz, 2017).

As a pragmatic phenomenon, metaphors involve a discrepancy between the encoded, 'literal' or conventional meaning of words and their context/occasion-specific use (Camp, 2009; Carston, 2010). In metaphor, two entities (the topic and the vehicle) from different conceptual domains or categories are linked based on some shared set of semantic properties or similarities or *common ground* (Van Herwegen & Rundblad, 2018). The comprehension of a metaphor requires a pragmatic inferential process (Wilson & Sperber, 2012) that draws upon the supportive context (Bambini, Bertini, Schaeken, Stella & Di Russo, 2016; Gernsbacher, Kaysar, Robertson & Werner, 2001) and lexical knowledge (Nippold, 2016).

It is noteworthy that metaphors exist in different structures, and accordingly, comprehension in the different structures is likely to differ. In addition, other variables such as context, experimental design and experimental setting also contribute to performance on metaphor tasks (Colston & Gibbs, 2002). These considerations bring us to the next section, which is about figurative and metaphorical language comprehension in ASD.

2.2.3 Comprehension of figurative language with emphasis on metaphors in ASD compared to TD

> *"A critical point in the comprehension process is reached when the listener understands not just what the speaker said, but what was meant"* (Bishop, 2014, p. 253).

Grasping what is meant by an utterance, which in figurative language does not coincide with what is explicitly said or stated, is key to comprehending figurative language. In addition, it is hard to imagine how figurative language can be comprehended without understanding the encoded words. In individuals with TD figurative language comprehension starts to develop in early childhood, alongside linguistic development (Pouscolous, 2011; 2014; Vosniadou, Ortony, Reynolds, & Wilson, 1984) and gradually improves throughout the school years, adolescence and even adulthood (Nippold & Duthie, 2003; Pexman & Glenwright, 2007). Comprehension is thought to become adult-like at around 10 years (Winner, 1988). Figurative

language tropes differ in terms of their time of acquisition too. For example, while some degree of metaphor competence has been reported in children aged 3-4 years if tested with age-appropriate tasks (Vosniadou et al., 1984), irony comprehension seems to develop later, between 6 and 10 years of age (Filippova & Astington, 2008; Pexman & Glenwright, 2007). Similarly, comprehension demands vary across tropes. For instance, comprehending of a metaphor "Sally is a chameleon" requires possession of some knowledge of the topic and the vehicle in order to grasp the similarity between these two domains. Irony comprehension has a different set of requirements: the recipient needs to draw a second-order inference about the speaker's beliefs and intentions (Andrés-Roqueta & Katsos, 2017; Colston & Gibbs 2002), understand the attitude and emotion of the speaker and also the impact they intend their statement to have (Nicholson, Whalen, and Pexman, 2013).

Even linguistically able individuals with ASD, regardless of their age (Ozonoff & Miller, 1996), have shown difficulties performing on different types of figurative language tasks compared to individuals with TD (e.g. Happé, 1993; Kaland et al., 2002; MacKay and Shaw; 2004; Ozonoff & Miller; 1996; Rundblad & Annaz, 2010; Vulchaova et al., 2015). These difficulties have traditionally been attributed to the social cognitive profile, especially the impaired ability of ToM (Baron-Cohen 2000; Fletcher & Happé, 2019; Happé, 1993; see however a study by Gernsbacher, & Frymiare, 2005, suggesting that ToM ability is not universally impaired in ASD). For example, in one of the pioneering studies on figurative language comprehension in individuals with ASD, Happé (1993) found that those individuals with ASD who failed a first-order ToM task (inferring a person's mental state, e.g. what he/she thinks) performed significantly poorer when the target was a metaphor relative to those who passed the ToM task.

A contradictory conclusion was reached by Norbury (2005) who adopted the task from that of Happé (1993) but modified it through simplifying vocabulary to avoid confounding effects of difficult words on performance. Norbury's study showed that only participants with impaired core language ability, regardless of their diagnostic status (ASD or not ASD) had problems comprehending metaphors. Furthermore, she concluded that ToM might be necessary but not sufficient to understand metaphors; semantic knowledge, on the other hand, was in this study found to be a significant predictor of metaphor comprehension.

Subsequent studies on metaphor comprehension have reported conflicting results with respect to group differences as well as the variables underlying figurative language comprehension (e.g., Kasirer & Mashal, 2016; Mashal & Kasirer, 2011; Norbury, 2005; Rundblad & Annaz,

2010). For example, Rundblad & Annaz (2010) reported statistically significant group difference in metaphor comprehension and no links between receptive vocabulary and metaphor comprehension. However, several caveats with this study warrant mention. First, the groups in this study significantly differed in terms of language ability even it was measured using a receptive vocabulary test, BPVS (Dunn, Dunn, Whetton, & Burley, 1997). Receptive vocabulary has been shown to be a relative strength in ASD and a poor proxy for language comprehension (Eigsti, 2017; Gensbacher & Pripas-Kapit, 2012). In addition, the parameters or properties of the metaphor task (i.e., asking participants an open question) itself may have introduced variation. The latter aspect has been a focus of this thesis and is explained in the next sub-section.

2.2.4   Metaphor comprehension assessment

In the literature on TD and clinical populations there is evidence that metaphor comprehension at least partially depends on the way it is measured (Perlini et al. 2018; Pouscoulous, 2011; 2014; Vosniadou et al., 1984). For example, when measured with a multiple-choice task or of non-verbal enactment task (i.e., listening to a story containing a metaphor and acting its meaning out with toys) has in fact revealed metaphorical comprehension skills in children for example at age three (Vosniadou et al., 1984). This means that multiple-choice task is easier than verbal explanation task used in the above-mentioned study by Rundblad & Annaz (2010) that involved metaphors embedded into picture stories required from the participants to answer open questions. Verbal explanation response format is challenging because it requires a higher degree of meta-linguistic competence (the ability to reflect upon and analyze language as an entity; Nippold, 2016) as well as expressive language skills. Moreover, verbal explanation tasks are considered as more socially demanding as they require more social interaction between the participant and the examiner than multiple-choice tasks do. The latter aspect is an important consideration in ASD due to the social communication and interaction difficulties that are a core feature of ASD. In addition, while using visuals (i.e., pictures) might generally be beneficial for understanding, some individuals with ASD might have problems in switching between modalities (i.e., visually and auditory delivered information; Reed & McCarthy, 2012). Without considering the different properties of the tasks in studies with ASD, the results from the studies cannot be interpreted validly. Despite the tremendous variety of tasks, with a myriad of properties, that have been used in earlier studies in ASD, the potential influence of

these properties on performance have not been a focus in ASD and figurative language research perhaps reflecting the established view that the origin of the difficulties was to be found in the diagnosis. 'It's become a caricature of autistic persons that they don't understand figurative language' (Gernsbacher & Pripas-Kapit, 2012, p.93), but is this correct?

2.2.5 Difficulties in figurative language: specific and unique to ASD?

The question that has increasingly been asked in recent literature are whether figurative language difficulties are universal for all individuals with ASD, as it traditionally was believed, or may impaired performance on tasks be related to factors that are known to cause difficulties in figurative language in TD.

More recent studies that sought to answer these questions have challenged the traditional view of universally impaired figurative language in ASD, as several groups of individuals with ASD did not perform significantly worse than controls with TD on figurative language tasks (Hermann et al., 2013; Whyte, Nelson, & Scherf, 2014). In particular, these recent findings indicate that a) figurative language are not impaired in all individuals on the spectrum; and, therefore, b) variables other than diagnostic status may explain the extent of group differences in studies comparing individuals with ASD and individuals with TD. Thus, the view that difficulties in figurative language is specific and unique to ASD has been steadily challenged (see Gernsbacher & Pripas-Kapit, 2012 for discussion).

2.2.6 Do language form and content and group matching strategies matter?

While the impact of metaphor comprehension task properties has not been investigated in ASD, several studies have correctly indicated the link between language form and content and group-matching strategies, and figurative language ability (e.g., Norbury, 2004; 2005; Whyte et al., 2014). However, there is still dearth of studies carefully examining these relationships. Furthermore, studies have yielded conflicting results, some finding this link, other failing to do so (see Gernsbacher & Pripas-Kapit, 2012 for an overview). Conflicting findings might be related to the ways language form and content have been assessed across the studies (e.g.

measuring of receptive vocabulary only). In addition, it might be that language form and content are differently related to different types of figurative language. For example, the already mentioned study by Rundblad & Annaz (2010) found that receptive vocabulary (as measured by BPVS) could reliably predict comprehension of metonymy which is another type of figurative language ("*Ibsen* is difficult to read"), but not metaphor. Therefore, more studies carefully examining participants' language abilities in comprehending different figurative language tropes are needed to advance understanding. A better understanding can eventually lead to identifying possible targets to improve social functioning through knowledge of words as building blocks for pragmatic and social communicative development.

3 Methodological considerations

Motivated by the need for better understanding of figurative language in ASD, this thesis aimed to inform future research, practice, and theory, paving the way for better-designed research and facilitating educational programs tailored for individuals with ASD. This aim could be achieved only through methodologically rigorous studies and, therefore, several considerations were made prior to and throughout the research process. The choices of methods in these studies were driven by the studies' aims and research questions. Systematic review and meta-analysis (Paper II and II) and cross-sectional design (Paper II) were used.

3.1   Why systematic review and meta-analysis?

> *"Now that we know how to summarize literature meta-analytically, it seems hardly justified to review a quantitative literature in the pre-meta-analytic, prequantitative manner"* (Rosenthal, 1994, p.131).

Systematic reviews can provide a basis for planning a new primary research by identifying gaps in the literature (Borenstein et al., 2009; Gough, Oliver, & Thomas, 2017). Therefore, primary studies should be based on a comprehensive knowledge of the existing research that should determine what kind of studies are needed to prevent unnecessary and/or low-quality studies (Borenstein et al., 2009; Gough et al., 2012). Logically, if this knowledge is unavailable, it would benefit the field of research for a review to be carried out.

An influential and excellent narrative review that to a great degree shaped my understanding of figurative language in ASD, authored by Gernsbacher and Pripas-Kapit (2012), has been a valuable contribution to the field and has had a great impact on subsequent research. However, being a narrative review, the study did not provide a systematic overview of earlier research and lacked the descriptions of the searching and screening process. Moreover, as this review has been influential, analysing the studies conducted after its publication would provide important knowledge about any potential changes in the ways figurative language has recently been studied in ASD.

No systematic reviews or meta-analyses had been conducted on figurative language in ASD previously. Because the aim of this thesis was to generate a comprehensive and trustworthy picture of figurative language in ASD, and as this is better done through systematic reviews

than through individual primary studies or narrative reviews, the decision to conduct the first study (paper I) was made.

The reason for undertaking another systematic review and meta-analysis (Paper III) was the need to consider how metaphor comprehension has been assessed in previous literature. My motive was to scrutinize the metaphor tasks to see if the different task properties could be related to differences in effect sizes across studies. Hypothetically, the demands on the performance across the tasks with different properties (i.e., multiple-choice vs. verbal explanation response format) would differ.

The reason for conducting a meta-analysis in both studies was to exploit the method's main advantages: statistically aggregating study findings, provide effect sizes of group differences, and increase power (Borenstein et al., 2009; Kavale, 1984). Increasing the power and generalizability is important for the field of figurative language in ASD as studies often include small samples that can be a threat to statistical power, leading to a rejection of the null hypothesis on an erroneous basis (Næss, Lyster, Hulme, & Melby-Lervåg, 2011).

3.2   Why a cross-sectional study?

The meta-analytic study (1) revealed several methodological flaws in previous research (see the Discussion part in Paper I for methodological discussions and implications).These methodological inconsistencies have restricted an understanding of which variables are related to figurative language, which consequently has impeded the selection of appropriate targets for intervention. For example, planning a randomized controlled trial without first observing which variables are related to the outcome is not appropriate (and is even unethical) and more carefully designed observational studies were required (Gough et al., 2017). Sometimes researchers should step back and examine the variables that have not been examined properly in observational studies to move the field forward. Therefore, the role of observational studies and studies with small samples, which are common practice in research with clinical populations, should not be underestimated (Etz & Arroyo, 2015; Smith & Little, 2018). Through using carefully designed metaphor tasks and measuring language form and content with multiple measures, the empirical study (Paper 2) contributes to a better understanding of metaphor comprehension in ASD and will facilitate future reviews and inform intervention studies.

3.3   Validity

The decisions made regarding the study design and method influenced the validity of the claims that were made based on the results. As a property of an inference, validity refers to the approximate truth or correctness of an inference (Shadish et al., 2002). In the following section, the considerations in terms of three types of validity (construct validity, external validity, statistical conclusion validity) that were relevant for these studies are discussed.

3.3.1   Construct validity

Construct validity depends on the score reliability, that is how accurately and precisely scores are measured (Thorndike & Thorndike-Christ, 2010). Because of the intimate relationships between the reliability of the test scores and the construct validity, they are discussed in the same sub-section here. Figurative language is a complex and multifaceted construct, a variable that is not directly observable. Making a proper conceptualization of this construct for a research study is therefore critical (Kline, 2009).

Prior conducting the systematic review and meta-analyses in Paper I and Paper III, together with the co-authors, we first established the understanding of the constructs we aimed to examine (different figurative language tropes in paper I and metaphor in paper III). In study III we specifically focused on how the studies assessed metaphors, therefore measurement issues were central for this paper.

For the empirical study (paper II) a new measure of metaphor comprehension was created since no validated or standardized Norwegian metaphor test existed for this age-group. A detailed description of the test development procedure is provided in Appendix 1 in paper II.

To develop a reliable measure and ensure the validity, task development process was discussed with several experts in the field of linguistics, psycholinguistics, pragmatics and psychology. Conducting the first meta-analysis (paper I) helped us to identify the weaknesses with the measures in earlier studies which we tried to avoid in our stimulus. Another critical point considered for the empirical study was that reliability of a test might be related to the differences in ability among the participants. Therefore, we attempted to create a task that would avoid ceiling or floor effects, but at the same time should show the variability in the

scores. This is an important consideration because if the test is very hard for some individuals, accuracy is likely to be at chance level (Devlin, 2015).

Test reliability and accordingly validity also depend on the length of the task as with increased length the scores depend increasingly on the ability of the person being measured and less on chance, providing a more accurate assessment (Devlin, 2015). However, there are practical factors that need to be considered such as time, fatigue and boredom. Pilot testing of the stimulus material was carried out to examine the length as well as the linguistic content of the items. Pilot-testing revealed no specific problem with the length. As to the content, only some of the items were slightly revised according to the feedback from the participants in the pilot-testing. For example, we changed the context for a few items as the contexts could have priming effect (would be possible to infer meaning from the context, without any need to comprehend metaphor).

To avoid the potential impact of the syntactic structure of the metaphor, only the items with the same syntactic structure ($X=Y$) were included. Also, since we did not test reading comprehension of the participants, the task was read aloud to them to avoid confounding effect of possible reading difficulties. However, reading aloud could also pose extra demands because of the known challenges with processing information from spoken language in ASD (see O'Connor, 2012 for a review).

One important aspect when selecting the properties for the metaphor task was to decide the response format, because from research in TD we know that choice of response formats may affect performance (Pouscoulous, 2011; 2014). There is no agreement on what kind of metaphor tasks are best to use in ASD research, and the decision about the response format has, therefore, to be determined by the research questions and the participants' characteristics. However, the extreme heterogeneity within the ASD group made to take this decision challenging. Finally, we decided to create a metaphor task with two sequential response formats: first multiple-choice response, then verbal explanation to justify the choices the participants made. The reason we also wanted the participants to justify the choices was that multiple-choice tasks may be susceptible for measurement error because of the possibility that participants would be guessing the answers (Kline, 2009), that would negatively impact the score reliability. Although verbal explanation/justification format is considered to be more sensitive measure, we wanted to avoid verbal explanation/justification format alone because of known extra social and expressive language demands they pose (Kwok et al., 2015;

32

McGregor et al., 2012; Lewis, Murdoch, & Woodyatt, 2007). However, we faced the following challenge: asking participants why they thought the alternative they choose was the best fit was found to be an ineffective way of obtaining an explanation because most participants simply repeated the answer they had selected from the multiple-choice. This made it impossible to determine whether the participants repeated answers with or without a real understanding of the items. Therefore, the answers were not included in the analysis.

3.3.2   'Take home messages' from the data collection process

Data collection process gave me several ideas on the issues of figurative language comprehension measurement in future studies. Importantly, during the testing sessions I received some important comments from the participants from both groups about the task properties that I will consider in future studies. The data collection process has shown me the importance of individuals with ASD participating in research that is about them and their lives, and how much researchers without ASD can learn through collaboration with these people (see Fletcher-Watson et al., 2018 for discussion about participatory research and its benefits). In general, the data collection process has given me invaluable experience with and knowledge about testing individuals with and without ASD that will facilitate our future studies.

3.3.3   External validity

External validity refers to the extent to which conclusions or inferences drawn from the study can be generalized across multiple cases or individuals with similar characteristics, settings, or measures (Kline, 2009; Orlikoff, Schiavetti & Metz, 2015; Shadish et al., 2002).

Paper I and Paper III both present research synthesis and meta-analyses and can therefore be considered more externally valid than single primary studies (Borenstein et al., 2009; Gough et al., 2009). This is because the core of research synthesis is the attempt to integrate empirical research for the purpose of creating generalizations (Cooper at al., 2009). Similarly, due to the large and multiple samples and multiple conditions and settings, meta-analyses provide estimates that are more representative of the whole population than those provided from single studies (Borenstein et al., 2009).

It is noteworthy that one of the most significant threats to meta-analysis is related to publication bias, referring to the tendency of publishing studies with significant findings which makes finding and including them in the meta-analysis easier (Borenstein et al., 2009; Shadish et al., 2002). Consequently, if the included studies are biased sample of all relevant studies, then the mean effect computed by the meta-analysis will reflect this bias (Borenstein et al., 2009). Several steps addressing the publication bias are described in the Papers I and II. Instead of iterating these steps, here I will briefly mention that in addition to searching for the 'grey literature' (studies that are not published in peer-reviewed academic journals), the methods of funnel plot and 'Trim-and Fill' were used to examine the publication bias.

In Study 2, several aspects were related to the external validity. First is participant recruitment as a main threat for the generalizing to other people concerns the degree to which the subjects selected for the study are representative of the population to which the researcher wishes to generalize (Orlikoff et al., 2015). Therefore, participants for this study were recruited throughout Norway to ensure recruitment of as many participants as possible to increase the statistical power. I ended with recruiting 29 individuals with ASD and 31 individuals with TD. Although the sample size is not very large, it is common in research with specific groups such as ASD and can be considered appropriate for this study type (Smith & Little, 2018). However, the overrepresentation of boys in the sample may have reduced representativeness and the generalizability of the results to the entire ASD population. Furthermore, the findings cannot be generalized on 'low-functioning' individuals with ASD. For example, we could not include one participant with ASD because the tests were difficult for the participant. Moreover, we should be careful to generalize the findings on other tropes of figurative language, such as for example idioms and irony because different tropes are understood differently, as discussed in relation to the construct validity earlier (paper I).

3.3.4   Statistical conclusion validity

Statistical conclusion validity refers to the appropriate use of statistical methods to estimate relations between variables we are interested in (Kline 2009). In Paper I and II, in case of overlapping sample, the study with a smaller sample was excluded. Furthermore, when studies reported multiple outcomes from the same sample, such as for example novel and conventional metaphors, the mean of the outcome for each study was computed and used this synthetic score as the unit of analysis (Borenstein et al., 2009). In this way, every study was

presented by one score in the meta-analysis and this approach solved the problem of more weight assigned to studies with more outcomes (Borenstein et al., 2009). The statistical analyses for Study 2 were selected and performed in close collaboration with a psychometrician at the Centre for Educational Measurement (CEMO), Faculty of Educational Sciences, University of Oslo.

3.4   Ethical considerations

3.4.1   Selection of study approach and treatment of participants

A high-quality study means better investment of time of the participants, and a good investment in science by society. Therefore, ethics and scientific quality are interrelated (Rosenthal, 1994).

In terms of ethical considerations related to this thesis I first want to mention the advantages of including two meta-analyses. In particular, meta-analyses increase the utility and benefits of individual studies, therefore the time and effort human participants contribute to individual studies is more justified when their data are included in meta-analyses (Rosenthal, 1994). Therefore, it is a responsibility of the researcher to ensure that necessary details about the methods and the results of an empirical study are reported in order to allow its inclusion in potential future reviews and meta-analyses.

The empirical study (Paper II) was conducted in accordance with the guidelines from the National Committee for Research Ethics in the Social Sciences and Humanities (NESH, 2006). Prior to the data collection, the ethical approval from the Norwegian Centre for Research Data (NSD) was obtained. As this study included young participants, their parents or a guardian gave the informed consent. The information letter included details about the purpose of the study, the testing procedure and the handling of the data. Because of the diagnostic status, the participants with ASD were defined as belonging to a vulnerable group, and they might not be as able to protect their own interests as participants with TD (NESH, 2006). Therefore, before testing sessions, all participants were informed that they could withdraw at any point without having to justify their decision. To make the participants feel

safe, parents, teachers or special teachers or another person they felt comfortable with were invited to attend the sessions. Locations for testing were suggested by the participants and parents. In addition to the written informed consent, oral consent was obtained from each participant before testing. During testing, care was taken to ensure that the participants felt comfortable and relaxed. The session started with explaining what was going to happen and what was expected from them. To avoid fatigue, the participants were given the time they needed to perform the tests and there was always the possibility for breaks whenever needed.

In a research study, keeping all personal information confidential is a requirement (NESH, 2006). Therefore, the signed consent forms and other information that could identify the participants were kept separate from the data-sets in locked filing cabinets. The data were anonymized by assigning each participant an ID-number. No information that could lead to identification of participants were included in any presentation of the data. The recordings, taken only for the metaphor task session, were transferred to the laptop that was not connected to the internet and was locked in a filing cabinet. They were deleted from the laptop by the time the project was finalized.

3.4.2 Selection of terms to refer to participants

Selection of terms used to refer to clinical populations is often a sensitive topic in research and ASD is a good example of how scientific classifications of people affect the people classified and how that effects on the people (Hacking, 1999; 2006). The language we use might define the people differently as they would define themselves (Hacking, 2006), and 'the language that we use has the power both to reflect and to shape people's perceptions of autism' (Kenny, Hattersley, Molins, Buckley, Povey, & Pellicano, 2016, p.1). ASD and its changing conceptualization is a good example as how differently people are classified. In particular, ASD has been earlier referred to 'childhood scizophrenia', 'infantile autism' and whereas now we know that ASD is neither the same as schizophrenia, nor it is a childhood but rather a life-long condition.

The way we see the people and the terms we use may differ from those preferred by different people with ASD (Fletcher-Watson & Happé, 2019). For example, many individuals with ASD prefer to be referred as 'autistic' (Gernsbacher 2017). Similarly, the term 'autism' is preferred by many individuals with ASD emphasizing that autism is a natural part of the human population (Fletcher-Watson & Happé, 2019) and a part of their identity.

36

Unfortunately, even the strengths these people have are often assumed as difficulties. As Gernsbacher, Dawson, & Mottron (2006) puts it: 'the empirical literature is replete with demonstrations of autistics' superiority in numerous perceptual, reasoning, and comprehension tasks', but 'quite compellingly, each of these statistically significant demonstrations of autistic superiority is labelled by its authors as a harmful dysfunction' (p.2).

I acknowledge and respect the views and preferences of people with ASD. However, I am aware that the terms used in this thesis (e.g., Autism spectrum disorder) will not be a preferable term of all individuals with ASD. In this thesis I adhered to the recommendations provided in the publication manual of the American Psychological Association (APA, 2010.) regarding using person-first language that is currently 'preferred when describing people with disabilities' (p.72). In person-first language, a noun referring to a person or persons precedes a phrase referring to a disability (e.g., individual with ASD). Person-first language contrasts with identity-first language (e.g., disabled person, autistic children). The core principle of person-first language is that persons with disabilities are treated linguistically the same way as persons without disabilities to avoid stigmatization (see Gernsbacher, 2017 for discussion). Therefore, throughout the thesis stigmatizing persons with ASD was avoided by using matched constructions for the two groups: 'individuals with ASD' and 'individuals with TD'. Summary of main findings of the studies. Also, the term 'Autism Spectrum Disorder' according to the DSM-5 recognises the variability of the presentation of the condition and this variation is also recognized in this thesis. Based on the justifications above, in this thesis person-first language and 'Autism Spectrum Disorder' is consistently used.

4 Summary of main results of the studies

The research questions this thesis aimed to answer were:

RQ1: What is the extent of the group difference in figurative language comprehension in ASD compared to individuals with TD?

RQ2: Which variables are related to figurative language comprehension in ASD?

Studies have answered both research questions using systematic review and/or meta-analysis (Study 1 and Study 3) and cross-sectional design (Study 2).

Study 1: Kalandadze, Norbury, Nærland and Næss (2016) addressed both research questions: We aimed to summarize and meta-analyze 41 studies (45 effect sizes) on figurative language comprehension in ASD compared to individuals with TD. In addition, the study examined the group-matching strategies, differences in figurative language tropes, chronological age and cross-linguistic difference as potential moderators of variation between effect sizes. Overall, individuals with ASD, exhibited inferior comprehension of figurative language than their peers with TD (Hedges' $g = -0.57$). Group matching strategies and types of figurative language were significantly related to differences in the effect sizes across studies. Importantly, the differences between the groups were small and nonsignificant when the groups were matched based on the language form and content ability. When comparing studies according to the figurative language tropes, surprisingly, comprehension of metaphors was more difficult for individuals with ASD compared with controls with TD than comprehension of irony and sarcasm, also compared with controls with TD. Chronological age and cross-linguistic differences could not explain the variation in effect sizes. Findings emphasize that group differences vary across figurative language tropes. Also, the crucial role of language form and content in interventions designed to improve social communication skills in individuals with ASD is highlighted.

Study 2: Kalandadze, Braeken, Falkum, Brynskov and Næss (submitted) addressed both research questions. This cross-sectional study aimed to investigate metaphor comprehension

as measured by a multiple-choice task in individuals with ASD (N= 28; aged 10-16 years) compared to peers with TD (N=31). The study also examined whether variation in language form and content could be related to variation in metaphor comprehension. Overall, individuals with ASD scored lower on metaphor comprehension than their TD peers (Δ = -5.63; SE = 2.31; p = .020; Cohen's d =-.66). However, after controlling for language form and content, the difference between the groups was no longer statistically significant (Δ= -2.55; SE = 2.34; p = .281; d = -.32). The results indicate that individuals with ASD performed worse on metaphor comprehension tasks than individuals with TD, but this difference was mainly due to the former group having impoverished ability of language form and content compared to the latter group. Thus, the observed difference in performance is partially related to differences in language form and content ability and not solely due to the group of individuals with ASD being impaired in metaphor comprehension.

Study 3: Kalandadze, Bambini and Næss (revised and resubmitted) addressed both research questions and reports a systematic review and meta-analysis of 14 studies and aimed to address the differences between groups of individuals with ASD and with TD in metaphor comprehension, as well the potential link between metaphor comprehension and task properties. Overall, groups of individuals with ASD showed poorer performance on metaphor comprehension than groups of individuals with TD (Hedge's g = -0.63). Although the pattern of tasks using verbal explanation response format posing most challenge for individuals with ASD has been observed, based on this study alone the role of task properties in variation between the studies' results cannot be elucidated. Therefore, it remains for future research to investigate the role of task properties in metaphor comprehension more carefully.

5 Discussion of the results

5.1 Is impaired figurative language comprehension universal in ASD?

Overall, the results of studies in this thesis indicate that individuals with ASD as a group show inferior performance on figurative language comprehension tasks than individuals with TD. However, the apparently impoverished figurative language comprehension is to a certain degree related to variables such as language form and content (lexical, semantic and syntactic knowledge), group-matching strategies (whether groups are matched for language ability, chronological age or both), differences in figurative language tropes (e.g., metaphor vs. irony) and likely to the figurative language task format.

5.1.1 Group matching strategies and language form and content

Language form and content (paper 1 and 3) and group-matching strategies (paper 1) were significantly related to the performance of individuals with ASD on figurative language tasks. In particular, when groups were matched based on language form and content ability, group difference was not statistically significant indicating that individuals with ASD and TD with comparable language skill also have a similar ability to perform on figurative language tasks. This finding is in line with earlier primary studies (e.g., Norbury, 2004; 2005) and resonates with the review by Gernsbacher and Pripas-Kapit, (2012). Accordingly, a significant group difference in language form and ability is associated with the significant groups differences on figurative language comprehension reported in individual studies (e.g. Rundblad & Annaz, 2010).

Regrettably, several earlier studies either did not control for language ability or failed to operationalize and assess language ability broadly (e.g., Happe, 1993; Gold & Faust, 2010). This aspect is critical because assessing only receptive vocabulary or even verbal intelligence (IQ) can obscure language impairment in individuals with ASD (Gernsbacher & Pripas-Kapit, 2012). Not surprisingly therefore, in our empirical study (Paper II), groups differed

significantly on metaphor comprehension at the baseline, but the difficulties individuals with ASD showed in metaphor comprehension could be ascribed to their impoverished receptive and expressive vocabulary, abstract semantic reasoning, and receptive syntax.

Thus, our findings resonate with the claims made by Gernsbacher & Pripas-Kapit (2012): 'If autistic persons have difficulty comprehending language, they'll also have difficulty comprehending metaphoric, idiomatic, inferential, potentially ambiguous, or otherwise complex language. And the same goes for non-autistic persons. More difficulty comprehending figures of language that are more difficult to comprehend is neither surprising nor unique to autism' (p.4). What about cognitive characteristics, such as impaired ToM ability and executive dysfunction (defined at p. 19 in this thesis) that have previously been suggested to account for difficulties in figurative language in ASD (Happe, 1993; Kasirer & Mashal, 2014; Mashal & Kasirer, 2011). Paper I contains information about the relationships between ToM and figurative language and language in general. Here I would like to add that the relationship between ToM and pragmatics is unclear. In particular, it is unknown whether ToM ability is always involved in pragmatic reasoning and may not be required for some kinds of figurative language comprehension such as metaphors (Andres-Roqueta & Katsos, 2017; Tendahl & Gibbs, 2008). This is because some sentences might be interpreted without the need to infer the speaker's mental state, and language form and content (vocabulary and grammar) might be sufficient to perform the task successfully (Andres-Roqueta & Katsos, 2017; see also Norbury, 2014 for the similar suggestion). As to the executive functions, examining this variable alongside language ability may provide valuable information.

5.1.2 What about differences in tropes and Task properties

Another important finding of this thesis (paper 1) is that differences in tropes are important to explain varying effect sizes between studies. This finding converges with previous results also suggesting that various types of figurative language are understood differently (Colston & Gibbs, 2002; Happe, 1993). However, in contrast to our findings, more difficulties between groups on irony than on metaphor studies have been suggested (Happe, 1993). Our unexpected finding can once again be explained by methodological factors, such as differences in ways metaphor and irony comprehension have been assessed across the studies. Hypothetically, metaphor studies could have used less carefully designed assessment tools

than irony studies. To assess this, study 3 focused on the metaphor comprehension task properties and their potential relationship to performance. Indeed, we found that metaphor comprehension tasks used in research in ASD differ tremendously in terms of their properties, but the potential influence of these properties are rarely considered in the individual studies. As we have not conducted any similar review on studies involving irony in ASD yet, we cannot make firm conclusions. Such a review would benefit the field.

The study with the largest effect size among the included studies used a verbal task (Rundblad & Annaz, 2010) indicating that verbal explanation tasks might be most challenging for individuals with ASD, especially for the participants with significantly low language ability. In our empirical study we used a multiple-choice task which could potentially could have facilitated performance since this task format is generally easier than verbal explanation tasks.

5.2   Limitations

Notwithstanding the strengths of these studies, the findings must be viewed in the light of the limitations that are discussed in the papers. One consideration related to the Paper II is that, due to the absence of a validated Norwegian metaphor comprehension task, a self-created task was used in the empirical study. Using a validated task would have increased the measurement of reliability and validity. Validating a task is time consuming and was not achievable within the time frame of this PhD-thesis. The detailed description of the task development procedure added as an appendix (1) in the paper will help readers to consider how the properties of the metaphor tasks could have contributed to the results.
One limitation related to the empirical study (paper 2) is the overrepresentation of boys in the sample. However, this is reflective of the current prevalence of ASD in boys and girls which is generally estimated at around 4:1 (Fombonne E 2005). Therefore, our sample can be considered representative of the ASD population as it is now.

Limitations related to Study 1 and Study 3 are thoroughly discussed in the papers (Paper I and Paper II).

5.3   Implications

5.3.1 Implications for research

The main implication of this thesis for future research is that studies should look beyond the diagnostic status and try to identify other variables, also those independent from the participants' characteristics or abilities (e.g., methodologies including assessment strategies) when studying figurative language comprehension in ASD. Language form and content and task properties are the main aspects that play a critical role in figurative language comprehension and it is important that future studies consequently measure language form and content with multiple measures to make it possible to compare the studies and draw firmer conclusions. In general, the detailed description of the methods including the task development procedure should be included in the papers where study-specific measures are involved. Longitudinal and intervention studies are needed.

5.3.2 Implications for practice

Given that individuals with ASD often have problems with figurative language, these difficulties should be a target for educational programs and interventions. Our findings suggest that individuals with ASD, due to different reasons, have problems in comprehending figurative language. Therefore, instead of avoiding using figurative language with these individuals, teachers, parents, clinicians, speech and language therapists should teach comprehension strategies. Avoiding to use figurative language will deprive these children from learning lexicalized figurative expressions (that are established in the language and are learned and stored in the mental lexicon as lexical units) or strategies to comprehend novel expressions (that are created online in a spontaneous speech). A plan for how to stimulate this type of language should be made. Detailed suggestions can be found in the implication sections of the papers in this thesis.

In addition to directly teaching figurative language comprehension, the focus should be on improving language form and content as it has been shown to be closely related to figurative comprehension. For example, defining words would be beneficial because it requires reflection and makes the person think of different aspects of the word's meaning (Nippold, 2016). Furthermore, the ability to define words is a metalinguistic skill closely related to cognition and academic achievement and inability to define the meanings of terms can result in poor communication. Working on improving meta-linguistic ability can facilitate an analysis of the words encoded in metaphoric expressions, and an examination of the linguistic

context the metaphor is embedded in (Nippold, 2016). However, as Brynskov et al., (2016) puts it, it is important that intervention should ' begin with a thorough analysis of the precise level of pragmatic skills in the individual child, so that intervention can start at a point where learning is within reach and intervention does not become stressful' (p.104).

5.4 Conclusive remarks

This article-based PhD-thesis investigated figurative language comprehension in individuals with ASD compared to individuals with TD using a systematic review, a meta-analysis and a cross-sectional design. Although individuals with ASD as a group scored lower on figurative language tasks, the differences were associated with variables such as lexical, semantic and syntactic knowledge and group matching strategies, as well as differences in figurative language tropes rather than the ASD *per se*. Although the influence of the task properties could not be examined statistically, the consequent ignorance of considering the potential influence of this important variable on the performance might have contributed to faulty interpretation of the findings of earlier studies. Not examining these variables and relying on only the diagnostic status, IQ or a single language ability measure has led to the shaping of an erroneous view of universally impaired figurative language in ASD. This has contributed in the strengthening of the absolutistic view of pragmatic language impairment in individuals with ASD and this thesis serves to challenge this view.

This thesis provides a better understanding of figurative language comprehension in ASD. Findings can ultimately inform the designing of effective interventions in Norway and internationally, aiming to improve social communication skills of individuals with ASD.

References

References

Adams, C. (2002). Practitioner Review: The assessment of language pragmatics. Journal of Child Psychology and Psychiatry, 438, 973–987. doi.org/10.1111/j.14697610.1970.tb01012.x

Adams, C. (2005). Social communication intervention for school-age children: rationale and description. Seminars in Speech and Language. 26(3), 181-188. doi:10.1055/s-2005-917123. PMID: 16155856

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders: DSM-5. Arlington, Va.: American Psychiatric Association

Andrés-Roqueta, C., & Katsos, N. (2017). The contribution of grammar, vocabulary and theory of mind in pragmatic language competence in children with autistic spectrum disorders. Frontiers in Psychology, 8, 996. doi:10.3389/fpsyg.2017.00996

Baird, G. (2008). Autistic symptomatology and language ability in autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry*, 49(11), 1184-1192

Baird, G., & Norbury, C.F. (2015). Social (pragmatic) communication disorders and autism spectrum disorder. Archives of Disease in Childhood, 101, 745-751. doi: 10.1136/archdischild-2014-306944

Bambini, V., Bertini, C., Schaeken, W., Stella, A., & Di Russo, F. (2016). Disentangling metaphor from context: An ERP study. *Frontiers in Psychology*, 7, 559. doi:10.3389/fpsyg.2016.00559

Baron-Cohen, S., Rebecca C. Knickmeyer, R.C., K. Belmonte, M.K. (2005). Sex Differences in the Brain: Implications for Explaining Autism. *Science*, 310, (5749), 819-823. doi:10.1126/science.1115455

Baron-Cohen, S. (2000). Theory of mind and autism: A fifteen year review. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), Understanding other minds: Perspectives from developmental cognitive neuroscience (pp. 3-20). New York, NY, US: Oxford University Press

Baron-Cohen,S., Leslie, A.M., Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21(1), 37-46. doi.org/10.1016/0010-0277(85)90022-8

Billstedt E, Gillberg I. C, Gillberg C. (2007). Autism in adults: symptom patterns and early childhood predictors. Use of the DISCO in a community sample followed from childhood. 48(11), 1102-10. doi.org/10.1111/j.1469-7610.1970.tb01012.x

Bishop D. V. (2010). Overlaps between autism and language impairment: phenomimicry or shared etiology? *Behavior genetics*, 40(5), 618–629. doi:10.1007/s10519-010-9381-x

Bishop, D.V.M. (2014). *Uncommon understanding. Development and disorders of language comprehension in children*. Classic edition published by Psychology Press

Bloom, L., & Lahey, M. (1978). Language development and language disorders. New York: Wiley

Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein H.R. (2009). *Introduction to meta-analysis*. Chichester, England: John Wiley & Sons

Bornstein, M. H., Hahn, C. S., & Suwalsky, J. T. (2013). Physically developed and exploratory young infants contribute to their own long-term academic achievement. Psychological science, 24(10), 1906–1917. doi:10.1177/0956797613479974

Bornstein, M. H., Hahn, C. S., Putnick, D. L., & Pearson, R. M. (2018). Stability of core language skill from infancy to adolescence in typical and atypical development. Science advances, 4(11), eaat7422. doi:10.1126/sciadv.aat7422

Boucher, J.M. (2017). *Autism Spectrum Disorder: characteristics, causes and practical issues*. Second edition. New York: SAGE Publications Ltd.

Brock, J. Norbury, C., Einav, S., & Nation, K. (2008). Do individuals with autism process words in context? Evidence from language-mediated eye-movements. *Cognition*, 108(3), 896-904. doi: 10.1016/j.cognition.2008.06.007

Brynskov, C., Krøjgaard P., and Eigsti, I. (2016). Language and communication in children with autism: do research and clinical practice converge in Denmark? *Nordic Psychology*, 68(2), 100–113. doi.org/10.1080/19012276.2015.1071199

Cage, E. & Troxell-Whitman, Z. J. (2019). Understanding the Reasons, Contexts and Costs of Camouflaging for Autistic Adults. *Autism and Developmental Disorders*, 1-13. doi.org/10.1007/s10803-018-03878-x

Camp, E. (2009). Metaphor. In L. Cummings (Ed.), *The Pragmatics Encyclopedia*, Oxford, London and New York: Routledge

Carston, R. (2010). Lexical pragmatics, ad hoc concepts and metaphor: a relevance theory perspective. *Italian Journal of Linguistics*, 22(1), 153–180

Carston, R. (2017). Relevance theory and metaphor. E. Semini & Z. Demien. (Eds.) Routledge handbook of metaphor and language (pp.42-55). Abingdon-on-Thames: Routledge

Colston H.L & Raymond W. Gibbs Jr. (2002) Are Irony and Metaphor Understood Differently? *Metaphor and Symbol*, 17:1, 57-80. doi:10.1207/S15327868MS1701_5

Cooper, H.M., Hedges, L.V., & Valentine, J.C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation

Dahlgren, S. O., & Gillberg, C. (1989). Symptoms in the first two years of life: A preliminary population study of infantile autism. European Archives of Psychiatry & Neurological Sciences, 238(3), 169-174. doi.org/10.1007/BF00451006

Giacomo, D., & Fombonne, E. (1998). Parental recognition of developmental abnormalities in autism. *European Child & Adolescent Psychiatry*, 7(3):131-6

Dennis, M., Lazenby A.L., & Lockyer, L. (2001). Inferential language in high-function children with autism. *Journal of Autism and Developmental Disorders*, 31(1), 47–54

Devlin, A.S. (2018). The research experience. Planning, conducting, and reporting research. Sage publications. Los Angeles. London. New Dehli. Singapore. Washingtone DC. Melbourne

Dunn, L., Dunn, L., Whetton, C., & Burley, J. (1997). The British Picture Vocabulary Scale II. Windsor, UK: GL Assessment

Dunn, M., Gomes, H., & Sebastian, M. J. (1996). Prototypicality of responses of autistic, language disordered, and normal children in a word fluency task. *Child Neuropsychology*, 2(2), 99-108. doi:10.1080/09297049608401355

Eigsti, I. M., & Schuh, J. M. (2016). Language acquisition in autism spectrum disorders: Beyond standardized language measures. In L. Naigles (Ed.), *Innovative Investigations of Language in Autism Spectrum Disorder* (pp.183-200). Berlin, Germany: APA/Walter de Gruyter

Eigsti, I. M., de Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders*, 5, 681-91. doi.org/10.1016/j.rasd.2010.09.001

Eigsti, I-M. (2017). Peabody Picture Vocabulary Test. In Volkmar F.R (ed.), *Encyclopedia of Autism Spectrum Disorders*, Springer Science+Business Media LLC 2017. doi: 10.1007/978-1-4614-6435-8_531-3

Eigsti, I.M., Bennetto, L. & Dadlani, M.B. J. (2007). Beyond Pragmatics: Morphosyntactic Development in Autism, Journal of autism and developmental disorders, 37 (6), 1007-1023. doi.org/10.1007/s10803-006-0239-2

Eigsti, I-M., Bennetto, L., & Dadlani, M.B. (2007). Beyond Pragmatics: Morphosyntactic Development in Autism. *Journal of autism and developmental disorders*, 37(6), 1007-23. doi.org/10.1007/s10803-006-0239-2

Eldar, E., Talmor, R., Wolf-Zukerman, T. (2010) Success and difficulties in the individual inclusion of children with autism spectrum disorder (ASD) in the eyes of their coordinators. International Journal of Inclusive Education, 14(1), 97–114. doi.org/10.1080/13603110802504150

Elsabbagh, M., Divan, G., Koh, Y. J., Kim, Y. S., Kauchali, S., Marcín, C., … Fombonne, E. (2012). Global prevalence of autism and other pervasive developmental disorders. Autism research : official journal of the International Society for Autism Research, 5(3), 160–179. doi:10.1002/aur.239

Eskes, G. A., Bryson, S. E., & McCormick, T. A. (1990). Comprehension of concrete and abstract words in autistic children. *Journal of Autism and Developmental Disorders*, 20(1), 61-73. doi.org/10.1007/BF02206857

Etz, K. E., & Arroyo, J. A. (2015). Small sample research: Considerations beyond statistical power. *Prevention Science*, 16(7), 1033–1036. doi:10.1007/s11121-015-0585-4

Filippova E, & Astington J.W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development*. 79(1), 126-38. doi:10.1111/j.1467-8624.2007.01115.x

Fletcher-Watson, S., & Happé, F. (2019). Autism. A new introduction to psychological theory and current debate. Routledge. Taylor & Francis Group. London and New York

50

Fletcher-Watson, S., Adams, J., Brook, K., Charman, T., Crane, L., Cusack, J., … Pellicano, E. (2018). Making the future together: Shaping autism research through meaningful participation. *Autism*. doi.org/10.1177/1362361318786721

Fombonne E (2005) Epidimiology of autistic disorder and other pervasive develomental disorders. *Journal of Clinical Psychiatry,* 66, 3-8

Fombonne E (2005) Epidimiology of autistic disorder and other pervasive develomental disorders. Journal of Clinical Psychiatry 66, 3-8

Fombonne, E.(1999). The epidemiology of autism. Psychological medicine. 29, 769-787

Gastgeb, O.Z., Strauss, M.S., Minshew, N.J., (2006). Do Individuals with Autism Process Categories Differently? The Effect of Typicality and Development, 77(6), 1717-1729, doi: doi.org/10.1111/j.1467-8624.2006.00969.x

Gernsbacher, M. A. (2017). Editorial Perspective: The use of person-first language in scholarly writing may accentuate stigma. *Journal of Child Psychology and Psychiatry*, 58(7), 859–861

Gernsbacher, M. A., & Frymiare, J. L. (2005). Does the Autistic Brain Lack Core Modules?. *The journal of developmental and learning disorders*, 9, 3–16. doi: 10.3389/fpsyg.2013.00691

Gernsbacher, M. A., Dawson, M., & Mottron, L. (2006). Autism: Common, heritable, but not harmful. *The Behavioral and brain sciences*, 29(4), 413–414. doi:10.1017/S0140525X06319097

Gernsbacher, M.A, Keysar, B., Robertson, R.R.W., & Werner, N.K. (2001). The role of suppression and enhancement in understanding metaphors. *Journal of Memory and Language*, 45(3), 433-450. doi:10.1006/jmla.2000.2782

Gernsbacher, M.A., & Pripas-Kapit, S.R. (2012). Who's missing the point? A commentary on claims that autistic persons have a specific deficit in figurative language comprehension. Metaphor and Symbol 27(1), 93–105. doi:10.1080/10926488.2012.656255

Gernsbacher, M.A., Dawson, M., & Goldsmith, H.H. (2005). The three reasons not to believe in autism epidemic. *Current Directions in Psychological Science*, 14, 55-58. doi: 10.1111/j.0963-7214.2005.00334.x

Gibbs, R. W., Jr., & Colston, H. L. (2012). Interpreting figurative meaning. New York, N Y, US: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139168779

Gluksberg, S. (2001). *Understanding Figurative Language: From Metaphor to Idioms* (Oxford Psychology Series). New York: Oxford University Press

Goodall, C. (2018). 'I felt closed in and like I couldn't breathe': A qualitative study exploring the mainstream educational experiences of autistic young people. Autism & Developmental Language Impairments. 3, 1-16. doi.org/10.1177/2396941518804407

Gough, D., Oliver S., & Thomas, J. (2017). An introduction to systematic reviews. Second edition. London. Sage Publications Ltd

Gould, J. and Ashton-Smith, J. (2011), "Missed diagnosis or misdiagnosis? Girls and women on the autism spectrum", *Good Autism Practice*, 12 (21), 34-41.

Hacking, J. (1999). *The social construction of what*. Harvard University Press. Cambridge. Massachussets. London, England

Hacking, J. (2006). "Kinds of people: moving targets". Web version. The British academy lecture. Read 11 April 2006 at the British Academy. Retrieved at https://www.thebritishacademy.ac.uk/sites/default/files/hacking-draft.pdf. 06.04.2019

Hale, C. M., & Tager-Flusberg, H. (2005). Social communication in children with autism: The relationship between Theory of Mind and discourse development. *Autism*, 9(2), 157-178. doi.org/10.1177/1362361305051395

Happé, F. & Frith, U. (2006). The Weak Coherence Account: Detail-focused Cognitive Style in Autism Spectrum Disorders, Journal of autism and developmental disorders, 36, 5. doi.org/10.1007/s10803-005-0039-0

Happé, F.G.E. (1993). Communicative competence and theory of mind in autism: a test of relevance theory. Cognition, 48(2), 101-1. doi:10.1016/0010-0277(93)90026-R

Happé, F.G.E. (1993). Communicative competence and theory of mind in autism: a test of relevance theory. *Cognition*, 48(2), 101-1. doi:10.1016/0010-0277(93)90026-R

Hermann, I., Haser, V., van Elst, L.T., Ebert, D., Müller-Feldmeth, D., Riedel, A., & Konieczny, L. (2013). Automatic metaphor processing in adults with Asperger syndrome: a metaphor interference effect task. *European Archives of Psychiatry and Clinical Neurosciences*, 263 (Suppl.2): S177-S187. doi:10.1007/s00406-013-0453-9

Hill, E.L. (2004). Evaluating the theory of executive dysfunction in autism. *Developmental Review*, 24(2), 189–233. doi:10.1016/j.dr.2004.01.001

Hoff, E. (2014). *Language development.* 5th edition. International edition. Wadsworth. Cengage learning. Australia. Brasil. Japan. Korea. Mexico. Singapore. Spain. United Kingdom. United States

Howlin, P. (2003). Outcome in High-Functioning Adults with Autism with and Without Early Language Delays: Implications for the Differentiation Between Autism and Asperger Syndrome. *Journal of Autism and Developmental Disorders*, 33(1):3-13. doi: 10.1023/A:1022270118899

Howlin, P. (2005). Outcomes in autism spectrum disorders. In: Volkmar F, Paul R, Klin A, Cohen D, editors. *Handbook of autism and pervasive developmental disorders*. 3rd ed. Vol. 1. New York: Wiley; pp. 201–222

Howlin, P., Mawhood, L., & Rutter, M. (2000). Autism and Developmental Receptive Language Disorder—a Follow-up Comparison in Early Adult Life. II: Social, Behavioural, and Psychiatric Outcomes. The journal of child psychology and psychiatry. 41(5), 561-578. doi.org/10.1111/1469-7610.00643

Hurlbutt, K., & Chalmers, L. (2004). Employment and Adults with Asperger Syndrome. *Focus on Autism and Other Developmental Disabilities*, *19*(4), 215–222. doi.org/10.1177/10883576040190040301

Isaksen, J., Diseth, T.H., Schjølberg, S., & Skjeldal, O.H. (2012). Observed prevalence of autism spectrum disorders in two Norwegian counties. *European journal of Peadeatric neurology*, 16(6), 592-598. doi: 10.1016/j.ejpn.2012.01.014

Kaland., N, Møller-Nielsen., A, Callesen., K, et al. (2002). A new 'advanced' test of theory of mind: evidence from children and adolescents with Asperger syndrome. *Journal of Child Psychology and Psychiatry,* 43(4), 517–528

Kalandadze, T., Norbury, C., Nærland, T., & Næss, B K-A. (2016). Figurative language comprehension in individuals with autism spectrum disorder: a meta-analytic review. *Autism*, 22, 99–117. doi:10.1177/1362361316668652

Kamio, Y., Robins, D., Kelley, E. et al. J (2007). Atypical Lexical/Semantic Processing in High-Functioning Autism Spectrum Disorders without Early Language Delay. *Journal of Autism and Developmental Disorders*, 37, 1116. doi.org/10.1007/s10803-006-0254-3

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

Katz, A.N. (2017). Psycholinguistic approaches to metaphor acquisition and use. In E. Semini & Z. Demien. (Eds.) *Routledge handbook of metaphor and language* (pp.472-485). Abingdon-on-Thames: Routledge

Kavale, K. A. (1984). Potential Advantages of the Meta-Analysis Technique for Research in Special Education. *The Journal of Special Education*, 18(1), 61–72. doi.org/10.1177/002246698401800107

Kelley, E., Paul, J.J., Fein, D. et al. Residual Language Deficits in Optimal Outcome Children with a History of Autism. *Journal of autism and developmental disorders*, 36, 807. doi.org/10.1007/s10803-006-0111-4

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism*, 20(4), 442–462. doi.org/10.1177/1362361315588200

Kim, S.H., Paul, R., Tager-Flusberg, H., & Lord, C. (2014). Language and communication in autism. In F. R. Volkmar, S.J. Rogers, R.Paul, & K. A. Pelpherey (eds.) *Handbook of autism and pervasive developmental disorders*. Fourth edition. Vol. 1. Diagnosis, Development, And Brain mechanisms. Wiley

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An Investigation of Language Impairment in Autism: Implications for Genetic Subgroups. *Language and cognitive processes*, 16(2-3), 287–308. doi:10.1080/01690960042000058

Kline, A.B. (2009). *Becoming a behavioural science researcher. A guide to producing research that matters*. New York. The Guilford Press

Kuhl, P. K., Coffey-Corina, S., Padden, D., Munson, J., Estes, A., & Dawson, G. (2013). Brain responses to words in 2-year-olds with autism predict developmental outcomes at age 6, *PLoSONE*, 8, e64967

Kwok, E.Y.L., Brown, H.M., Smyth, R.E., & Cardy, J.O. (2015). Meta-analysis of receptive and expressive language skills in autism spectrum disorder. *Research in Autism Spectrum Disorders*, 9, 202–222. doi: 10.1016/j.rasd.2014.10.008

Lewis, F. M., Murdoch, B. E., & Woodyatt, G. C. (2007). Linguistic abilities in children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 1(1), 85-100. doi.org/10.1016/j.rasd.2006.08.001

Lindsay, S., Proulx, M., Thomson, N., & Scott, H. (2013). Educators' Challenges of Including Children with Autism Spectrum Disorder in Mainstream Classrooms. *International Journal of Disability, Development and Education*, 60(4), 347-362, doi: 10.1080/1034912X.2013.846470

Lord C., Elsabbagh, M., Baird G., Veenstra-Vanderweele J. (2018). Autism spectrum disorder. *Lancet*, 392(10146), 508-520. doi: 10.1016/S0140-6736(18)31129-2

Lord, C., & Paul, R. (1997). Language and communication in autism. In D. J. Cohen & F. R. Volkmar (Eds.), *Handbook of autism and pervasive development disorders* (pp. 195-225).2nd edition. New York: John Wiley

Loucas, T., Charman, T., Pickles, A., Simonoff, E., Chandler, S., Meldrum, D., &

Loucas,T., Charman, T., Pickles, A., Simonoff, E., Chandler, S., Meldrum, D., Baird, G. (2008). Autistic symptomatology and language ability in autism spectrum disorder and specific language impairment. Journal of child psychology and psychiatry, 49(11),1184-1192. doi.org/10.1111/j.1469-7610.2008.01951.x

Lyall, K., Croen, L., Daniels, J., M. Daniele Fallin, Ladd-Acosta, C., Brian K. Lee, Bo Y. Park, Nathaniel W. Snyder, Diana Schendel, Heather Volk, Gayle C. Windham, Craig Newschaffer. (2017). The Changing Epidemiology of Autism Spectrum Disorders. *Annual Review of Public Health*, 38(1), 81-102. doi: 10.1146/annurev-publhealth-031816-044318

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

Magiati, I., Tay, X. W., & Howlin, P. (2014). Cognitive, language, social and behavioural outcomes in adults with autism spectrum disorders: A systematic review of longitudinal follow-up studies in adulthood. *Clinical Psychology Review*, 34(1), 73-86. doi.org/10.1016/j.cpr.2013.11.002

Mashal N., & Kasirer A. (2011). Thinking maps enhance metaphoric competence in children with autism and learning disabilities. *Research in developmental disabilities*, 32(6), 2045–2054. doi:10.1016/j.ridd.2011.08.012

Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual differences in children's pragmatic ability: A review of associations with formal language, social cognition, and executive functions, *Language Learning and Development*, 14(3), 186-223. doi: 10.1080/15475441.2018.1455584

Mayo, J., Chlebowski, C., Fein, D. A., & Eigsti, I. M. (2013). Age of first words predicts cognitive ability and adaptive skills in children with ASD. *Journal of autism and developmental disorders*, 43(2), 253–264. doi:10.1007/s10803-012-1558-0

McGregor, K. K., Berns, A. J., Owen, A. J., Michels, S. A., Duff, D., Bahnsen, A. J., & Lloyd, M. (2012). Associations between syntax and the lexicon among children with or without ASD and Language Impairment. *Journal of Autism and Developmental Disorders*, 42(1), 35-47

Nadig A. S., Lee I., Singh L., Kyle B., Ozonoff S. (2010). How does the topic of conversation affect verbal exchange and eye gaze? A comparison between typical development and high-functioning autism. *Neuropsychologia*, 48, 2730–2739. doi:10.1016/j.neuropsychologia.2010.05.020

Naigles & Chin (2015). Language in children with autism spectrum disorder. In L. Bavin and L. R. Naigles (eds.) The Cambridge handbook of child language (pp.637.658), Second edition. Cambridge University Press

Naigles L, Kelty E, Jaffery R,& Fein D. (2011. Abstractness and continuity in the syntactic development of young children with autism. *Autism Research*, 4(6), 422–437. doi: 10.1002/aur.223. Epub 2011 Oct 19

Minshew, N.J., Goldstein,G., Taylor H.G., & Siegel, D.J. (1994). Academic achievement in high functioning autistic individuals. *Journal of Clinical and Experimental Neuropsychology*, 16(2), 261-270. doi: 10.1080/01688639408402637

NESH, (2006). Forskningsetiske retningslinjer for samfunnsvitenskap, humaniora, juss og teologi. Retreived 06.04.2019: http://www.etikkom.no/Forskningsetikk/Etiskeretningslinjer/Samfunnsvitenskap-jus-og-humaniora

Nicholson, A., Whalen, J. M., & Pexman, P. M. (2013). Children's processing of emotion in ironic language. *Frontiers in psychology*, 4, 691. doi:10.3389/fpsyg.2013.00691

Nippold, M.A. (1991). Evaluating and enhancing idiom comprehension in language-disordered students. *Language, Speech, and Hearing Services in Schools*, 22(3), 100–106. doi.org/10.1044/0161-1461.2203.100

Nippold, M. A., & Duthie, J. K. (2003). Mental imagery and idiom comprehension: A comparison of school-age children and adults. *Journal of Speech, Language, and Hearing Research*, 46(4), 788–799

Nippold, M. A. (2016). *Later language development: School-age children, adolescents, and young adults* (4th ed.). Austin, TX: Pro-Ed.

Norbury, C.F. (2005). The relationship between theory of mind and metaphor: evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology*, 23(3), 383-399. doi:10.1348/026151005X26732

Norbury C.F. (2004). Factors supporting idiom comprehension in children with communication disorders. *Journal of Speech, Language, and Hearing Research*, 47(5), 1179–1193

Norbury C.F & Sparks, S. (2014). Language impairment within the autism spectrum. In Sue Robson, Suzanne Flannery Quinn (eds.) *The Routledge International Handbook of Young Children's Thinking and Understanding*. 1st edition. Routledge International Handbooks

Norbury, C.F.(2013). Practitioner Review: Social (pragmatic) communication disorder conceptualization, evidence and clinical implications. *Journal of Child Psychology and Psychiatry,* 55(3), 204–216. doi:10.1111/jcpp.12154

Norbury, C.F. (2014). Autism spectrum disorders and communication. In L. Cummings. (ed.). *The Cambridge handbook of communication disorders* (pp. 141-157). Cambridge: Cambridge University Press

Næss, K.A, Lyster, S.A, Hulme, C, Melby-Lervåg, M. (2011). Language and verbal short-term memory skills in children with down syndrome: a meta-analytic review. *Research in Developmental Disabilities*, 32(6), 2225–2234. doi.org/10.1016/j.ridd.2011.05.014

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: A review. *Neuroscience and Biobehavioral Reviews*, 36(2), 836-854. doi: 10.1016/j.neubiorev.2011.11.008

Orlikoff, O.F., Schiavetti, N., & Metz, D.E. (2015). *Evaluating Research in Communication Disorders. Seventh edition*. Pearson education

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

Perlini, C., Bellani, M., Finos, L., Lasalvia, A., Bonetto, C., Scocco, P.,… the GET UP group. (2018). Non-literal language comprehension in a large sample of first episode psychosis patients in adulthood. *Psychiatry Research*, 260, 78–89. doi:10.1016/j.psychres.2017.11.032

Pexman P. M. (2008). It's fascinating research: the cognition of verbal irony. *Current directions in psychological science*. 17, 286–290 10.1111/j.1467-8721.2008.00591.x

Pexman P.M., & Glenwright, M. (2007). How do typically developing children grasp the meaning of verbal irony? Journal of Neurolinguistics, 20(2), 178–196

Pijnacker, J., Hagoort, P., Buitelaar, J., Teunisse, J. P., & Geurts, B. (2009). Pragmatic inferences in high-functioning adults with autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, 39(4), 607–618. doi:10.1007/s10803-008-0661-8

Pouscoulous, N. (2011). Metaphor: for adults only? *Belgian Journal of Linguistics*, 25, 51-79. doi:10.1075/bjl.25.04pou

Pouscoulous, N. (2014). "The elevator's buttocks" metaphorical abilities in children, in Matthews, D (ed.) *Pragmatic development in first language acquisition*, John Benjamins Publishing Company, Amsterdam/Philadelphia

Rapp, A., & Wild, B. (2011). Nonliteral Language in Alzheimer Dementia: A Review. *Journal of the International Neuropsychological Society*, 17(2), 207-218. doi:10.1017/S1355617710001682

Reed, P., & McCarthy, J. (2012). Cross-modal attention-switching is impaired in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 42(6), 947-53. doi:10.1007/s10803-011-1324-8

Rogers, S.J., & Dilalla, D.L., (1990). Age of Symptom Onset in Young Children with Pervasive Developmental Disorders. *Journal of the American academy of child and adolescent psychiatry*. 29(6), 863-872. doi.org/10.1097/00004583-199011000-00004

Rowling, J.K. (2013). Harry Potter & the half-blood prince. Bloomssbury. London. New Dehli. New York. Sydney

Rundblad, G., & Annaz, D. (2010). The atypical development of metaphor and metonymy comprehension in children with autism. *Autism*, 14(1), 29–46. doi:10.1177/1362361309340667

Rutter, M. (1979). PSYCHOLOGICAL DEVELOPMENT—PREDICTIONS FROM INFANCY. *The journal of child psychology and psychiatry*. 11(1), 49-62

Rossetti, I., Brambilla, P., & Papagno, C. (2018). Metaphor comprehension in schizophrenic patients. Frontiers in Psychology, 9, 670. doi:10.3389/fpsyg.2018.00670

Rosenthal, R. (1994). Science and Ethics in Conducting, Analyzing, and Reporting Psychological Research. *Psychological Science*, 5(3), 127–134. doi.org/10.1111/j.1467-9280.1994.tb00646.x

Shakespeare. W. (1975). *As you like it.* Gramercy books. Random House Value Publishing.New York. Toronto. London. Sydney. Auckland (Original work first printed in 1623).

Schaeken, W., Van Haeren, M., & Bambini, V. (2018). The understanding of scalar implicatures in children with Autism Spectrum Disorder: Dichotomized responses to violations of informativeness. *Frontiers in Psychology*, 9, 1266. doi:10.3389/fpsyg.2018.01266

Seltzer, M. M., Shattuck, P., Abbeduto, L., & Greenberg, J. S. (2004). Trajectory of development in adolescents and adults with autism. *Developmental Disabilities Research Reviews*, 10(4), 234-247

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi experimental designs for generalized causal inference*. Houghton Mifflin Company

Short, A.B. & Schopler, E. J.(1988). Factors relating to age of onset in autism. *Journal of autism and developmental disorders*, 18, 207. doi.org/10.1007/BF02211947

Smith P.L, & Little D.R. (2018). Small is beautiful: in defense of the small-N design. *Psychonomic Bulletin and Review*. 25(6), 2093-2101. doi.org/10.3758/s13423-018-1451-8

Sperber, D., & Wilson, D. (2008). A deflationary account of metaphors. In R. Gibbs (ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 84-105). Cambridge: Cambridge University Press; Reprinted in Wilson, D. & Sperber, D. (2012). Meaning and Relevance (pp. 97-122). Cambridge: Cambridge University Press

Stone, W. L., & Yoder, P. J. (2001). Predicting Spoken Language Level in Children with Autism Spectrum Disorders. *Autism*, 5(4), 341 361. doi.org/10.1177/1362361301005004002

Swensen, L.D., Kelley, E., Fein, D., Naigles, L.R. (2007). Processes of language acquisition in children with autism: evidence from preferential looking. *Child developme*nt, 78(2), 542-557. doi.org/10.1111/j.1467-8624.2007.01022.x

Szatmari, P., Bryson, S.e., Boyle, M.H., Streiner, D.L., & Duku, E.(2003). Predictors of outcome among high functioning children with autism and Asperger syndrome. *The journal of child psychology and psychiatry.* 44(4), 520-522. doi.org/10.1111/1469-7610.00141

Tager-Flusberg, H., & Anderson, M. J. (1991). The development of contingent discourse ability in autistic children. *Child Psychology and Psychiatry*, 32(7), 1123-34

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. doi:10.1002/aur.1329

Tager-Flusberg, H., Rogers, S., Cooper, J., Landa, R., Lord, C., Paul, R., … Yoder, P. (2009). Defining spoken language benchmarks and selecting measures of expressive language

development for young children with autism spectrum disorders. *Journal of speech, language, and hearing research:J*SLHR, 52(3), 643–652. doi:10.1044/1092-4388(2009/08-0136)

Tager-Flusberg, H., & Joseph, R.M. (2003). Identifying neurocognitive phenotypes in autism. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 358(1430), 303–314. doi: 10.1098/rstb.2002.1198

Tager-Flusberg, H. (2000). Language and understanding minds: Connections in autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), Understanding other minds: Perspectives from developmental cognitive neuroscience (pp. 124-149). New York, NY, US: Oxford University Press

Tager-Flusberg H, Calkins S, Nolin T, Baumberger T, Anderson M, Chadwick-Dias A. A. (1990). longitudinal study of language acquisition in autistic and Down syndrome children. *Journal of Autism and Developmental Disorders*, 20, 1–21

Tager-Flusberg, H. (1992). Autistic children's talk about psychological states: Deficits in the early acquisition of a theory of mind. *Child Development*, 63(1), 161-172. doi:10.1111/j.1467-8624.1992.tb03604.x

Tek, S., Mesite, L., Fein, D., & Naigles, L. (2014). Longitudinal analyses of expressive language development reveal two distinct language profiles among young children with autism spectrum disorders. *Journal of autism and developmental disorders*, 44(1), 75–89. doi:10.1007/s10803-013-1853-4

Tek, S., Jaffery, G., Fein, D., & Naigles, L. R. (2008). Do children with autism spectrum disorders show a shape bias in word learning? *Autism research*, 1(4), 208-22

Thorndike, R. M., & Thorndike-Christ, T. (2014). *Measurement and evaluation in psychology and education* (8th ed). Harlow: Pearson

Tomblin, B. (2011). Co-morbidity of autism and SLI: kinds, kin and complexity. *International journal of language and communication disorders*. 46(2). 127-137. doi.org/10.1111/j.1460-6984.2011.00017.x

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

Volden, J., Coolican, J., Garon, N., White, J., Bryson, S. (2009). Brief report: pragmatic language in autism spectrum disorder: relationships to measures of ability and disability. *Journal of Autism and Developmental Disorders*,39(2):388-93. doi: 10.1007/s10803-008-0618-y

Volkmar, F. R., & Wolf, J. M. (2013). When children with autism become adults. *World psychiatry* (WPA), 12(1), 79–80. doi:10.1002/wps.20020

Vosniadou, S., Ortony, A., Reynolds, R.E., & Wilson, P.T. (1984). Sources of difficulty in the young child's understanding of metaphorical language. *Child Development*, 55(4), 1588-1606. doi:10.2307/1130028

Vulchanova M, Saldaña D, Chahboun S, et al. (2015). Figurative language processing in atypical populations: the ASD perspective. *Frontiers in Human Neuroscience*, 9, 24. doi: 10.3389/fnhum.2015.00024

Wei, X., Yu, J. W., Shattuck, P., McCracken, M., & Blackorby, J. (2013). Science, technology, engineering, and mathematics (STEM) participation among college students with an autism spectrum disorder. *Journal of autism and developmental disorders*, 43(7), 1539–1546. doi:10.1007/s10803-012-1700-z

Whalen, J.M., Pexman, P.M. Gill A.J:, & Nowson, S. (2013).Verbal irony use in personal blogs. *Behaviour & Information Technology*, 32(6), 560-569, doi: 10.1080/0144929X.2011.630418

Whalen, J. M., Pexman, P. M., & Gill, A. J. (2009). "Should Be Fun—Not!": Incidence and Marking of Nonliteral Language in E-Mail. *Journal of Language and Social Psychology*, 28(3), 263–280. doi.org/10.1177/0261927X09335253

Whyte, E.M,, Nelson, K.E., & Scherf, K.S. (2014). Idiom, syntax, and advanced theory of mind abilities in children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*, 57(1),120–130

Williams, D., Botting, N., & Boucher, J. (2008). Language in autism and specific language impairment: Where are the links? *Psychological Bulletin,* 134(6), 944-963. doi.org/10.1037/a0013743

Wilson, D. (2011). Parallels and differences in the treatment of metaphor in relevance theory and cognitive linguistics. *Intercultural Pragmatics*. 8(2),177–196. doi: 10.2478/v10148-011-0025-1

Wilson D and Sperber D (2012) *Meaning and Relevance*. Cambridge: Cambridge University Press

Winner, E. (1988). *The point of words: Children's understanding of metaphor and irony*. USA: Harvard University Press

World Health Organization. (1992). ICD-10: International statistical classification of diseases and related health problems (10th revision ed.). Geneva: World Health Organization

Ying Sng, C., Carter, M., & Jennifer Stephenson. (2018). A systematic review of the comparative pragmatic differences in conversational skills of individuals with autism. *Autism & Developmental Language Impairments*, 3, 1–24 doi: 10.1177/2396941518803806

Zubrick, S. R., Taylor, C. L., Rice, M. L., & Slegers, D. W. (2007). Late language emergence at 24 months: an epidemiological study of prevalence, predictors, and covariates. *Journal of speech, language, and hearing research: JSLHR*, 50(6), 1562–1592. doi:10.1044/1092-4388(2007/106)

Errata

Place	Original text	Corrected text	Description of the made correction
Page 11	Ingrid, F.,	Falkum, I.L.,	Corrected the name of the co-author
Page 14	[…] yet to be found; therefore	[…] yet to be found. Therefore	Changed ';' with '.' And, accordingly, 't' is replaced with 'T'
Page 15	DSM	DSM-5	Added missing '-5'
Page 15	DSM-4	DSM-IV	'-4' is replaced with '-IV'
Page 17	Nadig, Lee, Singh, Kyle, Ozonoff, 2010); Sng, Carter, & Stephenson, 2018)	Nadig, Lee, Singh, Kyle, Ozonoff, 2010; Sng, Carter, & Stephenson, 2018)	Deleted ')'

Note: paper 3 has been published and has the following reference:

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

Dissertational Papers

I

II

III

Paper I:

I

Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review

**Tamar Kalandadze[1], Courtenay Norbury[2], Terje Nærland[3]
and Kari-Anne B Næss[1]**

Abstract

We present a meta-analysis of studies that compare figurative language comprehension in individuals with autism spectrum disorder and in typically developing controls who were matched based on chronological age or/and language ability. A total of 41 studies and 45 independent effect sizes were included based on predetermined inclusion criteria. Group matching strategy, age, types of figurative language, and cross-linguistic differences were examined as predictors that might explain heterogeneity in effect sizes. Overall, individuals with autism spectrum disorder showed poorer comprehension of figurative language than their typically developing peers (Hedges' $g = -0.57$). A meta-regression analysis showed that group matching strategy and types of figurative language were significantly related to differences in effect sizes, whereas chronological age and cross-linguistic differences were not. Differences between the autism spectrum disorder and typically developing groups were small and nonsignificant when the groups were matched based on the language ability. Metaphors were more difficult to comprehend for individuals with autism spectrum disorder compared with typically developing controls than were irony and sarcasm. Our findings highlight the critical role of core language skills in figurative language comprehension. Interventions and educational programmes designed to improve social communication skills in individuals with autism spectrum disorder may beneficially target core language skills in addition to social skills.

Keywords

autism spectrum disorder, figurative language, meta-analysis, metaphor

Introduction

Autism spectrum disorder (ASD) is a pervasive developmental disorder that is characterized by impairments in social interaction and social communication, and by restricted, repetitive behaviours and interests (American Psychiatric Association (APA), 2013). While the relationship between ASD and core language skills has been debated for several years, many studies have consistently reported that the social use of language, particularly figurative language (the ability to go beyond what is explicitly stated), is universally impaired in individuals with ASD (e.g. Dennis et al., 2001; Happé, 1993, 1994). In particular, literal interpretations of utterances with intended nonliteral meanings have been considered characteristic of these individuals (e.g. Happé, 1993; MacKay and Shaw, 2004).

Despite this level of consensus, a growing body of research indicates that a subset of individuals with ASD do not differ significantly from typically developing (TD) controls with similar language ability (LA) in terms of selected measures of figurative language (Hermann et al., 2013; Norbury, 2004, 2005). Even in studies in which participants with ASD receive lower scores than their TD controls, performance is often above chance level (e.g. Wang et al., 2006), indicating that individuals with ASD are not consistently inclined towards literal meanings. These

[1]University of Oslo, Norway
[2]University College London, UK
[3]Oslo University Hospital, Norway

Corresponding author:
Tamar Kalandadze, Department of Special Needs Education, University of Oslo, PO Box 1140, Blindern, 0318 Oslo, Norway.
Email: tamar.kalandadze@isp.uio.no

findings indicate that deficits in figurative language comprehension may not be unique or/and universal among individuals with ASD, fuelling debates regarding the source and the extent of the difficulty associated with figurative language comprehension in people with ASD.

In this article, we present a meta-analysis of studies that have explicitly compared individuals with ASD with TD controls in terms of their figurative language comprehension. We focus on whether (a) the group matching strategy (i.e. whether the ASD and TD groups are matched based on chronological age (CA), LA or both), (b) differences with regard to CA, (c) the type of figurative language (*tropes*) measured and (d) cross-linguistic differences (the languages in which the studies are conducted) can explain between-study variance. A better understanding of these factors may have far-reaching implications for education and clinical interventions involving individuals with ASD.

What is figurative language?

Generally, in figurative language, the intended meanings of the words, sentences and expressions used do not coincide with their literal meanings (Gluksberg, 2001). When speaking figuratively, speakers mean something other than what they literally say (Gibbs and Colston, 2012). Therefore, to understand figurative language, an individual must be able to grasp the speaker's intention in a given context (Rapp and Wild, 2011).

The most common examples of figurative language include *metaphors* (e.g. 'Love is a journey'), which involves 'understanding and experiencing one kind of thing in terms of another' (Lakoff and Johnson, 2003: 5), and *verbal irony* (e.g. 'What nice weather' – to describe a rainy day), which 'consists in echoing a thought (e.g., a belief, an intention, a norm-based expectation) attributed to an individual, a group or to people in general, and expressing a mocking, skeptical or critical attitude to this thought' (Wilson and Sperber, 2012: 125). A sub-type of irony is *sarcasm*, which is often intended to insult or wound (Rapp and Wild, 2011) (e.g. 'I just love your blouse', said by someone who does not actually like the blouse).

Traditionally, figurative language has been regarded as deviant of so-called literal language. By contrast, it is now well established that figurative language is a ubiquitous part of daily language and social communication. For instance, approximately 8% of adult utterances in conversations among friends contain some irony (Gibbs, 2000), teachers frequently use figurative language when presenting their lessons to students (Kerbel and Grunwell, 1997), and literature is replete with metaphors (Colston and Kuiper, 2002). Consequently, figurative language comprehension influences social relationships, social participation and educational achievement (Cain et al., 2005; Kerbel and Grunwell, 1997; Nippold, 1991; Swineford

et al., 2014). As a result, deficits in figurative language comprehension may seriously affect an individual's life.

Figurative language comprehension in TD individuals

In TD individuals, the ability to comprehend figurative language emerges in early childhood, continues to develop steadily into adolescence, and improves throughout adulthood (e.g. Ackerman, 1982; Dews et al., 1996; Falkum et al., 2016; Hancock et al., 2000; Pexman et al., 2005; Pexman and Glenwright, 2007; Rundblad and Annaz, 2010a; Semrud-Clikeman and Glass, 2010; Winner, 1988; Winner et al., 1976). However, findings vary regarding the age at which the ability to comprehend different tropes emerges, which may partially be due to theoretical and methodological inconsistencies across studies (Pouscoulous, 2011; Winner et al., 1976).

The developmental achievements that underpin figurative language comprehension are the subject of theoretical debate. Some claim that figurative language develops in parallel with the acquisition of a theory of mind (ToM: the capacity to attribute one's own mental states and those of others) due to the presumed need to appreciate a speaker's intended message, which may not be literal (Happé, 1993). Pouscoulous (2011) suggests that the cognitive means for understanding figurative language (e.g. metaphor) are present by the time a child starts to speak. As children age, their language skills, world knowledge, and cultural expertise develop, which may account for their improved figurative language comprehension. Importantly, because figurative language tropes differ in terms of structure and processing demands (Colston and Gibbs, 2002), they seemingly require different mechanisms to be understood.

Notably, when considering the role of ToM and/or language skills in figurative language comprehension, the well-documented close relationships between TD children's linguistic and communicative skills (semantic, syntax and pragmatics) and ToM make teasing out the independent contributions of ToM and language skills in figurative language comprehension difficult. For instance, accuracy in standard false-belief tasks has been found to be strongly correlated with participants' receptive and expressive language abilities (Pons et al., 2009).

Figurative language comprehension in individuals with ASD

A number of studies have shown that individuals with ASD have difficulty understanding figurative language (e.g. Happé, 1993; Kaland et al., 2002; Rundblad and Annaz, 2010b). Two primary explanations for such difficulties have been proposed. First, the social cognitive profile that is suggested to be characteristic of individuals with ASD, including deficits in ToM development (Baron-Cohen

et al., 1985), is thought to account for specific deficits in their figurative language comprehension. Happé (1993) conducted the first experimental research to make explicit connections between ToM and figurative language comprehension in individuals with ASD. She examined relevance theory (Sperber and Wilson, 1986) and ToM explanations for individuals with ASD and found that only the individuals with ASD who passed the first-order ToM tasks (inferring a person's mental state, e.g. what he/she thinks) performed well on metaphor tasks, whereas performance on irony tasks required that these individuals be capable of passing second-order ToM tasks (considering embedded mental states (Baron-Cohen, 2001), e.g. what he thinks that she thinks). The study concluded that ToM understanding predicted performance on metaphor and irony tasks.

Second, figurative language comprehension deficits are neither universal nor specific to individuals with ASD (Gernsbacher and Pripas-Kapit, 2012), but they can be related to an individual's structural language skills (vocabulary and syntax) (Norbury, 2004, 2005; Whyte et al., 2014). For example, Norbury (2004) found that children and adolescents with ASD did not show impairments in figurative language comprehension compared with TD controls when the former's structural language abilities (vocabulary and syntax) were within the normal range.

This hypothesis is supported by recent findings that reveal impairments in structural language skills (e.g. syntax) in some linguistically able individuals with ASD (Brynskov et al., 2016; Eigsti et al., 2011). Given the evidence that syntax is one of the most important predictors of success in, for example, metaphor comprehension in TD individuals (Pouscoulous, 2014), the impairments in syntactic ability in many individuals with ASD may affect their figurative language comprehension (e.g. Whyte et al., 2014).

Additionally, comparisons between distinct neurocognitive phenotypes within ASD reveal different patterns of language comprehension. More specifically, the autism language impairment (ALI) phenotype includes individuals with ASD who meet standard diagnostic criteria for language impairment (Tager-Flusberg, 2006; Tager-Flusberg and Joseph, 2003). The problems that individuals with ALI experience in comprehending structural language vary, and they are not unique to ASD or necessarily related to the severity of core ASD symptoms or overall cognitive functioning (Tager-Flusberg and Joseph, 2003). When individuals with ALI are compared with individuals with ASD, who score within the normal range on standard language tests (Autism language normal (ALN)), differences in figurative language comprehension are evident (Gernsbacher and Pripas-Kapit, 2012; Norbury, 2004, 2005).

Gernsbacher and Pripas-Kapit (2012) argued that, when researchers control for language comprehension in ToM tasks, differences in figurative language comprehension

between ASD and TD groups disappear. Similarly, studies have shown that core language skills remain a significant predictor of figurative language comprehension in individuals with ASD after ToM has been considered (Norbury, 2005).

In addition to these two primary explanations about the involvement of ToM or core/structural language skills in figurative language comprehension, several studies have concluded that both ToM and language (syntax) are uniquely related to figurative language comprehension (idioms) in individuals with ASD (e.g. Whyte et al., 2014). However, given that relatively few studies have examined ToM ability alongside language skills in individuals with ASD, the extent to which language skills and ToM independently contribute to figurative language comprehension remains to be explained.

Poor figurative language comprehension seems to be a persistent challenge for individuals with ASD (De Villiers et al., 2011; MacKay and Shaw, 2004), with deficits reported in adults as well (e.g. Ozonoff and Miller, 1996). However, due to the paucity of longitudinal studies investigating figurative language development in individuals with ASD, conclusions on the developmental path of this skill in ASD are difficult to draw. A cross-sectional study by Whyte and Nelson (2015) found that, for 7- to 12-year-old children with ASD, performance on nonliteral language tasks increased significantly with CA. Therefore, we hypothesize that CA may be an important determinant of figurative language comprehension in individuals with ASD, as it is in TD individuals.

Figurative language comprehension deficits in individuals with ASD have been reported in studies conducted in different languages (e.g. Adachi et al., 2004; Kaland et al., 2002). Given the differences between languages and the influence of culture on language, figurative language may differ depending on the language used. However, until now, most studies have been conducted in English-speaking countries, which may lead to language and cultural biases.

In summary, there is a significant gap in our knowledge regarding the extent of figurative language comprehension deficits in individuals with ASD relative to TD controls. Moreover, there is little consensus regarding the factors that underlie figurative language abilities in individuals with ASD – a situation that is exacerbated by inconsistent and contradictory findings in the literature. Resolving these issues is crucial to identify appropriate targets and methods for interventions.

Previous reviews

Although several narrative reviews have focused on figurative language comprehension in individuals with ASD (e.g. Gernsbacher and Pripas-Kapit, 2012; Hobson, 2012; Lyons and Fitzgerald, 2004; Melogno et al., 2012; Passanisi

and Di Nuovo, 2015; Pexman, 2008; Samson, 2013; Vulchanova et al., 2015), no prior meta-analysis has summarized group differences and similarities in figurative language comprehension between individuals with ASD and TD individuals.

The current study

The present meta-analytic review examines (a) the differences and similarities between individuals with ASD and TD controls with regard to figurative language comprehension and (b) variables (group matching strategy, CA, tropes, and cross-linguistic differences) that may explain the differences in the results across studies.

Given that research on figurative language comprehension in individuals with ASD often involves small samples, which can threaten statistical power (Næss et al., 2011), meta-analyses are useful because they statistically aggregate study findings, provide the effect sizes of group differences and thus offer increased statistical power (Borenstein et al., 2009).

Method

To ensure methodological quality, the present meta-analysis was designed and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (www.prisma-statement.org).

Literature search

The electronic database search was conducted under a University Librarian's supervision. First, we searched through electronic databases (PsychInfo, Eric, Embase, Web of Science, Medline, ScienceDirect, Linguistics and Language Behavior Abstracts (LLBA) and NoRART) for studies published until February 2016, using combinations of keywords related to ASD – *ASD, Asperger Syndrome or asperger**, *Autism or autis**, and *Pervasive Developmental Disorders or 'pervasive developmental disorder*'* – crossed with keywords related to figurative language – *Figurative Language or 'figurative language'*, *'figure* of speech'*, *humor or humour, hyperbole**, *idiom**, *'indirect speech'*, *irony, metaphor**, *metonymy*, *'non-literal language'*, *sarcasm* and *simile*. Second, we examined the references of the included studies as well as relevant narrative reviews to retrieve additional sources that were not already included in the search results. Third, we manually searched through issues of the *Autism, Autism Research*, the *Journal of Autism and Developmental Disorders, Research in Developmental Disabilities*, and the *International Journal of Language and Communication Disorders* as well as the Publications in Research page on autism. We also examined a special issue of *Metaphor and Symbol* (2012; 27(1))

that was devoted to ASD. Fourth, to minimize potential publication bias, we searched for grey literature and also emailed key authors in the field, asking them for unpublished studies. Including grey literature is important as, in general, evidence shows that studies that report large effect sizes and significant results are more easily published than studies that report null findings or small effect sizes (Borenstein et al., 2009).

Study inclusion criteria

Selected articles were required to meet the following predetermined inclusion criteria: (1) the results of any trope(s) of figurative language comprehension, measured with either standardized or nonstandardized assessment tasks, had to be reported. Several studies involved one or more measures of figurative language tropes as an aspect of ToM, but only pure measures of figurative language were included in this study. When insufficient data were available to calculate the effect sizes and when necessary data could not be obtained from the author(s), studies were excluded. (2) Participants had to be diagnosed with ASD using the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) or *International Classification of Diseases* (ICD) diagnostic criteria, and they had to have a verbal IQ (VIQ) score of >70. Thus, studies that examined participants with normal intelligence (i.e. high-functioning ASD (HFASD), high-functioning autism (HFA), or Asperger syndrome or Pervasive Developmental Disorder–Not Otherwise Specified (PDD-NOS)) were included. (3) A comparison group of CA- and/or LA-matched TD controls had to be included, and (4) studies had to be reported in English, Russian, Norwegian, Swedish or Danish because at least one of the authors master one or more of these languages.

Screening process

The first author and a research assistant judged the relevance of the abstracts obtained from the search results. When an abstract contained insufficient information, the full-text article was reviewed. The papers that met the inclusion criteria based on the abstracts were examined. Finally, 41 studies were coded. For further information on the screening process and the reasons that studies were excluded, see Figure 1.

Coding

Study characteristics. Study characteristics (title, author(s) and publication year) were coded for descriptive purposes. For the main analysis, we coded the number of ASD participants and the number of TD participants in addition to inferential statistics based on means and standard deviations (SDs), mean rates of correct answers, *p*-values or chi-square values. Because computing an overall effect
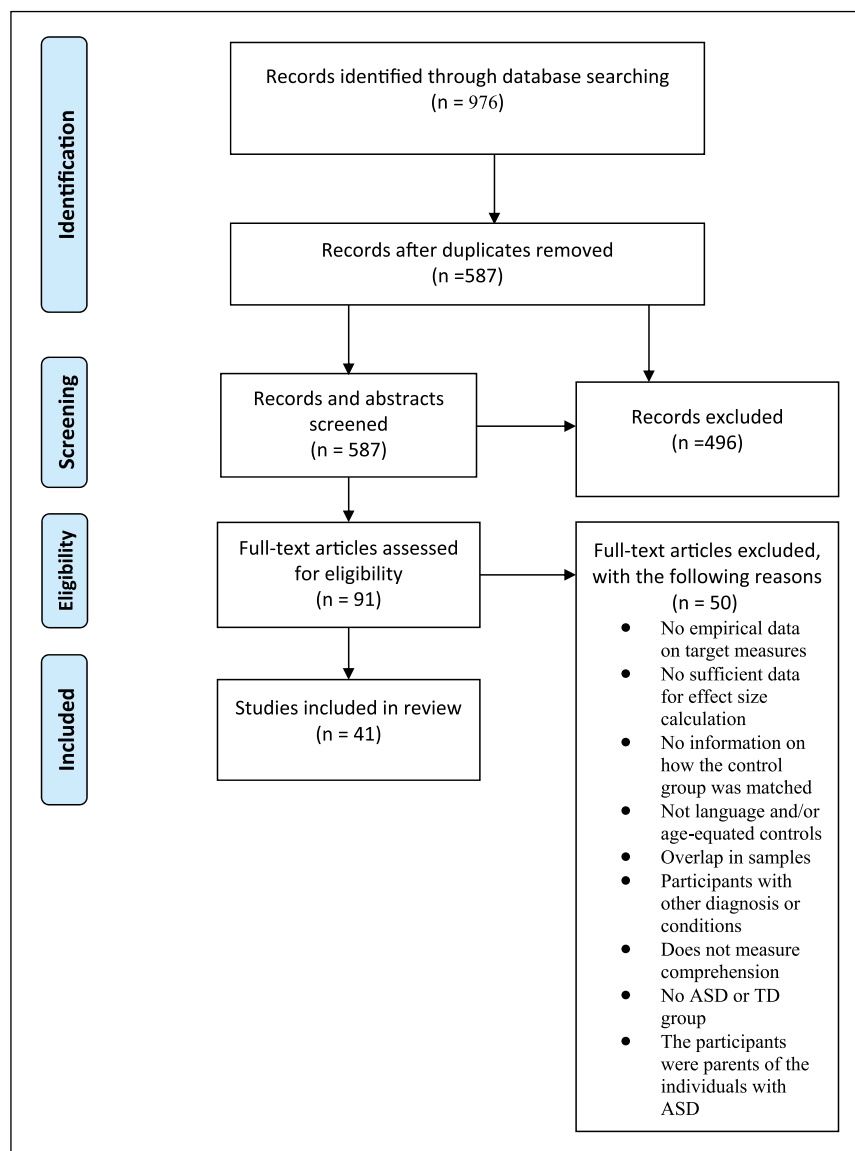
Figure 1. Flow diagram for the search and inclusion of studies.

size more than once based on information from the same sample can lead to incorrect estimates (Borenstein et al., 2009), for the studies that included multiple data collection points (e.g. intervention studies), only the first data collection point was coded. The first time point usually provides the largest sample due to attrition over time and because the results may not be influenced by any intervention effects; in descriptive studies with overlapping samples, the data from the study with the largest sample were coded; for studies that included measures of neuroimaging, only behavioural results were coded.

Predictor variables. The following predictor variables were selected and coded.

Age. The mean ages of the participants with ASD and the participants with TD were coded.

Group matching strategy. We coded whether the ASD and TD groups were matched based on CA, LA or both. Details about group matching variables are presented in Appendix 1.

Tropes. All figurative language tropes that were examined in the included studies were coded.

Cross-linguistic differences. The languages in which the studies were conducted were coded into two categories: (1) the English language (26 studies) and (2) other languages: Hebrew (4 studies), Japanese (2 studies), Taiwanese (2 studies), German (1 study), Danish (1 study), Korean (1 study), French (1 study), Chinese (1 study), Cantonese (1 study) and Dutch (1 study).

We also planned to use ToM as a predictor variable, but, due to the small number of studies that examined ToM

independently of figurative language ($n = 7$; Adachi et al., 2004; De Villiers et al., 2011; Huang et al., 2015; Martin and McDonald, 2004; Norbury, 2004; Rundblad and Annaz, 2010b; Whyte et al., 2014), no analysis including ToM could be conducted.

Multiple subgroups and multiple outcomes within studies. The included studies involve complex data structures: (a) *independent subgroups* within studies (e.g. children, adolescents and adults), (b) *dependent subgroups* within studies (e.g. one target group compared with two TD control groups or two target groups compared with one TD control group) and (c) *multiple outcomes* within studies (e.g. various tropes of figurative language or various forms of the same trope, such as novel and conventional metaphors).

Treating dependent effect sizes as independent introduces bias by giving more weight in the meta-analysis to the studies with multiple outcomes or more than two groups (Scammacca et al., 2014). To avoid this problem, several considerations were made prior coding. All the *independent and dependent subgroups* within the included studies were coded and they could contribute either as separate scores or as composite scores in the analysis. The choices made with regard to treating the information that comes from independent and dependent groups are explained in detail in the section on the meta-analytic procedure and analysis.

Two independent raters coded 100% of the included studies: the first author of this article and a doctoral candidate who is trained in meta-analysis. The inter-rater Cronbach's alpha was $\alpha = 0.99$ for CA, $\alpha = 1.00$ for tropes, and $\alpha = 0.98$ for the outcome measures. The disagreements between the raters were resolved by consulting original papers and/or via discussions between the raters.

Meta-analytic procedures and analysis

All statistical analyses were conducted using *Comprehensive Meta-Analysis* (CMA) Software, version 3 (Biostat). The data reported in different formats were combined in the same analysis because CMA computes effect size from different formats.

The effect sizes for all studies were computed using Hedges' *g*, which is corrected for small sample sizes and thus tends to be unbiased for small samples (Hedges, 1981). Cohen's general guidelines were used to judge the effect size. Based on the suggestions from Cohen (1968), effect sizes were referred as small ($d = 0.2$), medium ($d = 0.5$) and large ($d = 0.8$). However, note that such tentative guidelines should be interpreted with caution and preferably in relation to the practical consequences that the effect size may have (Lakens, 2013). When Hedges' *g* value was positive, individuals with ASD had the highest group mean; when its value was negative, the group difference favoured the TD group. A 95% confidence interval

(CI) was calculated for each effect size to indicate whether it was statistically significantly greater than zero. The effect is statistically significant if the CI does not cross zero.

The overall effect size was estimated by calculating a weighted average of individual effect sizes. Effect size calculations were based on a random-effects model, which, unlike the fixed-effects model, assumes that between-study variations in effect size not only result from random error but also are systematic and that the variations depend on variables, which are likely to vary from study to study (Borenstein et al., 2009). The assumption that effect sizes are heterogeneous suggests that other factors beyond an ASD diagnosis have an impact on the differences in effect sizes between studies. In particular, the effect sizes were predicted to vary based on CA, group matching strategy, tropes, and the languages in which the included studies were conducted.

Heterogeneity

We examined between-study heterogeneity in effect-size distributions and degrees of homogeneity using Q and I^2 (Hedges and Olkin, 1985). The Q-statistic index shows heterogeneity in effect size across the studies, and I^2 explains that a proportion of the variance is real and not simply due to random error (Borenstein et al., 2009). A significant result on this test indicates significant heterogeneity between the studies' effect sizes.

Publication bias

A funnel plot was used to determine whether publication bias was present. A symmetrical funnel would indicate the absence of publication bias.

Multiple subgroups and multiple outcomes analysis

The *independent subgroups* within the studies were treated as separate studies because they provide independent/unique information from the different participants. By contrast, the *dependent subgroups* do not provide unique information because the same participants are included in each comparison and the resulting effect sizes are statistically dependent. The inclusion of statistically dependent effect sizes in a meta-analysis can present a serious threat to its validity (Borenstein et al., 2009; Scammacca et al., 2014). First, the inclusion of dependent effect sizes leads to incorrect estimates of the variance in the summary effect; second, in the meta-analysis, it gives more weight to studies that have multiple measures. Therefore, in this study, the results from the *dependent subgroups* were aggregated as a composite score, contributing to only one effect size and thus eliminating the dependence from, for example, the shared control group.

Creating a composite score in our meta-analysis resulted in combining the measures from the dependent subgroups that compared either one ASD to two TD groups – one matched based on LA and another matched based on CA – or compare a TD group with both ASL (ASD with language impairment) and ASO (ASD only – ASD without language impairment) groups (Norbury, 2004). Although it is considered an appropriate way to treat dependent measures, this approach obscures important findings in these studies that are directly relevant to this meta-analysis. Namely, when the ASD and TD groups are matched based on LA, inter-group differences are usually very small and nonsignificant in contrast to the findings of a comparison of the same ASD group with a CA-matched TD group. Similarly, the group difference is very small and nonsignificant when the TD group is compared with the ASO group, and large and significant when the TD group is compared with the ASL group (Norbury, 2004).

Similar to the case of dependent subgroups, all *multiple outcomes* were collapsed into a composite score, thereby contributing to only one effect size. The composite score is defined as the mean effect size in a study, with a variance that considers the correlation among the different outcomes. Thus, every study is represented by one score in the main analysis, regardless of the number of outcomes or dependent subgroups included in the mean.

Meta-regression analysis

Random-effects models were used in the meta-regression analysis to avoid the problem of the difference in the weighting of small studies and large studies in the fixed-effects models. Using random-effects models ensures that each study is weighted more evenly, irrespective of the study's sample size. To determine the strength of the predictors of the outcomes, R^2 (a percentage of the explained between-study variance) was used as an effect size. R^2 index quantifies the proportion of variance explained by the covariates (Borenstein et al., 2009).

Analysis by subgroups

To examine how the differences in effect sizes across studies varied according to the group matching strategies, we ran an *analysis by subgroup* that involved the following matching variables: CA, LA, or both CA and LA. Two studies (Pexman et al., 2011; Whyte et al., 2014) were excluded from this analysis because they included a dependent ASD group and two TD groups, and, as explained earlier, a composite score of the results was created.

Analysis by outcomes – metaphors and irony/sarcasm

Given the evidence that different tropes of figurative language are likely to be comprehended differently by individuals with ASD (Happé, 1993), we compared the studies involving metaphor with the studies involving irony and sarcasm (combined). These tropes were chosen because they were the most frequently studied tropes in the included studies.

To examine group differences across these tropes, we used the *select by outcome analysis* function of CMA. Three studies (Adachi et al., 2004; De Villiers et al., 2011; Huang et al., 2015) examined both metaphor and irony or sarcasm and were therefore included only with metaphor studies because their inclusion in both groups would lead to validation problem due to statistically dependent effect sizes. They were included with the metaphor studies because there were fewer metaphor studies compared with irony and sarcasm studies.

Two studies did not differentiate metaphors from idioms and figures of speech (Dennis et al., 2001; Landa and Goldberg, 2005); therefore, they were excluded from the analysis. Overall, 13 independent effect sizes examined different types of metaphor, and 20 independent effect sizes examined irony and sarcasm.

Results

The characteristics of each study included in the meta-analysis are presented in Appendix 1. Figure 2 shows the group differences with CIs between individuals with ASD and TD individuals in terms of the comprehension of figurative language.

Figurative language comprehension in individuals with ASD compared with that in TD controls

A total of 45 independent effect sizes, involving 1119 individuals with ASD (mean sample size = 27.97, SD = 21.12, range = 8–164) and 978 TD controls (mean sample size = 24.45, SD = 15.77, range = 8–164), examined the differences in figurative language comprehension between the two groups. The standardized mean effect size was moderate and significant (g = –0.57, 95% CI = [–0.72, –0.41], $p < 0.001$) in favour of TD individuals. The heterogeneity between studies was significant (Q (44) = 138.56, $p = < 0.001$, $I^2 = 68.24$). The funnel plot examining the publication bias in the analysis showed symmetrical distribution, indicating no publication bias.

Impact of age, tropes, matching strategy and cross-linguistic differences

The meta-regression analysis including group matching strategy, age, different tropes, and cross-linguistic differences generated a significant result (Q (9) = 28.85; $p < 0.001$), indicating that effect size is related to at least one of the covariates. The model reliably explained 41% of the variance in the effect sizes between the studies.
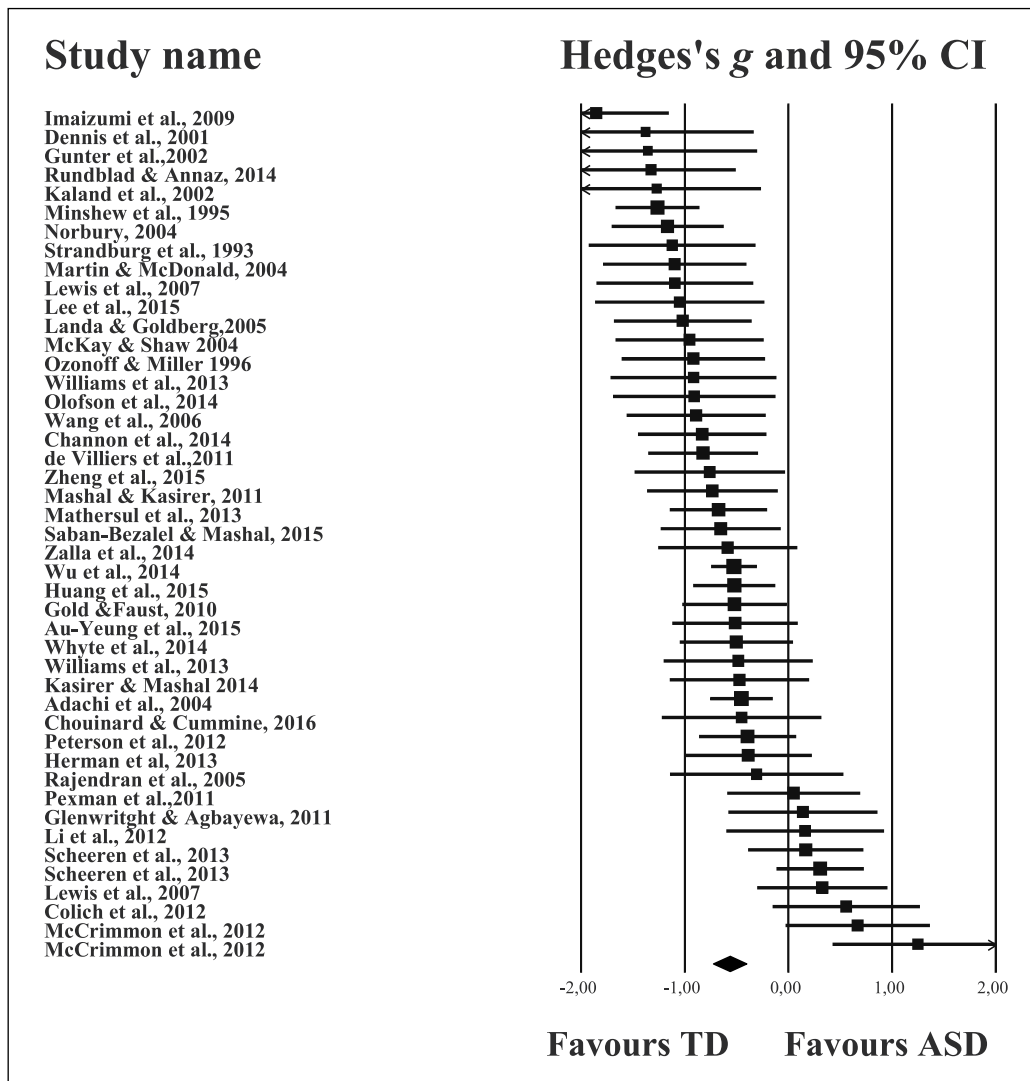
Figure 2. Overall mean effect size for group differences in figurative language comprehension comparing individuals with ASD and TD controls (Hedges' *g*, displayed by ♦) and effect sizes with confidence intervals for each study represented by horizontal lines.

However, an examination of the impact of each covariate revealed that age and cross-linguistic differences do not have a significant impact on the differences in effect sizes across studies ($p = 0.140$; $p = 0.543$, respectively). By contrast, tropes and group matching strategy were significantly related to the effect sizes ($p < 0.002$ and $p < 0.035$, respectively).

Although we could not include ToM as an independent measure due to an insufficient number of studies, a narrative summary of the results showed that some of the studies that measured ToM independently from figurative language tropes found correlations between ToM and one or more trope (Adachi et al., 2004; De Villiers et al., 2011; Martin and McDonald, 2004; Whyte et al., 2014), while others did not (Norbury, 2005; Rundblad and Annaz, 2010b). Huang et al. (2015) found that ToM understanding was partially related to figurative language comprehension

in children with ASD. Namely, children with no ToM competence showed lower scores in figurative language comprehension than children who achieved first-order or second-order ToM. However, no differences were found between first-order ToM achievers and second-order ToM achievers. Notably, different ToM tasks have been used across these studies, which could have resulted in different findings in relation to figurative language comprehension.

Impact of group matching strategy

The results of the *analysis by subgroups* showed that mean effect sizes varied across the studies depending on whether the ASD and TD groups were matched based on CA, LA or both ($g = -0.92$, $g = -0.06$ and $g = -0.55$, respectively). In particular, the group differences in studies that included LA-matched groups were small and

nonsignificant, they were large in studies that included CA-matched groups, and they were moderate in studies that included both CA- and LA-matched groups. The test to compare the difference in the results for these matching variables yielded a *Q*-value of 10.33 with 3*df* and *p* = 0.016, indicating that the group matching strategy significantly influences between-study differences.

Narrative summary of the comparisons of the dependent subgroups. Due to the small number of the studies that compared more than two dependent subgroups (*n* = 3; Norbury, 2004; Pexman et al., 2011; Whyte et al., 2014), we could not conduct a quantitative analysis. Instead, we report a brief narrative summary of the results of these studies.

Norbury (2004) found large group difference in the figurative language (idioms) comprehension of the TD group and the groups with children with ASL. However, the differences between children with ASO and the TD controls were small and nonsignificant. The children were assessed with the British Picture Vocabulary Scale (BPVS II; Dunn et al., 1997), the Concepts and Directions Subtest of the Clinical Evaluation of Language Fundamentals (CELF-III; Semel et al., 2000), and the Recalling Sentences Subtest of the CELF-III (Conti-Ramsden et al., 2001).

Pexman et al. (2011) found small differences between the CA-matched ASD and TD groups. Even smaller difference in favour of the participants with ASD was found between individuals with ASD and their LA-matched TD counterparts. The picture vocabulary subscale of the Test of Language Development–Primary, Third Edition (TOLD-P: 3; Newcomer and Hammill, 1997) was used as an LA measure to match the ASD and TD groups.

Whyte et al. (2014) showed that the ASD group had more deficits in figurative language (idioms) comprehension than did the CA-matched TD group. However, when the same ASD group was compared to another LA-matched TD group, the group difference was found to be small in favour of ASD. The participants in the LA-matched ASD and TD groups were matched based on their syntactic abilities as assessed with Syntax Construction subtest of the Comprehensive Assessment of Spoken Language (CASL; Carrow-Woolfolk, 1999).

Impact of tropes – metaphor and irony/sarcasm. The *select by outcome analysis* comparing TD and ASD groups on measures of metaphor, on the one hand, and irony and sarcasm, on the other, showed that group differences were larger in studies that measured metaphor comprehension (Hedges' *g* = −0.72) compared with the studies that measured irony and sarcasm (Hedges' *g* = −0.48).

Discussion

In this study, we investigated figurative language comprehension in individuals with ASD compared with CA- or/

and LA-matched TD controls. In addition, we examined variables that could reliably explain variations in effect size across studies. The results showed that (1) on average, individuals with ASD fall behind in their comprehension of various figurative language tropes relative to TD peers; (2) the between-study differences can reliably be explained by the group matching strategy and the trope measured; (3) age and cross-linguistic differences cannot account for between-study differences. These findings and their implications for future research and practice are discussed in the following sections.

Figurative language comprehension is challenging for many individuals with ASD

Overall, individuals with ASD exhibited moderately poorer figurative language comprehension skills compared with their TD controls. This finding is consistent with a number of studies (e.g. Happé, 1993; Kaland et al., 2002) and indicates that the social communication problems that individuals with ASD generally have may partially covary with their poor understanding of figurative language. However, the studies that included LA-matched groups displayed nonsignificant differences between the ASD and TD groups. This finding indicates that the figurative language deficit is seemingly neither universal nor unique to individuals with ASD; instead, it appears to be related to participants' language skills. This finding will be discussed in the next section.

Is figurative language comprehension related to core language skills in individuals with ASD?

Although individuals with ASD, in general, showed greater difficulties in understanding figurative language than TD controls, a high level of heterogeneity in the effect sizes across the studies examined indicates that factors other than an ASD diagnosis per se have an impact on the performance of figurative language tasks. Indeed, the results of the meta-regression analysis revealed that the group matching strategy might explain some of the between-study variance. A further examination of the matching strategy showed differences in effect size in terms of whether the ASD and TD groups were matched based on CA, LA or both. In particular, the mean effect size was small (Hedges' *g* = −0.06) for studies that used LA as a matching variable, whereas the effect size was large (Hedges' *g* = −0.92) for studies that included groups that were matched according to CA. Thus, when individuals with ASD and TD individuals are compared based on their performance on core language tests, the former and the latter receive comparable scores on figurative language tasks. This means that figurative language comprehension in individuals with ASD is closely related to their core language skulls.

When interpreting these findings, it is important to note that several studies with CA-matched groups revealed that the ASD and TD groups exhibited significant inter-group differences in language skills (e.g. Rundblad and Annaz, 2010b). These differences would have influenced these groups' performances on figurative language tasks. In addition, some studies that used LA as a group matching variable showed at least slight group difference in CA. Namely, the participants with ASD were at least slightly older than the TD controls. In these studies, the high performance of participants with ASD on the figurative language tasks may well reflect the impact of age. This finding is supported by the results of studies that used both CA- and LA-matched groups to examine the effect size. Namely, when individuals with ASD are compared with both CA- and LA-matched TD controls, the former show poorer figurative language comprehension.

However, this result may also be explained by the verbal ability tests used to match the ASD and TD groups. In particular, the most frequently used tests in studies that involved groups matched based on both CA and LA were the verbal subtests of Wechsler intelligence tests. Only a few studies used other tests such as the Peabody Picture Vocabulary Test (PPVT) or the British Picture Vocabulary Scale (BPVS), the Test of Word Knowledge (TOWK), and the National Adult Reading Test (Nelson and Wilson, 1991). Whereas the studies that included LA-matched groups used BPVS, the Concepts and Directions Subtest of the CELF-III (Semel et al., 2000), the Recalling Sentences Subtest of the CELF-III (Conti-Ramsden et al., 2001), the picture vocabulary subscale of the TOLD-P:3 (Newcomer and Hammill, 1997) and the Syntax Construction subtest of the CASL as a measure of LA to match the ASD and the TD groups.

Gernsbacher and Pripas-Kapit (2012) argue that verbal IQ (VIQ) scores may overestimate the language abilities of individuals with ASD and, in turn, obscure the language impairment of many individuals with ASD. Therefore, testing vocabulary alone is important but insufficient when assessing the language abilities of individuals with ASD. Consequently, when individuals with ASD are matched to TD controls based on VIQ scores only, a poor understanding of figurative language may be a result of problems with language comprehension that may not be easily detected by the VIQ.

The narrative summary of the results of the studies that included dependent subgroups also supports the findings with respect to the close relationship between core language skills and figurative language comprehension. Namely, when the same group of individuals with ASD is compared with both the CA- and LA-matched groups of TD individuals, the group differences are evident. In particular, the group differences between the ASD group and LA-matched TD group are very small or almost nonexistent, indicating that language skills – not an ASD diagnosis

per se – are related to problems with figurative language comprehension by individuals with ASD.

The finding that a subset of individuals with ASD, who have comparable language skills to those of TD individuals, do not have problems with figurative language comprehension indicates that figurative language comprehension is seemingly not universal or unique to individuals with ASD. This finding is supported by studies that report deficits in figurative language comprehension over a broad range of disorders, including learning disabilities, aphasia, Alzheimer's disease and Williams syndrome (e.g. Lee and Kamhi, 1990; Papagno and Caporali, 2007; Rapp and Wild, 2011).

As ToM was not included as a predictor variable in the meta-regression, we cannot make any claims about the independent contribution of ToM to figurative language comprehension. It is important to note that the finding with respect to the close relationships between core language skills and figurative language does not mean that ToM is not required to understand figurative language. Given the close relationship between core language skills and ToM, we can posit that high scores on core language tests may also be related to high scores on ToM tasks.

Our narrative summary has revealed that the findings of studies that examine the relationship between ToM and figurative language comprehension in individuals with ASD are inconsistent. On the one hand, some studies find that, once language is considered, ToM, as measured by false-belief tasks, does not explain the unique variance in, for example, metaphor comprehension (e.g. Norbury, 2004, 2005). On the other hand, a study by Whyte et al. (2014) concluded that ToM, as measured by, for example, 'Strange Stories' (Happé, 1994; O'Hare et al., 2009) and the children's version of the Reading the Mind in the Eyes (RMTE) task (Baron-Cohen et al., 2001), makes a unique contribution to figurative language (i.e. idioms) comprehension, even when core language skills are considered.

The use of different ToM measures may explain the varying results. For example, false-belief tasks are known to strongly depend on basic aspects of language that may be delayed or impaired in individuals with ASD (Hale and Tager-Flusberg, 2003; Milligan et al., 2007). In addition, false-belief tasks may not be sensitive to the continued development of ToM into adolescence (Wellman et al., 2001). The 'Strange Stories' (Happé, 1994; O'Hare et al., 2009) and the children's version of the RMTE task (Baron-Cohen et al., 2001) are considered to be more developmentally sensitive than false-belief tasks (Whyte et al., 2014).

However, the 'Strange Stories' are also closely related to verbal abilities. Teasing out the effects of cognitive and linguistic factors on figurative language comprehension is difficult (Norbury, 2005), partly because ToM tasks are verbally loaded and language plays an important role in the development of ToM (Lohmann and Tomasello, 2003; Nilsson and de López, 2016).

Is metaphor comprehension more difficult than irony and sarcasm comprehension for individuals with ASD relative to TD individuals?

In line with the theoretical literature on figurative language in TD individuals that claims that different tropes, for example, metaphor and irony, involve different pragmatic mechanisms and are thus understood differently by TD individuals (e.g. Wilson and Sperber, 2012), our study found that individuals with ASD also comprehend different types of figurative language differently. However, in contrast to, for example, Happé's (1993) study, we found that, compared with TD controls, individuals with ASD found metaphor comprehension more difficult than irony and sarcasm comprehension. One potential explanation for this unexpected finding may be the methodological characteristics of the studies, particularly the differences in the tasks used to measure different aspects of the same tropes of figurative language comprehension.

Research involving TD individuals has continually shown that an understanding of metaphor critically hinges on the task's complexity and on the effort that a person needs to make to complete this task. For example, metaphor comprehension tasks that require metalinguistic abilities (verbal definition tasks) are more demanding than, for example, multiple-choice tasks (Pouscoulous, 2014). However, multiple-choice tasks have also been questioned with respect to their ecological validity because they include literal interpretations that make them either too simplistic or too confusing (Norbury, 2004). The studies included in this meta-analysis used both multiple-choice (e.g. Adachi et al., 2004) and verbal definition tasks (e.g. Norbury, 2004; Whyte et al., 2014). In addition, many metaphor studies did not differentiate between novelty and conventionality, which are variables that may reliably explain the variance in performance.

Irony and sarcasm comprehension studies also used different tasks and different ways of administering these tasks. Although individuals with ASD are often reported to perform poorly on these tasks, some studies have found that these individuals can comprehend verbal irony in the context of computer-mediated tasks, which impose minimal social and verbal demands (e.g. Glenwright and Agbayewa, 2012). Several of the included studies used irony/sarcasm tests without too many verbal demands, which may explain the less pronounced differences between ASD and TD groups.

Are age and cross-linguistic differences less important?

The results of our study revealed that the group difference was stable across the age range studied. However, this finding should be interpreted with caution because some earlier studies involved, for example, a wide age range

of participants, spanning from childhood through late adolescence (e.g. Landa and Goldberg, 2005). Another potential explanation for this finding may be that, as claimed by Pouscoulous (2011), children's linguistic competence, world knowledge and cultural expertise may account for the improvement in their figurative language abilities with age – not age per se. Thus, the association between language and figurative language does not seem to merely reflect the influence of age.

The size of the gap between the ASD and TD groups was the same in studies conducted in English and in those conducted in other languages, which suggests that, in the included studies, seemingly no cross-linguistic differences exist with regard to the figurative language comprehension of individuals with ASD. This finding conflicts with some findings in the cognitive linguistic literature, which discusses the variation of metaphor across languages and cultures in a TD context. In particular, evidence has shown that comprehension of some aspects of, for example, metaphor may vary between languages (e.g. Özçalışkan, 2003). More studies in different languages will be beneficial to understand figurative language comprehension in individuals with ASD from a cross-linguistic perspective.

Methodological issues

Some limitations observed in the included studies must be considered when interpreting the findings of this review. In a number of earlier studies, the main methodological shortcoming involved not performing appropriate tests to control for language comprehension in individuals with ASD. In addition, the wide age range, significant intergroup differences in language skills (e.g. Rundblad and Annaz, 2010b) and the lack of reading skill assessments when the task is administered in written form could have a considerable impact on the participants' performances on figurative language tasks.

How the studies operationalize the figurative language trope that they examine is often unclear, as are hypotheses based on theories of figurative language comprehension. This may drawback results in tasks that measure different aspects of figurative language, which require different pragmatic processes, thereby making it difficult to summarize the figurative language comprehension deficits in individuals with ASD.

However, in more recent studies, the methodology has improved. For instance, when examining figurative language comprehension, more studies use language measures that are more advanced than VIQ and vocabulary tests. Another example of recent methodological improvements involves including one ASD group and two TD control groups – one matched based on CA and another matched based on LA. This approach is useful in examining whether figurative LA is developmentally delayed or deviant in individuals with ASD.

Moreover, several studies used analysis of variance (ANOVA) and regression analyses (e.g. Norbury, 2004), and few studies used cross-sectional developmental trajectory analysis, which is based on developmental trajectories or growth models (Thomas et al., 2009). The latter is seemingly a well-justified alternative for group matching in ASD research because it does not require a priori group matching, which has been shown to be challenging for individuals with ASD (Jarrold and Brock, 2004).

Limitations of this meta-analysis

Meta-analyses are generally criticized for ignoring important between-study differences by creating a summary of the outcomes (Borenstein et al., 2009). However, the potential sources of between-study differences may be formally addressed by examining potential predictor variables. Unfortunately, one actual predictor that could have reliably explained between-study differences, ToM, could not be included in the meta-regression analysis because of the insufficient number of studies.

Collapsing multiple outcomes in the studies and creating a composite score to avoid threats related to using dependent data resulted in the following limitation: in the studies that included dependent subgroups (one target group and two control groups (e.g. Whyte et al., 2014) or one control group and two ASD groups (Norbury, 2004), we combined the outcome measures to ensure that dependent data contributed to one effect size only. The main point of the methodologies used in these studies was initially to show that, when a group of individuals with ASD is compared with an LA-matched TD group, the group difference is generally small and nonsignificant. By combining the outcomes into one composite score, these important differences are not highlighted in this meta-analysis. The meta-analysis instead asks whether figurative language deficits are specific to individuals with ASD relative to TD individuals in general. However, we addressed this issue by narratively summarizing and discussing the findings of the three studies that included dependent subgroups.

Implications for practice and research

Given that a subset of individuals with ASD display greater deficits compared with CA-matched TD controls in comprehending various figurative language tropes, these difficulties should be addressed in clinical and educational settings.

It is important that teachers, parents and clinicians do not avoid using figurative language in their interactions with children and adolescents with ASD, as figurative language frequently occurs in daily life settings. Instead, parents, teachers and clinicians should use and monitor figurative language with great awareness, provide explanations when required, and reinforce comprehension across settings (see also Kerbel and Grunwell, 1997, for similar recommendations with respect to idioms).

Because the deficits in figurative language appear to be related to core language skills, improving core language skills through educational and clinical interventions may lead to improved figurative language comprehension.

Given that controlling for language comprehension and vocabulary and VIQ measures is crucial to obtain valid results in figurative language comprehension studies on individuals with ASD (Gernsbacher and Pripas-Kapit, 2012), future studies on individuals with ASD should control for general language comprehension to eliminate the impact of this covariate when studying figurative language. Likewise, more studies are needed to investigate the role of syntax in figurative language comprehension in individuals with ASD.

More research is needed to examine the independent contributions of core language skills and ToM in figurative language comprehension. It is important that studies examining the role of ToM use developmentally sensitive and appropriate ToM tasks, preferably ones that are not highly correlated with verbal ability (Whyte and Nelson, 2015).

The causal relationships between, for example, ToM and LA and figurative language comprehension can be established by explicitly targeting figurative language skills through training studies. A few interventions that are designed for figurative language comprehension have been shown to be successful in children with ASD (e.g. Mashal and Kasirer, 2011; Persicke et al., 2012; Whyte et al., 2013). Figurative language skill training would also benefit adults and adolescents with ASD.

The quality of future studies should be optimized by considering the challenges related to small samples as well as those with wide age ranges. Cross-sectional developmental trajectory analysis is one possible approach to overcome the latter challenge (Thomas et al., 2009). However, the most informative way to study the development of figurative language comprehension, is to longitudinally examine the development of the same participants over time. Notably, no longitudinal studies were found in our systematic literature search.

Although we conclude that group matching strategy and trope differences could explain some of the variance in effect size, a large proportion of the variance has yet to be explained. Several critical variables, such as task content, task demands and task administration, deserve more careful consideration in future research.

Conclusion

In conclusion, individuals with ASD show deficits in figurative language comprehension compared with their TD controls. The significant differences between the studies can reliably be explained by trope differences, indicating that different tropes require different comprehension processes in individuals with ASD. Moreover,

the group matching strategy was a reliable predictor of figurative language comprehension. Namely, the studies that matched groups based on LA yielded small and nonsignificant effect sizes, indicating that figurative language comprehension in individuals with ASD is closely related to core language abilities but simultaneously disputing the view of the uniqueness and universality of figurative language comprehension deficits in individuals with ASD. Therefore, interventions and educational programmes that aim to improve the social communication skills of individuals with ASD should target core language skills in addition to social skills.

References

References marked with an asterisk denote studies included in the meta-analysis.

Ackerman BP (1982) On comprehending idioms: do children get the picture? *Journal of Experimental Child Psychology* 33(3): 439–454.

*Adachi T, Koeda T, Hirabayashi S, et al. (2004) The metaphor and sarcasm scenario test: a new instrument to help differentiate high functioning pervasive developmental disorder from attention deficit/hyperactivity disorder. *Brain & Development* 26(5): 301–306.

American Psychiatric Association (APA) (2013) *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. 5th ed. Arlington, VA: APA.

*Au-Yeung SK, Kaakinen JK, Liversedge SP, et al. (2015) Processing of written irony in autism spectrum disorder: an eye-movement study. *Autism Research* 8: 749–760.

Baron-Cohen S (2001) Theory of mind in normal development and autism. *Prisme* 34: 174–183.

Baron-Cohen S, Leslie AM and Frith U (1985) Does the autistic child have a 'theory of mind'. *Cognition* 21(1): 37–46.

Baron-Cohen S, Wheelwright S, Hill J, et al. (2001) The 'reading the mind in the eyes' test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry* 42(2): 241–251.

Borenstein M, Hedges LV, Higgins JPT, et al. (2009) *Introduction to Meta-Analysis*. Chichester: John Wiley & Sons.

Brynskov C, Krøjgaard P and Eigsti I (2016) Language and communication in children with autism: do research and clinical practice converge in Denmark? *Nordic Psychology* 68(2): 100–113.

Cain K, Oakhill J and Lemmon K (2005) The relation between children's reading comprehension level and their comprehension of idioms. *Journal of Experimental Child Psychology* 90(1): 65–87.

Carrow-Woolfolk E (1999) *Comprehensive Assessment of Spoken Language*. Circle Pines, MN: AGS Publishing.

*Channon S, Crawford S, Orlowska D, et al. (2014) Mentalising and social problem solving in adults with Asperger's syndrome. *Cognitive Neuropsychiatry* 19(2): 149–163.

*Chouinard B and Cummine J (2016) All the world's a stage: evaluation of two stages of metaphor comprehension in people with autism spectrum disorder. *Research in Autism Spectrum Disorders* 23: 107–121.

Cohen J (1968) Multiple regression as a general data-analytic system. *Psychological Bulletin* 70(6): 426–443.

*Colich NL, Wang A, Rudie JD, et al. (2012) Atypical neural processing of ironic and sincere remarks in children and adolescents with autism spectrum disorders. *Metaphor and Symbol* 27(1): 70–92.

Colston HL and Gibbs RW (2002) Are irony and metaphor understood differently? *Metaphor and Symbol* 17(1): 57–60.

Colston HL and Kuiper MS (2002) Figurative language development research and popular children's literature: why we should know, 'where the wild things are'. *Metaphor and Symbol* 17(1): 27–43.

Conti-Ramsden G, Botting N and Faragher B (2001) Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry and Allied Disciplines* 42: 741–748.

*De Villiers PA, de Villiers JG, Diaz S, et al. (2011) Non-literal language and theory of mind in autism spectrum disorders. In: *Poster presented at the American speech-language hearing association convention*, San Diego, CA, 16–19 November.

*Dennis M, Lazenby AL and Lockyer L (2001) Inferential language in high-function children with autism. *Journal of Autism and Developmental Disorders* 31(1): 47–54.

Dews S, Winner E, Kaplan J, et al. (1996) Children's understanding of the meaning and functions of verbal irony. *Child Development* 67: 3071–3085.

Dunn LM, Dunn LM, Whetton C, et al. (1997) *The British Picture Vocabulary Scale*. 2nd ed. Windsor: NFER-Nelson.

Eigsti I, de Marchena AB, Schuh JM, et al. (2011) Language acquisition in autism spectrum disorders: a developmental review. *Research in Autism Spectrum Disorders* 5(2): 681–691.

Falkum IL, Recasens M and Clark EV (2016) 'The moustache sits down first': on the acquisition of metonymy. *Journal of Child Language* 19: 1–33.

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

Gibbs RW (2000) Irony in talk among friends. *Metaphor and Symbol* 15(1–2): 5–27.

Gibbs RW and Colston HL (2012) *Interpreting Figurative Meaning*. Cambridge: Cambridge University Press.

*Glenwright M and Agbayewa AS (2012) Older children and adolescents with high-functioning autism spectrum disorders can comprehend verbal irony in computer-mediated communication. *Research in Autism Spectrum Disorders* 6(2): 628–638.

Gluksberg S (2001) *Understanding Figurative Language: From Metaphor to Idioms* (Oxford Psychology Series). New York: Oxford University Press.

*Gold R and Faust M (2010) Right hemisphere dysfunction and metaphor comprehension in young adults with Asperger syndrome. *Journal of Autism and Developmental Disorders* 40(7): 800–811.

*Gunter HL, Ghaziuddin M and Ellis HD (2002) Asperger syndrome: tests of right hemisphere functioning and interhemispheric communication. *Journal of Autism and Developmental Disorders* 32(4): 263–281.

Hancock JT, Dunham PJ and Purdy K (2000) Children's comprehension of critical and complimentary forms of verbal irony. *Journal of Cognition and Development* 1(2): 227–248.

Happé FG (1993) Communicative competence and theory of mind in autism: a test of relevance theory. *Cognition* 48(2): 101–119.

Happé FG (1994) An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders* 24(2): 129–154.

Hedges LV (1981) Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics* 6(2): 107–128.

Hedges LV and Olkin I (1985) *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.

*Hermann I, Haser V, van Elst LT, et al. (2013) Automatic metaphor processing in adults with Asperger syndrome: a metaphor interference effect task. *European Archives of Psychiatry and Clinical Neurosciences* 263 (Suppl. 2): S177–S187.

Hobson RP (2012) Autism, literal language and concrete thinking: some developmental considerations. *Metaphor and Symbol* 27(1): 4–21.

*Huang S, Oi M and Taguchi A (2015) Comprehension of figurative language in Taiwanese children with autism: the role of theory of mind and receptive vocabulary. *Clinical Linguistics & Phonetics* 29(8–10): 764–775.

*Imaizumi S, Furuya I and Yamasaki K (2009) Voice as a tool communicating intentions. *Logopedics, Phoniatrics, Vocology* 34(4): 196–199.

Jarrold C and Brock J (2004) To match or not to match? Methodological issues in autism-related research. *Journal of Autism and Developmental Disorders* 34(1): 81–86.

*Kaland N, Møller-Nielsen A, Callesen K, et al. (2002) A new 'advanced' test of theory of mind: evidence from children and adolescents with Asperger syndrome. *Journal of Child Psychology and Psychiatry* 43(4): 517–528.

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

Kerbel D and Grunwell P (1997) Idioms in the classroom: an investigation of language unit and mainstream teachers' use of idioms. *Child Language Teaching & Therapy* 13(2): 113–123.

Lakens D (2013) Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology* 4: 863.

Lakoff G and Johnson M (2003) *Metaphors We Live by*. Chicago, IL: University of Chicago Press.

*Landa RJ and Goldberg MC (2005) Language, social, and executive functions in high functioning autism: a continuum of performance. *Journal of Autism and Developmental Disorders* 35(5): 557–573.

Lee RF and Kamhi AG (1990) Metaphoric competence in children with learning disabilities. *Journal of Learning Disabilities* 23(8): 476–482.

*Lee SB, Song SH, Ham JH, et al. (2015) Idiom comprehension deficits in high-functioning autism spectrum disorder using a Korean autism social language task. *Yonsei Medical Journal* 56(6): 1613–1618.

*Lewis FM, Murdoch BE and Woodyatt GC (2007) Communicative competence and metalinguistic ability: performance by children and adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 37(8): 1525–1538.

*Li JP, Law T, Lam GY, et al. (2013) Role of sentence-final particles and prosody in irony comprehension in Cantonese-speaking children with and without autism spectrum disorders. *Clinical Linguistics & Phonetics* 27(1): 18–32.

Lyons V and Fitzgerald M (2004) Humor in autism and Asperger syndrome. *Journal of Autism and Developmental Disorders* 34(5): 521–531.

*McCrimmon AW, Schwean VL, Saklofske DH, et al. (2012) Executive functions in Asperger's syndrome: an empirical investigation of verbal and nonverbal skills. *Research in Autism Spectrum Disorders* 6: 224–233.

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

*Martin I and McDonald S (2004) An exploration of causes of non-literal language problems in individuals with Asperger syndrome. *Journal of Autism and Developmental Disorders* 34(3): 311–328.

*Mashal N and Kasirer A (2011) Thinking maps enhance metaphoric competence in children with autism and learning disabilities. *Research in Developmental Disabilities* 32(6): 2045–2054.

*Mathersul D, McDonald S and Rushby JA (2013) Understanding advanced theory of mind and empathy in high-functioning adults with autism spectrum disorder. *Journal of Clinical and Experimental Neuropsychology* 35(6): 655–668.

Melogno S, D'Ardia C, Pinto MA, et al. (2012) Explaining metaphors in high-functioning autism spectrum disorder children: a brief report. *Research in Autism Spectrum Disorders* 6(2): 683–689.

*Minshew NJ, Goldstein G and Siegel DJ (1995) Speech and language in high-functioning autistic individuals. *Neuropsychology* 9(2): 255–261.

Næss KA, Lyster SA, Hulme C, et al. (2011) Language and verbal short-term memory skills in children with down syndrome: a meta-analytic review. *Research in Developmental Disabilities* 32(6): 2225–2234.

Nelson HE and Wilson JR (1991) *The Revised National Adult Reading Test*. Windsor: NFER-Nelson.

Newcomer PL and Hammill DD (1997) *Test of Language Development, Primary*. 3rd ed. Austin, TX: Pro-Ed.

Nippold MA (1991) Evaluating and enhancing idiom comprehension in language-disordered students. *Language, Speech, and Hearing Services in Schools* 22(3): 100–106.

*Norbury CF (2004) Factors supporting idiom comprehension in children with communication disorders. *Journal of Speech, Language, and Hearing Research* 47(5): 1179–1193.

Norbury CF (2005) The relationship between theory of mind and metaphor: evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology* 23(3): 383–399.

*Olofson EL, Casey D, Oluyedun OA, et al. (2014) Youth with autism spectrum disorder comprehend lexicalized and novel primary conceptual metaphors. *Journal of Autism and Developmental Disorders* 44(10): 2568–2583.

Özçalışkan S (2003) Metaphorical motion in crosslinguistic perspective: a comparison of English and Turkish. *Metaphor and Symbol* 18(3): 189–228.

*Ozonoff S and Miller JN (1996) An exploration of right-hemisphere contributions to the pragmatic impairments of autism. *Brain and Language* 52(3): 411–434.

Papagno C and Caporali A (2007) Testing idiom comprehension in aphasic patients: the effects of task and idiom type. *Brain and Language* 100(2): 208–220.

Passanisi A and Di Nuovo S (2015) Social and pragmatic impairments in individuals with autism spectrum disorder. A lack of theory of mind? *Life Span and Disability* 18(1): 75–99.

Persicke A, Tarbox J, Ranick J, et al. (2012) Establishing metaphorical reasoning in children with autism. *Research in Autism Spectrum Disorders* 6(2): 913–920.

*Peterson CC, Wellman HM and Slaughter V (2012) The mind behind the message: advancing theory-of-mind scales for typically developing children, and those with deafness, autism, or Asperger syndrome. *Child Development* 83(2): 469–485.

Pexman PM (2008) It's fascinating research: the cognition of verbal irony. *Current Directions in Psychological Science* 17(4): 286–290.

Pexman PM and Glenwright M (2007) How do typically developing children grasp the meaning of verbal irony? *Journal of Neurolinguistics* 20(2): 178–196.

Pexman PM, Glenwright M, Krol A, et al. (2005) An acquired taste: children's perceptions of humor and teasing in verbal irony. *Discourse Processes* 40(3): 259–288.

*Pexman PM, Rostad KR, McMorris CA, et al. (2011) Processing of ironic language in children with high-functioning autism spectrum disorder. *Journal of Autism and Developmental Disorders* 41(8): 1097–1112.

Pons F, de Rosnay M, Harris P, et al. (2009) Theory-of-mind and language in children. *Impuls Tidsskrift for Psykologi* 3: 30–34.

Pouscoulous N (2011) Metaphor: for adults only? *Belgian Journal of Linguistics* 25: 51–79.

Pouscoulous N (2014) The elevator's buttocks metaphorical abilities in children. In: Matthews D (ed.) *Pragmatic Development in First Language Acquisition*. Amsterdam: John Benjamins, pp.239–259.

*Rajendran G, Mitchell P and Rickards H (2005) How do individuals with Asperger syndrome respond to nonliteral language and inappropriate requests in computer-mediated communication? *Journal of Autism and Developmental Disorders* 35(4): 429–443.

Rapp AM and Wild B (2011) Nonliteral language in Alzheimer dementia: a review. *Journal of the International Neuropsychological Society* 17(2): 207–218.

Rundblad G and Annaz D (2010a) Development of metaphor and metonymy comprehension: receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology* 28(3): 547–563.

*Rundblad G and Annaz D (2010b) The atypical development of metaphor and metonymy comprehension in children with autism. *Autism* 14(1): 29–46.

Saban-Bezalel R and Mashal N (2015) Hemispheric processing of idioms and irony in adults with and without pervasive developmental disorder. *Journal of Autism and Developmental Disorders* 45(11): 3496–3508.

Samson AC (2013) Humor(lessness) elucidated – sense of humor in individuals with autism spectrum disorders: review and introduction. *Humor* 26(3): 393–409.

Scammacca N, Roberts G and Stuebing KK (2014) Meta-analysis with complex research designs: dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research* 84(3): 328–364.

*Scheeren AM, de Rosnay M, Koot HM, et al. (2013) Rethinking theory of mind in high-functioning autism spectrum disorder. *Journal of Child Psychology and Psychiatry* 54(6): 628–635.

Semel EM, Wiig EH and Secord W (2000) *Clinical Evaluation of Language Fundamentals*. 3rd ed. London: The Psychological Corporation.

Semrud-Clikeman M and Glass K (2010) The relation of humor and child development: social, adaptive, and emotional aspects. *Journal of Child Neurology* 25(10): 1248–1260.

Sperber D and Wilson D (1986) *Relevance: Communication and Cognition*. Oxford: Blackwell Publishing House.

*Strandburg RJ, Marsh JT, Brown WS, et al. (1993) Event-related potentials in high-functioning adult autistics: linguistic and nonlinguistic visual information processing tasks. *Neuropsychologia* 31(5): 413–434.

Swineford LB, Thurm A, Baird G, et al. (2014) Social (pragmatic) communication disorder: a research review of this new DSM-5 diagnostic category. *Journal of Neurodevelopmental Disorders* 6(1): 41.

Tager-Flusberg H (2006) Defining language phenotypes in autism. *Clinical Neuroscience Research* 6(3–4): 219–224.

Tager-Flusberg H and Joseph RM (2003) Identifying neurocognitive phenotypes in autism. *Philosophical Transactions of the Royal Society B: Biological Sciences* 358(1430): 303–314.

Thomas MS, Annaz D, Ansari D, et al. (2009) Using developmental trajectories to understand developmental disorders. *Journal of Speech, Language, and Hearing Research* 52(2): 336–358.

Vulchanova M, Saldaña D, Chahboun S, et al. (2015) Figurative language processing in atypical populations: the ASD perspective. *Frontiers in Human Neuroscience* 9: 24.

*Wang AT, Lee SS, Sigman M, et al. (2006) Neural basis of irony comprehension in children with autism: the role of prosody and context. *Brain* 129(4): 932–943.

Whyte EM and Nelson KE (2015) Trajectories of pragmatic and nonliteral language development in children with autism spectrum disorders. *Journal of Communication Disorders* 54: 2–14.

Whyte EM, Nelson KE and Khan KS (2013) Learning of idiomatic language expressions in a group intervention for children with autism. *Autism* 17(4): 449–464.

*Whyte EM, Nelson KE and Scherf KS (2014) Idiom, syntax, and advanced theory of mind abilities in children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research* 57(1): 120–130.

*Williams DL, Cherkassky VL, Mason RA, et al. (2013) Brain function differences in language processing in children and adults with autism. *Autism Research* 6(4): 288–302.

Wilson D and Sperber D (2012) *Meaning and Relevance*. Cambridge: Cambridge University Press.

Winner E (1988) *The Point of Words: Children's Understanding of Metaphor and Irony*. Cambridge, MA: Harvard University Press.

Winner E, Rosenstiel AK and Gardner H (1976) The development of metaphoric understanding. *Developmental Psychology* 12(4): 289–297.

*Wu C, Tseng L, An C, et al. (2014) Do individuals with autism lack a sense of humor? A study of humor comprehension, appreciation, and styles among high school students with autism. *Research in Autism Spectrum Disorders* 8(10): 1386–1393.

*Zalla T, Amsellem F, Chaste P, et al. (2014) Individuals with autism spectrum disorders do not use social stereotypes in irony comprehension. *PLoS ONE* 9(4): e95568.

*Zheng Q, Jia Z and Liang D (2015) Metaphor and metonymy comprehension in Chinese-speaking children with high-functioning autism. *Research in Autism Spectrum Disorders* 10: 51–68.

Appendix 1. Characteristics of the included studies.

Study Author (year)	N ASD (TD)	Mean CA ASD (TD)	Trope	Effect size (g)	95% CI	Equating strategy
*Adachi et al. (2004)	54 (199)	118 (120)	Metaphor; Sarcasm (Combined)	−0.45	[−0.75, −0.15]	No difference in CA and VIQ
*Au-Yeung et al. (2015)	22 (20)	389 (286)	Irony	−0.51	[−1.11, 0.09]	Matched based on VIQ
*Channon et al. (2014)	21 (21)	480 (524)	Sarcasm	−0.83	[−1.45, −0.21]	Matched based on CA
*Chouinard and Cummine (2016)	13 (12)	401 (396)	Metaphor	−0.45	[−1.21, 0.31]	Matched based on CA and semantic knowledge
*Colich et al. (2012)	15 (15)	171 (158)	Irony	0.55	[−0.15, 1.26]	Matched based on CA and VIQ
Dennis et al. (2001)	8 (8)	119 (113)	Metaphor/idiom	−1.37	[−2.42, −0.33]	Matched based on CA
*De Villiers et al. (2011)	30 (28)	149 (151)	Irony Metaphor (Combined)	−0.82	[−1.35, −0.29]	Matched based on CA and VIQ
*Glenwright and Agbayewa (2012)	14 (14)	148 (139)	Ironic criticism	0.14	[−0.58, 0.86]	Matched based on verbal mental age and CA
*Gold and Faust (2010)	27 (36)	275 (296)	LVF/RH Conventional metaphor LVF/RH Novel metaphor RVH/LH Conventional metaphor RVH/LH Novel metaphor (Combined)	−0.52	[−1.02, −0.01]	Matched based on CA and VIQ
*Gunter et al. (2002)	8 (8)	195 (203)	Humour Written metaphor Novel metaphor (Combined)	−1.35	[−2.40, −0.30]	Matched based on CA and VIQ
*Hermann et al. (2013)	20 (20)	509 (421)	Metaphor	−0.38	[−1.00, 0.22]	Matched based on vocabulary
*Huang et al. (2015)	50 (50)	122 (127)	Irony Metaphor Sarcasm (Combined)	−0.52	[−0.91, −0.12]	Matched based on CA and vocabulary
*Imaizumi et al. (2009)	20 (24)	119 (114)	Sarcasm	−1.85	[−2.55, −1.15]	Matched based on CA
*Kaland et al. (2002)	21 (20)	189 (186)	Figure of speech Irony (Combined)	−1.26	[−2.27, −0.26]	Matched based on CA
*Kasirer and Mashal (2014)	17 (17)	253 (273)	Conventional metaphor Novel metaphor (Combined)	−0.47	[−1.14, 0.20]	Matched based on CA
Landa and Goldberg (2005)	19 (19)	132 (132)	Metaphoric expressions and figures of speech	−1.01	[−1.68, −0.35]	Matched based on CA and VIQ
Lee et al. (2015)	16 (10)	111,72 (111,6)	Matched idiom task Mismatched idiom task (Combined)	−1.04	[−1.86, −0.23]	Matched based on age and IQ
Lewis et al. (2007) (Adults)	17 (13)	418 (416)	Figurative language	−1.09	[−1.85, −0.33]	Matched based on CA
Lewis et al. (2007) (Children)	20 (18)	139 (138)	Figurative language	0.32	[−0.30, 0.95]	Matched based on CA
*Li et al. (2013)	13 (13) 12 (12)	125 (125)	Irony belief Irony intention (Combined)	0.16	[−0.64, 0.84]	Matched based on LA and CA

Appendix 1. (Continued)

Study	N	Mean CA	Trope	Effect size (g)	95% CI	Equating strategy
Author (year)	ASD (TD)	ASD (TD)				
MacKay and Shaw (2004)	19 (21)	116 (123)	Hyperbole; indirect request; irony; metonymy; rhetorical questions; understatement (Combined)	−0.95	[−1.66, −0.23]	Matched based on CA and LA
*Martin and McDonald (2004)	14 (24)	236 (237)	Irony	−1.09	[−1.78, −0.40]	Matched based on CA
Mashal and Kasirer (2011)	20 (20)	156	Idiom Metaphor (Combined)	−0.73	[−1.36, −0.10]	Matched based on CA and LA
*Mathersul et al. (2013)	40 (33)	446 (500)	Sarcasm	−0.67	[−1.14, −0.20]	Matched based on CA and vocabulary
McCrimmon et al. (2012) (Cluster A)	24 (12)	222 (222)	Proverb	0.66	[−0.02, 1.36]	Matched based on CA and VIQ
McCrimmon et al. (2012) (Cluster B)	9 (21)	222 (222)	Proverb	1.25	[0.42, 2.07]	Matched based on CA
*Minshew et al. (1995)	62 (50)	213 (203)	Metaphoric expressions	−1.26	[−1.66, −0.85]	Matched based on CA and VIQ
Norbury (2004)			Idiom TD vs ASL Idiom TD vs ASO (Combined)	−1.16	[−1.70, −0.62]	Matched based on CA
*Olofson et al. (2014)	13 (13)	155 (153)	Conventional metaphor Novel metaphor (Combined)	−0.90	[−1.69, −0.12]	Matched based on CA
Ozonoff and Miller (1996)	17 (17)	314 (287)	Humour	−0.91	[−1.60, −0.22]	Matched based on CA and VIQ
*Pexman et al. (2011)	18 (18)	132 (132) 132 (94)	Ironic compliment ASD vs TD CAM Ironic compliment ASD vs TD LAM Ironic criticism ASD vs TD CAM Ironic criticism ASD vs TD LAM (Combined)	0.05	[−0.59, 0.69]	Matched based on LA
*Peterson et al. (2012)	44 (29)	108 (105)	Sarcasm	−0.39	[−0.86, 0.07]	Matched based on CA
*Rundblad and Annaz (2010b)	11 (17)	101 (100)	Metaphor Metonymy (Combined)	−1.32	[−2.13, −0.51]	Matched based on CA. Significant group difference in verbal skills
*Rajendran et al. (2005)	9 (12)	198 (201)	Figure of speech Sarcasm (Combined)	−0.30	[−1.14, 0.52]	Matched based on CA and VIQ
Saban-Bezalel and Mashal (2015)	23 (24)	316 (327)	Idiom Irony (Combined)	−0.65	[−1.23, 0.07]	Matched based on CA and VIQ
*Scheeren et al. (2013) (Adolescents)	84 (16)	184 (172)	Sarcasm	0.30	[−0.11, 0.72]	Matched based on receptive IQ (significantly older ASD group)
*Scheeren et al. (2013) (Children)	19 (7)	122 (114)	Sarcasm	0.16	[−0.39, 0.72]	
Strandburg et al. (1993)	13 (13)	299 (314)	Idiom	−1.12	[−1.92, −0.31]	Matched based on CA and VIQ

Appendix 1. (Continued)

Study	N	Mean CA	Trope	Effect size (g)	95% CI	Equating strategy
Author (year)	ASD (TD)	ASD (TD)				
Wang et al. (2006)	18 (18)	143 (143)	Idiom	−0.88	[−1.56, −0.21]	Matched based on CA and VIQ
Whyte et al. (2014)		116 (114)	Idiom ASD vs CAM	0.50	[−1.04, 0.04]	Matched based on CA and LA
		116 (114)	Idiom ASD vs LAM (Combined)			
*Williams et al. (2013) (Adults)	13 (12)	299 (252)	Irony	−0.91	[−1.71, −0.11]	Matched based on CA and VIQ
*Williams et al. (2013) (Children)	15 (14)	156 (150)	Irony	−0.48	[1.20, 0.23]	Matched based on CA and VIQ
Wu et al. (2014)	164 (164)	165 (165)	Incongruity Nonsense joke (Combined)	−0.52	[−0.74, −0.30]	Matched based on CA and IQ
*Zalla et al. (2014)	17 (17)	328 (361)	Irony	−0.58	[−1.25, 0.08]	Matched based on CA and VIQ
*Zheng et al. (2015)	15 (15)	78 (75)	Conventional metaphor Conventional metonyms Novel metaphors Novel metonyms (Combined)	−0.75	[−1.48, −0.03]	Matched based on CA and VIQ

ASD: autism spectrum disorder; ASL: ASD with language impairment; ASO: ASD only; CA: chronological age; CI: confidence interval; LA: language ability; LH: left hemisphere; LVF: left visual field; RH: right hemisphere; RVF: right visual field; TD: typically developing.
References marked with an asterisk denote studies included in subgroup analysis of metaphor and irony and sarcasm.

Paper III:

Kalandadze, T., Bambini, V., & Næss K.-A.B. (submitted). 'A Systematic Review and Meta-analysis of Studies on Metaphor Comprehension in Individuals with Autism Spectrum Disorder: Do Task Properties Matter?'

III

A Systematic Review and Meta-analysis of Studies on Metaphor Comprehension in Individuals with Autism Spectrum Disorder: Do Task Properties Matter?

Tamar Kalandadze[1,2], Valentina Bambini[3], Kari-Anne B.Næss[1]

[1] University of Oslo, Norway

[2] Knowledge Centre for Education, The Research Council of Norway

[3] University School for Advanced Studies IUSS Pavia, Italy

Correspondence concerning this article should be addressed to

Knowledge Centre for Education, The Research Council of Norway,

P.O Box 564 N-1327 Lysaker, Norway

Email: tamar.kalandadze@isp.uio.no/tka@rcn.no

Abstract

Individuals with autism spectrum disorder (ASD) often experience difficulty in comprehending metaphors compared to individuals with typical development (TD). However, there is a large variation in the results across studies, possibly related to the properties of the metaphor tasks. This pre-registered systematic review and meta-analysis (a) explored the properties of the metaphor tasks used in ASD research, and (b) investigated the group difference between individuals with ASD and TD on metaphor comprehension, as well as the relationship between the task properties and any between-study variation. A systematic search was undertaken in seven relevant databases. Fourteen studies fulfilled our predetermined inclusion criteria. Across tasks, we detected four types of response format and a great variety of metaphors in terms of familiarity, syntactic structure, and linguistic context. Individuals with TD outperformed individuals with ASD on metaphor comprehension (Hedge's g = -0.63). Verbal explanation response format was utilized in the study showing the largest effect size in the group comparisons. However, due to the sparse experimental manipulations, the role of task properties could not be established. Future studies should consider and report task properties to determine their role in metaphor comprehension, and to inform experimental paradigms as well as educational assessment.

Keywords: Autism Spectrum Disorder, figurative language, response format, experimental pragmatics

Introduction

A metaphor is a paradigmatic type of figurative language involving discrepancy between the encoded, 'literal' meaning of words, and their occasion-specific use (Camp, 2009; Carston, 2010). Metaphors can appear in many forms, such as "Sally is a chameleon", or "Your theory is falling apart". Accordingly, different accounts of metaphor comprehension have been proposed (see Bowdle & Gentner, 2005; Gibbs, 2011; Gibbs & Tendahl, 2006; Gluksberg, 2001; Wilson, 2011). Among them, pragmatic accounts (e.g., Relevance Theory) focus on metaphor in communication, highlighting the inferential mechanisms that lead to adjusting the linguistically encoded concepts to arrive at the speaker's intended meaning (Sperber & Wilson, 2008). For instance, in "Sally is a chameleon", the adjustment results in the broadening of the concept CHAMELEON to include not only a species of lizard but also individuals with certain psychological features (Carston, 2012). On the other hand, cognitive linguistics accounts (e.g., Conceptual Metaphor Theory) emphasize the role of metaphor in thought, seeing it in terms of conceptual mappings across cognitive domains (Gibbs, 2011; Lakoff & Johnson, 1980). The conceptual mappings emerge in our metaphorical use of language, as in "Your theory is falling apart" for the mapping of theories onto physical constructs such as buildings (THEORIES ARE BULDINGS).

Regardless of the theoretical approach, there is an agreement that metaphors are a ubiquitous part of language and appear frequently in everyday communication, school-books, academic texts, literature, and media communications (Golden, 2010; Steen, Dorst, & Hermann, 2010). Hence, difficulty in understanding metaphors may impede social communication, ability to obtain information as well as academic attainment.

In individuals with typical development (TD) metaphor comprehension skills mature throughout childhood till adolescence, and it is commonly assumed that the age of 10 represent a crucial moment (Lecce et al., 2019; Winner et al., 1976). There is, however, also awareness that metaphorical competence is evident earlier – if assessed with age-appropriate tasks (Poscoulous, 2011; 2014; Vosniadou, Ortony, Reynolds, & Wilson, 1984). In contrast, profound and lasting difficulty in metaphor comprehension has traditionally been considered characteristic for individuals with autism spectrum disorder (ASD; Adachi et al., 2004; Happé, 1993; Rundblad & Annaz, 2010a), a neurodevelopmental condition characterized by impairments in social communication and interaction, as well as restricted and stereotyped behaviours (American Psychiatric Association, 2013). In particular, individuals with ASD have been reported to interpret metaphors literally (Happé, 1993), a phenomenon referred to as the 'literality bias' or concretism (see Rossetti, Brambilla, & Papagno, 2018, for explanation of these terms).

However, there is a discrepancy in study findings. For example, several studies show no statistically significant difference between ASD and TD groups in figurative language comprehension, including metaphors (Hermann et al., 2013; Kasirer & Mashal, 2014; Mashal & Kasirer, 2011; Norbury, 2005). These findings indicate that variables other than characteristics intrinsic to ASD may explain the variation in results across studies. Indeed, group matching strategy and general language ability have previously been found to explain some of the between-study variance in figurative language comprehension (see Kalandadze, Norbury, Nærland, & Næss, 2016 for a review). Yet, the remaining unexplained variance requires an investigation of additional relevant variables.

In the behavioural and neurological literature in TD and clinical populations there is an agreement that the ability to understand metaphors hinges on the task properties such as response format (i.e., multiple-choice vs. verbal explanation task), or absence of linguistic

context (see Pouscoulous, 2011; 2014 for discussion of studies with TD participants, and Rosetti et al., 2018 for discussion of literature on schizophrenia). For instance, children with TD show earlier competence in metaphor comprehension when tested with act-out rather than verbal explanation task, perhaps due to the differences in linguistic and cognitive demands that verbal and other types of tasks pose (Pouscoulous, 2011). Similarly, response format could explain how individuals with ASD perform on metaphor tasks. For example, individuals with ASD might understand metaphors comparably to individuals with TD but have more difficulties in explaining the meaning verbally due to difficulties with expressive language (Kwok, Brown, Smyth, & Oram Cardy, 2015). The same might be true for other properties of the metaphors, e.g., the amount and type of context available to interpret the expression, or the familiarity of the expression (Pouscoulous, 2011; 2014).

Despite this knowledge, the properties of metaphor comprehension assessment tasks in studies that compare individuals with ASD to individuals with TD have yet to be comprehensively and systematically explored. Neither the potential inter-relationships between the task properties and any between-study variation have been systematically investigated. Reviews that have been conducted focused on ASD and figurative language in general, rather than on metaphor specifically (Gernsbacher & Pripas-Kapit, 2012; Kalandadze et al., 2016; Melogno, Pinto, & Levi, 2012; Vulchanova et el., 2015). Importantly, the comprehension of metaphor might differ from the comprehension of other figurative language types in several respects (Vulchanova et al., 2019). Moreover, the majority of the existing reviews utilized a narrative approach (Gernsbacher & Pripas-Kapit, 2012; Melogno et al., 2012; Vulchanova et al., 2015), which differs from our systematic approach in fundamental ways, especially regarding transparency and systematicity of methods used (Borenstein et al., 2009).

Here, we provide a novel and thorough systematic review and meta-analysis of the properties of the metaphor tasks used in ASD research. We quantitatively compared performance on metaphor comprehension tasks between groups of individuals with ASD and TD and investigated the potential role of the task properties in between-study variation.

By systematically summarizing and synthesizing the available research in the field fulfilling certain inclusion criteria, our study provides robust results which will ultimately have implications when designing future figurative language/metaphor comprehension research, for advancing assessment practices as well as for guiding the research-based intervention paradigms for individuals with ASD.

The following section includes an overview of metaphor task properties that have been identified as critical for metaphor comprehension in TD and clinical populations (e.g. Pouscoulous, 2011; Rosetti et al., 2018). These are (1) response format (e.g. multiple-choice, meaningfulness decision, etc.), and (2) linguistic characteristics (metaphor familiarity, syntactic structure of the metaphor, linguistic context, and stimulus modality).

Response format

Evidently, the different ways of eliciting the responses when measuring metaphor comprehension pose diverse cognitive and linguistic demands. For example, earlier studies that tested metaphor comprehension of young children by asking them to explain or paraphrase a metaphor concluded that metaphor comprehension was not fully acquired until later in development (e.g. Winner, Rosenstiel, & Gardner, 1976; see Winner, 1988 for an overview; see Pouscoulous, 2011; 2014 for discussion). Alternatively, these findings may be explained by other variables such as response format demands (Pouscoulous, 2014). For example, metaphor explanation or justification tasks require a participant to articulate associations between metaphor topic and vehicle (e.g., *sister* and *butterfly* in *'My sister is a*

butterfly'). Therefore, performance also depends on metalinguistic judgment as well as expressive language and executive control skills. In addition, verbal explanation tasks require participants to explain the meaning of a metaphor to another person, and are therefore more socially demanding than written or computer-based tasks. Explanation tasks might also trigger the processing of the other person's reactions indicating whether the message was understood or not, thus engaging social-communication skills. By contrast, multiple-choice tasks do not rely on expressive language or meta-linguistic skills and require minimal social interaction with the examiner. However, multiple-choice tasks might be more costly in terms of need for inhibiting the false alternative(s) and selecting the correct one, as suggested by evidence from patients with brain lesions (Rapp, Felsenheimer, Langohr, & Klupp, 2018). The important role of the response format in metaphor comprehension is also supported by studies explicitly comparing different tasks. For instance, a study by Perlini et al. (2018) showed that only results from verbal explanation (but not multiple-choice) tasks yielded statistically significant difference between patients in the early phases of psychosis and controls.

Linguistic characteristics

Here, we present available evidence regarding the role played by different linguistic characteristics of the metaphor: metaphor familiarity, syntactic structure of the metaphor; linguistic context, and stimulus modality.

Metaphor familiarity. Metaphors are often differentiated according to whether they are conventional (i.e., well-established and often encountered in a language), or novel (i.e., not familiar, based on creative invention) (Bowdle & Gentner, 2005; Rosetti et al., 2018; Varga et al., 2014). For instance, a metaphor like "The sky's scarf is coloured" (Melogno et al., 2012) is considered novel, while "There is a flood outside the museum" (Rundblad & Annaz, 2010a),

where flood refers to 'lots of people', is considered a lexicalised/conventional metaphor. Both behavioural and neuroimaging evidence from different populations suggests different processing patterns for metaphor familiarity modulation, and, in particular, a facilitation for conventional compared to novel metaphors (Bambini, Gentili, Ricciardi, Bertinetto, & Pietrini, 2011; Blasko & Connine, 1993; Gluksberg, Gildea, & Bookin, 1982; Lee & Dapretto, 2006; Mashal, Faust, Hendler, & Jung-Beeman, 2009; Rapp et al., 2018; Rossetti et al., 2018; Varga et al., 2014). This might be because at least highly conventional metaphors are to be retrieved from the long-term memory where they are stored as learned lexical units, whereas novel metaphors might to a greater degree depend on the pragmatic ability to make context-relevant inferences (see Pouscoulous, 2011; 2014; Wilson & Carston, 2006 for discussions). Conventional metaphors may, therefore, be understood more quickly and with less cognitive effort, whereas the online processing required for novel metaphors could result in longer processing time involving pragmatic ability to a greater extent. Nevertheless, the exact nature of the difference in comprehension of conventional vs. novel metaphors is still debated (Cardillo, Watson, Schmidt, Kranjec, & Chatterjee, 2012).

Syntactic structure of the metaphors. Metaphors in the literature and discourse appear in various syntactic structures. For example, nominal metaphors express the metaphoric meaning using a noun (e.g., "Caroline is a princess", Wilson & Carston, 2006), predicate metaphors use a verb (e.g., "The rumor flew through the office", Utsumi & Sakamoto, 2011), and adjective metaphors use an adjective (e.g., *"sharp tongue",* Kasirer & Mashal, 2014). The cognitive effort required for the comprehension of the metaphors of different syntactic structure is likely to diverge (Cardillo et al., 2012; Chen, Widick, & Chatterjee, 2008). For instance, understanding nominal metaphors is suggested to entail either comparison (the assumption that metaphors convey similarities between semantically distinct

concepts; Gentner, Bowdle, Wolff, & Boronat, 2001), categorization, (the establishment of taxonomic relations between semantically distinct concepts; Glucksberg, 2003), or both comparison and categorization (Bowdle & Gentner, 2005). On the contrary, predicate metaphors may be understood through a process of highlighting core abstract conceptual features of a verb (Chen et al., 2008). Adjective metaphors are also said to be comprehended through categorization (Gluksberg, 2001; Glucksberg & Keysar, 1990) or by a two-stage categorization process (Utsumi & Sakamoto, 2007). This variation resulting from the different syntactic structures of metaphors may impact study outcomes.

Linguistic context. Metaphors in real life are usually encountered in sentences and/or discourse. Therefore, presenting metaphors with little or no context creates an artificial situation and may obscure the individual's ability to interpret a metaphorical expression. A number of studies on figurative language in individuals with TD as well as clinical populations (i.e., schizophrenia) suggest that the presence of a supportive context can significantly facilitate access to non-literal meaning (Chakrabarty et al., 2014; Pouscoulous, 2011; 2014). In line with this, Event Related Brain Potential (ERP) studies have shown that, in the earlier phases of processing, higher integration efforts are required for metaphoric expressions presented in minimal context compared to supportive context (Bambini, Bertini, Schaeken, Stella, & Di Russo, 2016).

Stimulus modality. The mode of the metaphor stimuli (i.e., auditory, written/visual) may also impact performance. For example, young children are usually tested with auditory tasks where they listen to the verbal metaphors and instructions because of their not-yet-adequate reading ability to complete written tasks or read instructions. However, it is not entirely clear whether and how the stimulus modality impacts metaphor comprehension in older children. In addition, metaphor tasks often incorporate a picture/image component to facilitate

comprehension of verbal metaphor (e.g., in Rundblad & Annaz, 2010b). Evidence from brain damaged patients suggests that right-hemisphere damaged patients performed better on a verbal than on a visuo-verbal test relative to the control group of participants without brain damage (Rinaldi, Marangolo, & Baldassarri, 2004). This might be explained by a disadvantage in processing visual information or by the challenges associated with cross-modal processing.

In sum, evidence suggests that task properties are essential to performance on metaphor comprehension tasks. This may give rise to different processing strategies in individuals with TD and ASD and affect statistical differences between clinical and control groups. Since the task properties are often associated with changes in behavioural and neural response in processing metaphors, psycho- and neuro-linguistic studies are increasingly based on extensive ratings of metaphor materials. To this end, norms have been established offering metaphorical expression characterizations along several linguistic dimensions, such as familiarity, interpretability, naturalness, imageability (e.g., Bambini, Resta, & Grimaldi, 2014; Cardillo, Schmidt, Kranjec, & Chatterjee, 2010; Cardillo, Watson, & Chatterjee, 2017; Jacobs & Kinder, 2017). These linguistic dimensions, however, are much less established in the literature on metaphor comprehension in ASD.

Metaphor comprehension task properties in studies with participants with ASD

Studies that compare individuals with ASD to individuals with TD on metaphor comprehension have employed a variety of tasks with different properties. For example, both Happé (1993) and Norbury (2005) employed a sentence completion task where the participants were asked to finish each sentence with a word they could choose from a list. Another type of multiple-choice format was used in the study conducted by Adachi and colleagues (2004). They tested metaphor comprehension by metaphoric scenarios where the

children were asked to read the questions silently and choose from the four response options (one correct and three incorrect). In their study, Rundblad & Annaz (2010a) employed a different format, whereby open verbal responses were given in response to short stories that were accompanied by images/pictures to aid comprehension.

Intriguingly, these studies yielded distinct results regarding the magnitude of group-level differences in metaphor comprehension between individuals with ASD and controls with TD. In particular, Adachi et al. (2004), Happe (1993) and Rundblad and Annaz (2010a) found significantly lower ability to understand metaphors, whereas Norbury (2005) found no statistically significant difference between language-ability matched groups. Presumably, the open verbal response format used in Rundblad & Annaz' (2010a) study could be more challenging for at least some individuals with ASD with impaired meta-linguistic, expressive language or executive function-related skills (Bishop & Norbury, 2005; Kwok et al., 2015; Lewis, Murdoch, & Woodyatt, 2007; Melogno, Pinto, & Levi, 2015).

Furthermore, including pictures in metaphor task may also influence performance of individuals with ASD. In individuals with TD, including pictures in a metaphor task can be an advantage because visualization can aid comprehension of verbal metaphors. Using visual support properly, for example pictures accompanying verbal instruction to aid comprehension, is generally also encouraged in work with individuals with ASD (e.g., Dettmer, Simpson, Smith Myles, & Granz, 2000; Nelson, McDonnell, Johnston, Crompton, & Nelson, 2007; Rao & Gagie, 2006). Indeed, there is evidence from a priming study of probable benefit of using pictures over words to access meaning in ASD (Kamio & Toichi, 2000). Nevertheless, it should be noted that a task presented in two modalities may be more challenging for individuals with ASD as they may struggle to switch between visual and auditory information. This can be hypothesized on the basis of studies such as Reed and McCarthy (2011), where individuals with ASD, compared with participants with TD, showed greater

11

difficulty when different modalities were employed than when only one modality was required. However, the individual needs of the students vary (Rao & Gagie, 2006), resulting in some individuals with ASD benefiting most from picture support, while others from the written support.

Certain task properties might be more suitable than others for individuals across the spectrum, given the cognitive and linguistic strengths (i.e., unimpaired rote memory, interest in details) and differences or challenges (i.e., executive functions) often observed in this population. For example, with regard to the metaphor familiarity, individuals with ASD might have more difficulties than individuals with TD in understanding novel metaphors because comprehension of novel metaphors involves pragmatic operations to a greater degree than conventional ones (Pouscoulous, 2011). In particular, by being innovative and occasion-specific, novel metaphors rely on pragmatic inference involving context-specific meaning adjustments (Recanati, 2004; Sperber & Wilson, 2008; Wilson & Carston, 2006; 2007), while conventional metaphors should depend less on inferencing and more on lexical knowledge (Pouscoulous, 2014). Nevertheless, because they are likely to be stored in the lexicon and thus linked to vocabulary knowledge, conventional metaphors might also pose problems for individuals on the spectrum (Pouscoulous, 2011). Individuals with ASD have often been shown to have compromised or biased vocabulary (Tager-Flusberg, 1992; Tager-Flusberg et al., 1990). As vocabulary knowledge is closely related to metaphor comprehension in individuals with TD (Nippold, 2016), compromised vocabulary knowledge might be linked to difficulties in metaphor comprehension in individuals with ASD with poorer vocabulary.

Some examples of the different task properties employed in the ASD literature on metaphor comprehension are provided in Table 1.

INSERT TABLE 1 HERE

The substantial variability in the assessment tasks employed may account for differences in the results of the studies, making it critical to inspect the properties of these tasks. This issue has been highlighted in a few narrative reviews. For instance, Melogno, D'ardia, Pinto, and Levi (2012) stressed that the heterogeneity of the tasks requiring diverse comprehension skills as the main difficulty in assessing the contribution of different tasks/variables, and they emphasized the urgent need of a careful review of the literature. Likewise, a more recent review by Siqueira, Marques, and Gibbs (2016) claimed that contrasting findings across studies of figurative language (including metaphors) in different clinical populations (including ASD) may be related more to issues related to data collection than to a specific difficulty one population may have in understanding a certain type of figurative language.

The current study: objectives and research questions

The overarching aim of this study was to advance the knowledge and awareness of the impact of task properties on metaphor comprehension performance in individuals with ASD compared to individuals with TD. We aimed to accumulate the existing knowledge by synthesizing the earlier research using the methods of systematic review and a meta-analysis.

The present study (i) explored the properties of the metaphor tasks used in ASD research; (ii) investigated the group difference between individuals with ASD and TD on metaphor comprehension, as well as the relationship between the task properties and any between-study variation. We anticipated larger between-study differences in studies employing verbal explanation formats than studies using alternative response formats (e.g. multiple-choice response format).

Method

This review was pre-registered in the International Register of Systematic Reviews, PROSPERO, with the registration number: CRD42017057231. (Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42017057231).

First, we systematically reviewed the included studies in terms of metaphor task properties (response format and linguistic characteristics). We then undertook a meta-analysis to compare individuals with ASD to individuals with TD on metaphor comprehension, as well as to examine the relationship between response format and any between-study variation.

Data collection, study inclusion and coding

A systematic literature search was initially conducted on 14.04.2016 and was then updated on 04.04.2017. The words for the literature search and the search strategies were selected after discussions in the authors' team and in close collaboration with two librarians at the University of Oslo library with expertise in literature searching. The librarians' responsibility was to ensure that the right search strategies were used and adapted correctly to the different data bases. The following electronic databases were searched: Psychinfo, LLBA, Eric, Embase, Norart, Medline, Web of science. The following terms were used as keywords: *ASD OR asperger* OR autis* OR "pervasive developmental disorder* combined with *allegor* OR analogy OR analogies OR "figure* of speech" OR "figurative language" OR imagery OR imageries OR metaphor* OR simile*.* No restrictions in terms of the publication year were applied.

In addition to the searches in the databases, the key terms (*ASD and metaphor comprehension*; *Asperger and metaphor comprehension*) were applied to Google scholar to identify any grey literature (literature that are not published in scientific journals, e.g., working papers, conference proceedings) to minimize potential publication bias in the meta-

analysis. This step is important because studies with significant results and large effect sizes are more easily published than studies that report non-significant findings or small effect sizes (Borenstein, Hedges, Higgins, & Rothstein, 2009). Also, we manually searched the tables of contents of the following key journals: *Journal of Autism and Developmental Disorders* and *Autism.* Finally, we went through the reference lists of the included articles and book chapters.

To be included in both the systematic review and the meta-analysis, articles were required to meet the following predetermined criteria: (1) the studies had to report on metaphor comprehension separately (when results on metaphor comprehension were part of the results on one global figurative language variable the study was excluded); (2) only participants with ASD were included. Notably, although we consistently use the term 'ASD' according to the DSM-5 (APA, 2013), we expected that diagnoses in the included studies would be based on the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; APA, 1994 or DSM-IV-TR, 2000),* or *International Classification of Diseases (ICD-10; WHO, 1992)* criteria, which prevailed at the time the studies was conducted. Thus participants might have been diagnosed with Autistic Disorder, Asperger's syndrome/disorder, or Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS); (3) only the studies involving participants with primary diagnosis of ASD (without any co-occurring conditions) were included to avoid the influence of other conditions on the outcome; (4) study design had to compare individuals with ASD to individuals with TD (the groups could either be equated for chronological age (CA), CA and other variables including verbal abilities, or verbal abilities only). No CA restrictions were applied since metaphor comprehension difficulties in ASD are also found in adults with ASD (Happe, 1993); (5) studies had to report data necessary to calculate effect sizes such as mean and standard deviation or *p*-values as well as information and / or examples about the metaphor stimuli that were used; (6) studies could be reported in English,

Norwegian, Italian, Russian, Swedish, or Danish because at least one of the authors is competent in each of these languages. By including several languages, we aimed to avoid the language bias often observed in systematic reviews, which is characterized by overrepresentation of English studies (Borenstein et al., 2009). Titles and abstracts obtained from the search were screened for relevance based on the pre-determined inclusion criteria by the first author. In case of insufficient information to decide the relevance on the study in the title and abstract, the full-text was reviewed. Finally, 14 studies met the inclusion criteria. For further information on the screening process and a summary of the reasons that studies were excluded see Figure 1.

INSERT FIGURE 1 HERE

We coded the following study characteristics: author(s), publication year, diagnostic status, comparison group, CA of the participants (mean and standard deviations), sample sizes in each group, and means and standard deviations or *p* -values for measures of metaphor comprehension. The following information about the task properties was coded: response format, metaphor familiarity, syntactic structure of the metaphor, linguistic context, and stimulus modality.

Several considerations were made when extracting the means, standard deviations or *p*-values for calculating effect sizes in the meta-analysis. First, for the studies with multiple data collection points (e.g., intervention studies), only data from the first time-point was coded. This was to ensure the results were not influenced by any intervention effects. Second, to avoid estimate dependency, the data from the largest sample was extracted when overlapping samples existed (Borenstein et al., 2009). Third, to avoid the problems with assigning more weight to studies with more outcome variables (Borenstein et al., 2009), we calculated a

composite score of multiple outcomes (e.g., novel and conventional metaphors) within each of the studies. The composite score is the mean effect size, with a variance that considers the correlation among the difference outcomes. Thus, every study including multiple outcomes was represented by one score which was used as the unit of analysis.

As pre-determined, the first and the second authors initially discussed the coding procedure, then double-coded the data from 10 randomly selected papers and discussed the coding of the remaining four papers. The inter-rater agreements for the coded variables in the 10 randomly selected papers were as follows: 100% for author, publication year, ASD and comparison group, age of the participants, sample size in each group, metaphor familiarity, syntactic structure of the metaphor, and linguistic context; 97% for response format and stimulus modality; 93.10% for the metaphor comprehension measures (mean with SDs and *p*-values). Notably, a divergence on the metaphor comprehension measures emerged with regard to the study by Kasirer and Mashal (2014). The divergence was due to the inverted values for the ASD and TD groups reported on the table in the original article. The last author of the original paper has confirmed the typo in email correspondence. The correct values were used for calculating the effect sizes. The other disagreements between the raters were resolved by discussion and/or by consulting the original papers.

The procedure of systematically reviewing the task properties

A comprehensive coding scheme was developed for the scrutiny of the relevant data from the included studies. Data on metaphor properties were analysed in detail for response format and linguistic characteristics (metaphor familiarity, syntactic structure, linguistic context, and stimulus modality). The exact number of studies reporting on each of these properties was identified. The findings of the studies that experimentally examined a property of interest are presented in the results section descriptively. Lack of taking into account the properties was

also considered a noteworthy finding. If the studies did not report task properties, we tried to obtain the necessary information by locating a description of the task from previous studies through searching Google web by the task name.

Meta-analytical procedure

Statistical analyses were conducted using the Comprehensive Meta-Analysis software (CMA) Version 3 (Borenstein, Hedges, Higgins, & Rothstein, 2014).

Because of the importance of adjusting a meta-analysis to the studies examined (Borenstein et al., 2009), we made some considerations for effect size computations. In particular, we used the Hedge's formula for standardized mean difference (SMD) with a confidence interval of 95% to report effect sizes. Hedges' *g* was selected since it is corrected for sample sizes (Hedges, 1981) and studies on metaphor comprehension in ASD often include small samples. A positive Hedges' *g* value indicated that individuals with ASD had the higher group mean; a negative Hedge's *g* value indicated that the groups differed in favour of TD group. A 95 % confidence interval (CI) was calculated for each effect size to indicate whether it was significantly greater than zero. The effect is statistically significant if the CI does not cross zero. The effect sizes were interpreted based on Cohen's (1988) benchmarks, with effect $d = \leq 0.2$ reflecting a small effect, $d = \leq 0.5$ considered medium effect, and $d = \leq 0.8$ indicating a large effect. However, these values are relative and somewhat arbitrary both to each other and to the specific study and research method employed (Cohen, 1988; Thompson, 2007). Therefore, interpreting these guidelines in relation to the clinical consequences that the effect size may have (Lakens, 2013) is important to avoid misleading suggestions to the practice. Hence, reporting the effect sizes in the results section of this paper is complemented by a descriptive review.

Effect sizes across studies were averaged using a random-effects model, which does not assume that all studies in the meta-analysis share a common true effect size (Borenstein et al., 2009).

To visualise the distribution of effect sizes and CIs, and to detect possible outliers, a forest plot was used. We also performed sensitivity analysis to determine the impact of potential outliers. Sensitivity analysis makes it possible to estimate the adjusted overall effect size after removing studies one by one when extreme effect sizes are detected.

Heterogeneity. We used the *Q* test of homogeneity (Hedges & Olkin, 1985) to examine the heterogeneity in effect sizes. The *Q*-statistic with its *p*-value in a random effect model is a test of significance and reflects whether the variance is significantly different from zero. In addition, we used *I²,* which reflects the extent of overlap of confidence intervals and is considered a measure of inconsistency.

Publication bias. Despite our efforts to identify grey literature, low-effect or non-significant studies could still be missing from the meta-analysis. To detect and statistically estimate the potential retrieval bias, we examined a funnel plot, in which a sample-size dependent statistic is plotted on the Y-axis and the effect size is plotted on the X-axis. In the absence of publication bias, this plot should form a symmetrical funnel (Cooper, Hedges, & Valentine, 2009). However, the funnel plot can be difficult to interpret visually when using a random effects model (Lau et al., 2006). Therefore, in addition, a "Trim and Fill" analysis (Duval & Tweedie, 2000) was applied. In eventual presence of publication bias, the "Trim and Fill" analysis would be used to impute values in the funnel plot to make it symmetrical and an adjusted overall mean effect size would be calculated.

Results

First, the results from the literature search are reported, followed by the description of results from the systematic review of the task properties. Finally, we present the results from the meta-analysis.

Results from the literature search

The electronic search yielded 1219 references. In addition, one study was identified through searching in the references. All hits were screened and 14 studies (13 published papers and one conference proceeding (de Villiers et al., 2011) that met the inclusion criteria were included in the systematic review and meta-analysis. Information on the screening process and the reasons for study exclusion are reported in Figure 1.

Results from the systematic review of metaphor task properties

The detailed description of the task properties of the included studies is presented in Table 2.

INSERT TABLE 2 HERE

Response format. The answers across the tasks were elicited by the following response formats: *verbal explanation or justification*, where participants were asked to explain the meaning of the expression (*n*=2; Melogno et al., 2012; Rudblad & Annaz, 2010a); *multiple-choice*, where participants had to choose the correct answer among a series of 3, 4 or 5 options (*n* =7; Adachi et al., 2004; Huang, Oi, & Taguchi, 2015; Kasirer & Mashal, 2014; Kasirer & Mashal, 2016; Mashal & Kasirer, 2011; Olofson et al., 2014; Zheng, Jia, & Liang, 2015); *meaningfulness decision*, where participants were asked to decide whether the

expression makes sense or not (yes/no) (*n*=4; Chouinard & Cummine, 2016; Gold & Faust, 2010; Gunter et al., 2002; Hermann et al., 2013;). Two studies (Gunter et al., 2002; de Villiers et al., 2011;) combined *multiple-choice or meaningfulness decision* and *verbal explanation/justification formats*. De Villiers et al. (2011) used multiple-choice picture modality followed by the question requiring verbal explanation. Metaphor explanation responses were reported in the results. However, the scoring strategy is not explained in their paper and, therefore, is not clear whether the responses from the multiple-choice task have also been merged in the reported results. Gunter et al. (2002) used three tasks (multiple-choice combined with verbal explanation and meaningfulness decision task requiring to decide whether metaphors were plausible or not). However, the tasks were not described in detail in the paper, so we obtained the necessary information about the task properties by searching previous studies that employed the same tasks (Bottini et al., 1994; Jodzio, Lojek, & Bryan, 2005). Furthermore, Gunter et al. (2002) did not explain how the answers were scored and how the results obtained from the multiple-choice and verbal explanation tasks were presented in relation to each other.

None of the included studies manipulated response format in order to investigate its impact on performance.

Metaphor familiarity. Most studies employed tasks that included novel as well as conventional metaphors (*n* =7; Gold & Faust, 2010; Gunter et al., 2002; Kasirer & Mashal, 2014; Kasirer & Mashal, 2016; Mashal & Kasirer, 2011; Olofson et al., 2014; Zheng et al., 2015;), while others included only novel (*n* =2; Hermann et al., 2013; Melogno, D'Ardia, Pinto, & Levi, 2012) or conventional metaphors (*n* =1; Rundblad & Annaz, 2010a). Four studies (Adachi et al., 2004; Chounard & Cummine, 2016; Huang et al., 2015; de Villiers et al., 2011) did not specify metaphor familiarity.

Based on the results of the included studies, the impact of familiarity varied across studies, with some studies reporting group differences for conventional, but not for novel metaphors (Mashal & Kasirer, 2011; Kasirer & Mashal, 2016), whilst others reported no group differences based on familiarity (Kasirer & Mashal, 2011). For example, some studies found that individuals with ASD could interpret both conventional metaphors (e.g., "Susan is a warm person") and novel metaphors (e.g., "Susan is a toasty person"; Olofson et al., 2014), and others found that novel metaphors were more difficult for individuals with ASD than conventional metaphors, yet this was also the case for individuals with TD (Gold & Faust, 2010; Zheng et al., 2015).

Syntactic structure. Based on those studies that provided information about syntactic structure or examples of metaphor items, the tasks varied greatly according to this variable as well. Six studies (Adachi et al., 2004; Chounard & Cummine, 2016; Gunter et al., 2002; Hermann et al., 2013; Huang et al., 2015; Zheng et al., 2015) involved (mostly) nominal or mixed syntactic structure. Five studies (Gold & Faust, 2010; de Villiers et al., 2011; Kasirer & Mashal, 2014; Kasirer & Mashal, 2016; Mashal & Kasirer, 2011) involved word pairs (noun-adjective pairs). Note that word-pair metaphors in de Villiers et al. (2011) were incorporated in interrogative sentence ("Which one is the blind house?"), while other studies did not embed word-pair metaphors in any context. Syntactic structure for conventional metaphors in Gunter et al. (2002) was not specified. Melogno et al. (2012) and Rundblad & Annaz (2010a) included sentences. Olofson et al. (2014) also included sentences with either verbs (predicate metaphors) or adjectives and was the only included study that explicitly focused on conceptual metaphors. Notably, the syntactic structure might be linked to different theoretical accounts of metaphor. For example, pragmatics-oriented scholars mostly consider "X is Y" expressions, while the literature in Cognitive Linguistics focuses on the multiplicity of

linguistic structures that might reflect underlying conceptual metaphors and considers metaphorically used verbs or longer expressions. However, this kind of theory-driven distinction has not been considered in the literature on ASD.

Overall, since some studies failed to provide information on syntactic structure, and several papers included only a few examples of metaphors without indicating whether the metaphor task was consistent in terms of the syntactic structure, the exact number of studies using any specific syntactic structure is impossible to report. Moreover, it is important to note that there might have been inconsistent items in the datasets. For instance, Gunter et al. (2002) adopted novel (or unusual, as they are call them in the paper) metaphors from Bottini et al. (1994), which were mostly nominal (X is Y): so –following our methodological choice of basing the review on what reported by the authors in the paper– we made a judgement based on this information and classified the items used in this study as nominal. However, we are aware that at least some metaphor items are not nominal (see the metaphor examples provided by Bottini et al., 1994). Similarly, Adachi et al. (2004) used metaphors with mixed structures. Huang et al. (2015) translated the same stimuli used by Adachi et al. (2004) from Japanese into Taiwanese. One of the example items in both Adachi et al. (2004) and Huang et al. (2015) studies is however translated into English as a simile. Although metaphors and similes are different figurative types and are understood differently (Happé, 1993), we decided to maintain these studies in the analysis both to be consistent with our methodological approach (basing the review on what was reported by the authors) and because the other example items in Adachi et al. 2004 were indeed metaphors.

For all the above reasons, and also because none of the included studies manipulated syntactic structure, the impact that variation in this linguistic variable might have on the group differences in metaphor comprehension is not clear.

Linguistic context. The type of context across the studies varied from none or minimal context (word pairs or sentence-level, $n = 9$; Chouinard & Cummine, 2016; Gold & Faust, 2010; Gunter et al., 2002; Hermann et al., 2013; Kasirer & Mashal, 2014; Kasirer & Mashal, 2016; Mashal & Kasirer, 2011; Melogno et al., 2012;) to scenarios or short stories with or without accompanying pictures ($n = 5$; Adachi et al., 2004; de Villiers et al., 2011; Huang et al., 2015; Olofson et al., 2014; Rundblad & Annaz, 2010a; Zheng et al., 2015). The task employed by Melogno et al. (2012) involved metaphors presented both in decontextualized sentences and in short story context. However, no results relating to the influence of the context are reported in that study. Other studies did not manipulate the context experimentally. Accordingly, no results regarding the impact of linguistic context on group differences in metaphor comprehension are reported in this review.

Stimulus modality. Five studies (Adachi et al., 2004; Gold & Faust, 2010; Hermann et al., 2013; Huang et al., 2015; Zheng et al., 2015) presented the stimuli in written modality. Four studies (Gunter et al., 2002-for the conventional metaphor task only; Chounard & Cummine, 2016; Olofson et al., 2014; Rundblad & Annaz, 2010a) delivered metaphor comprehension aurally. Computer-based tasks were administered either aurally (Olofson et al., 2014) or in writing (Gold & Faust, 2010; Hermann et al., 2013). Five studies (Gunter et al., 2002; Kasirer & Mashal, 2011; Kasirer & Mashal, 2014; Mashal & Kasirer, 2016; Melogno et al., 2012) did not specify the modality. Gunter et al. (2002) did not report information about the stimulus modality, but we could identify the modality (for conventional metaphors only) in the previous study (Jodzio et al., 2005). Stimulus modality is not specified in Melogno et al. (2012). De Villiers et al. (2011) employed stimuli with pictures, but without any indication whether participants were asked to read the metaphors or whether the questions were asked

aurally. Three additional studies included stimulus material with pictures (Olofson et al., 2014; Rundblad & Annaz, 2010a; Zheng et al., 2015).

The only study that reported that they used a test validated for the age-group of the participants in the study was of Melogno et al. (2012).

Metaphor comprehension in individuals with ASD and TD controls: a meta-analysis

Fourteen independent effect sizes, involving 336 individuals with ASD (mean sample size = 24, SD 15,01, range 8-54) and 498 individuals with TD (mean sample size = 35.57, SD 48.47, range 8-199), examined the differences in metaphor comprehension between the two groups. The standardized mean effect size was moderate, g = -0.63, 95 % CI [-0.80, -0.46], $p=$ <0.001) in favour of individuals with TD. This indicates that individuals with ASD on average have more difficulties in metaphor comprehension compared to individuals with TD.

The heterogeneity between studies was not significant (Q (13) = 16.50, p = 0.22), and 21.20% of true variability (I^2) could be explained by individual study characteristics. Higgins, Thompson, Deeks, and Altman (2003) provide some rough benchmarks for I^2, which refer to the question of what proportion of the observed variation is real. They suggest considering values below 25% as low.

Sensitivity analysis showed that the overall effect size ranged from g = -0.66, 95% CI [-0.85, -0.46] to g = -0.57, 95% CI [-0.71, -0.42]. The funnel plot showed symmetrical distribution indicating no publication bias. No studies were imputed in a Trim and Fill analysis indicating again that no publication bias was detected. The forest plot (Figure 2) shows the group differences and CIs between individuals with ASD and TD in terms of the metaphor comprehension.

INSERT FIGURE 2 HERE

Impact of response format on between-study variance

We intended to examine the response format as a potential moderator of between-study variation. However, due to the limited number of the studies on each response format category (e.g., only two studies on verbal explanation format), a meta-regression or a subgroup analysis (which may be considered as a special case of meta-regression; Fu et al. 2011) would yield non-reliable results because of low statistical power. Specifically, it is recommended that for a categorical subgroup variable (response format in our case), each subgroup should include a minimum of 4 studies (Fu et al., 2011). Therefore, we qualitatively report the observed effect sizes with CIs to identify the patterns of possible relationships between response format and the heterogeneity between studies. Although not aggregated, the descriptively reported effect sizes can still guide interpretation of results and inform future studies.

Among the four types of *response format* identified in the included studies, the two studies that required verbal explanations showed moderate to large effect sizes (Melogno et al., 2012; Rundblad & Annaz, 2010a). One of these studies (Rundblad & Annaz, 2010a) generated the largest effect size from the included studies: $g = -2.20$, 95% CI [-3.14—1.27]. This study employed an open verbal explanation task in which the short stories were accompanied with simple, hand drawn pictures (hence, two modalities were involved). The experimenter read each story while presenting the child with one simple picture showing one story character. The child was asked to report what that character saw. In the other study that used verbal explanation response format (Melogno et al., 2012), the yielded effect size was moderate ($g = -0.62$, 95% CI [-1.18, -0.04]). This study assessed metaphor comprehension using the Junior Metaphor Comprehension Test, a validated tool for use with a pediatric population (Junior MCT; Pinto, Melogno, & Iliceto, 2008).

Large group differences were found in the two studies that combined verbal explanation with other response formats. De Villiers et al. (2011) combined verbal justification/explanation and picture multiple-choice response formats in the same task and yielded large effect size: g =-0.84, 95% CI [-1.41, -0.27]. In Gunter et al. (2002) the combined effect size for the three tasks used (multiple-choice combined with verbal explanation and meaningfulness decision) was large: g = -1.14, 95% CI [-2.17, 0.11]. Two caveats related to this study must be mentioned: first, this study included a very small sample (n = 8), and second, the stimulus material in the meaningfulness task involved linguistically complex language (i.e., "The politician who didn't give straight answers was jumping ditches"; "The meaning of life is an itch you can't scratch"; or "The old man had a head full of dead leaves"; see Bottini et al. (1994) for more examples).

For the seven studies that employed multiple-choice approach only (Adachi et al., 2004; Huang, Oi, & Taguchi, 2015; Kasirer & Mashal, 2014; Kasirer & Mashal, 2016; Mashal & Kasirer, 2011; Olofson et al., 2014; Zheng, Jia, & Liang, 2015), effect sizes varied from small to large: g = -0.37, 95 % CI [-0.83, -0.09] to g = -0.91, 95 % CI [-1.69; -0.12]. In the three studies that used meaningfulness decision tasks effect sizes ranged from small to moderate: g = -0.39, 95% CI [-1.00, 0.23] to g = -0.52, 95 % CI [-1.03, -0.01].

Discussion

The aim of the present systematic review and meta-analytic study was twofold: first, we sought to explore the properties of the metaphor tasks used in research involving individuals with ASD in a systematic manner. Second, we intended to examine the extent to which the groups of individuals with ASD differed from individuals with TD on metaphor

comprehension, and whether any between-study variation could be explained by the properties of metaphor comprehension tasks.

We found that the included studies employed different types of materials and tasks either invented by the researchers who designed the studies or adopted (and sometimes translated) from previous studies. Although the task properties varied greatly, the potential impact of the task properties was rarely considered. Regarding the group differences, overall, individuals with ASD fell behind their TD controls in comprehension of metaphors. The patterns show that verbal explanation response format – either pure verbal explanation or in combination with other response formats– resulted in the large effect sizes. However, due to the scarce experimental manipulation of task properties, their moderating role could not be established based on the included studies.

Properties of the tasks are seldom considered and/or controlled in the studies

In terms of the *response format,* different approaches are adopted across the studies, with the most common being multiple-choice format. Less often used response formats include verbal explanation, followed by verbal explanation combined with another response format, and meaningful decision format. It is possible that the studies involving individuals with ASD avoid using verbal explanation tasks because of known challenges related to this type of response format (i.e., cognitively, linguistically and socially more demanding). Since the impact of response format has been associated with the between-group difference in other populations (see for example, Perlini et al., 2018), we anticipated detecting similar patterns in studies comparing individuals with ASD to those with TD. However, the included studies did not experimentally manipulate the response format. Therefore, firm conclusions based on the results reported in these studies cannot be drawn.

A noteworthy finding of this review is that the impact of some of the properties, such as metaphor *familiarity,* are more frequently considered than others. The reason might be that ASD is a suitable condition for studying the distinction between novel and conventional metaphors, due to the common impairment observed in pragmatic language in this population (Paul & Norbury, 2012). Specifically, individuals with ASD should have more problems with comprehending novel as compared to conventional metaphors due to the involvement of inferential pragmatic ability to a greater extent in novel metaphors than in conventional metaphors. Yet results regarding the impact of metaphor familiarity on group difference between individuals with ASD and individuals with TD are mixed and inconclusive. This might partially be explained by different ways in which the studies have rated the degree of familiarity. For example, familiarity is often assessed based on the ratings collected from a limited number of participants, which might be not reliable given the large differences in subjective judgment on familiarity. Accordingly, Thibodeau, Sikos, and Durgin (2018) have questioned construct validity of sentence-level subjective ratings of metaphors collected from native speakers and argued that familiarity ratings are likely to be confounded with processing fluency (i.e., how easily people understand the sentences). Moreover, it may also be that other properties that co-vary with familiarity, such as word-level psycholinguistic characteristics (e.g., frequency, concreteness, length), as well as metaphors' characteristics such as interpretability, naturalness, and imageability may account for distinct results (Cardillo et al., 2010). Therefore, we argue that using stimuli for which these properties have been rated by large samples of participants, and controlled for, could offer a more robust benchmark to explore the difference between familiar and unfamiliar metaphors. In addition, the use of controlled materials will favour the comparison of the findings across studies, which will be a great advantage for future systematic reviews and meta-analytic studies.

Syntactic structure is the least explored property and was not normally controlled for in the studies. Given that individuals with ASD frequently show impairments in syntax (Brynskov, Krøjgaard, & Eigsti, 2016), and given the evidence that metaphors in different syntactic structures are comprehended differently (Cardillo et al., 2012; Chen et al., 2008), the syntactic structure of the metaphoric items may have been important to take into account in research with individuals with ASD. However, since most studies do not report on syntactic structure (or offer inconsistent examples), we cannot conclude that the stimuli did consistently display the same structure throughout the task. Based on our results, reporting the number of studies according to the syntactic structure of the metaphors should be therefore considered with caution.

The impact of *context* on the between group difference is also poorly explored in ASD research. This finding is striking given that inferring meaning from context has been reported to be challenging for individuals with ASD due to a cognitive difference in the normal drive for coherence (Frith, 1989; Frith & Happé, 1994; Happé, 1999; see, however, Brock, Norbury, Einav, & Nation, 2008, suggesting that differences in processing linguistic context in individuals with ASD are actually related to individual differences in their core language abilities). This implies that, although context may facilitate comprehension in TD, it may in fact pose problems in individuals with ASD, which is in line with the 'context blindness' hypothesis referring to a lack of contextual sensitivity in ASD (Vermulen, 2014). Interestingly, one study that was screened within this review (but excluded in the full-text screening stage due to the reported co-occuring conditions among individuals with ASD) found that context facilitates metaphor comprehension in ASD (Giora, Gazal, Goldstein, Fein, & Stringaris, 2012). However, one study is not enough to infer a pattern concerning the role of context. Notably, not only the presence or absence of context, but also the type of context may matter, because context with a large amount of information could hamper comprehension

by overloading participants' working memory and affecting attention (see Boxhoorn et al., 2018; Pennington & Ozonoff, 1996).

Regarding *stimulus modality,* most included studies used written tasks. This could be preferable when measuring metaphor comprehension in individuals with ASD since written tasks do not pose high social interaction demands and are less taxing for memory. In addition, aurally delivered tasks might be difficult for individuals with ASD due to their characteristics in processing auditory semantic information from spoken language (see O'Connor, 2012 for a review). On the other hand, it is still unclear whether written words facilitate comprehension processes for individuals with ASD in general. There is some evidence that young individuals with ASD benefit more from written word priming (not metaphorical, but 'conventional, 'literal' words) in their lexical access than young TD controls and older individuals with ASD (Harper-Hill, Copland, & Arnott, 2014; see however Kamio & Toichi, 2000 suggesting the possible advantage of pictures over words in access to semantics in ASD). It is unknown if similar effects encompass the case of metaphor.

In sum, there is a lack of attention to the role of task properties in performance on metaphor comprehension tasks in the existing ASD research. We observed the discrepancy in the task properties across the studies, as well as the limited number of studies experimentally manipulating the task properties. Therefore, strong conclusions about the extent to which task properties can explain the distinct findings in the ASD literature cannot be drawn from this study. Nevertheless, our study offers new insight into how studies in ASD have assessed metaphor comprehension and directs the focus towards the importance of acknowledging the substantial variability in tasks and their properties when interpreting the results from the existing studies. Also, it calls for careful consideration when designing and reporting of task properties in metaphor studies.

Metaphor Comprehension is more challenging for individuals with ASD than for individuals with TD

Overall, individuals with ASD as a group exhibited more difficulties in metaphor comprehension than the comparison group of individuals with TD. This finding is consistent with the results from prior studies (i.e., Happé, 1993; Rundblad & Annaz, 2010a; van Herwegen & Rundblad 2018), as well as with the findings from a recent meta-analysis (Kalandadze et al., 2018).

Taken together, this evidence indicates that, as a group, individuals with ASD more frequently experience problems in metaphor comprehension. Nevertheless, we need to acknowledge that there are several possible explanations for the significant group difference. For example, the meta-analysis (Kalandadze et al., 2016) and single studies have found that group-matching strategies could explain the between-study variation on figurative language comprehension. In particular, if ASD and TD groups were matched for language ability, the groups have been found to not differ significantly on metaphor comprehension (Norbury, 2005). These variables should necessarily be taken into account when explaining the difficulties with metaphor comprehension in individuals with ASD, together with the role of the metaphor task properties, which, despite its well-documented importance for metaphor comprehension, has not been examined until now.

Observed pattern of the associations between the response format and between-study variation.

As hypothesized, verbal explanation tasks (pure verbal explanation or combined with other response formats) are, based on the observed effect sizes, most challenging for individuals with ASD as compared to TD controls. This is not surprising because explaining metaphorical meaning is both cognitively, linguistically, and socially demanding, as it requires planning

and formulating utterances, and thus relies on expressive language as well as metalinguistic and executive skills, which have often been found to be challenging for individuals with ASD (Hill, 2004; Kwok et al., 2015; Lewis et al., 2007). This finding also converges with results from an irony processing study (another type of figurative language) in individuals with ASD, in which minimizing the verbal and pragmatic demands of the task resulted in the similar accuracy in judging speaker intent for ironic criticism between the groups of individuals with ASD and TD (Pexman et al., 2011).

However, using a verbal explanation task, if validated for use with the target group, might reduce the magnitude of the group difference. For example, Melogno et al. (2012) used the test Junior MCT designed for the specific age-range (4-6 years) and validated it for use with 'pediatric population'. Using the validated tool likely resulted in smaller effect size compared to other studies that used not validated verbal explanation tasks. The distinction between the results based on assessing metaphor comprehension with validated vs. not validated tool is fundamental, but, unfortunately, our result is based on one only study (Melogno et al., 2012). In order to draw clearer conclusions more studies using validated materials to study metaphor comprehension are needed.

Meaningfulness decision tasks seem to be the least challenging for individuals with ASD when compared to TD individuals. Although somehow surprising, this might be because this type of task does not require expressive language skills, planning and formulating the responses as in verbal explanation tasks, nor inhibiting the incorrect alternatives as in multiple-choice tasks. In addition, meaningfulness decision tasks might be less socially demanding since they require less interaction with the examiner than verbal tasks. Meaningfulness decision tasks might, therefore, be less taxing for individuals with ASD than verbal explanation tasks.

Limitations

Several limitations of this study should be considered when interpreting the findings. First, the inconsistency in using response format across the studies made it unfeasible to examine the potential moderating effect of this variable. Although descriptively reported effect sizes are informative, a meta-regression or a subgroup analysis would allow for a more accurate examination of the relationships between the response format and between-study variation. Second, none of the included studies attached the stimulus materials in appendices. Some studies presented a few examples of the metaphorical items, whereas others did not even report the examples. Although we did not have access to the full list of stimuli, the information provided in the papers was sufficient for our purposes. Future reviews that want to examine the consistency of the metaphor stimuli in the existing studies, which is definitely worth investigating, should contact the authors and request the full set of stimuli. Future reviews should also examine what types of metaphors the stimuli contain (e.g., nominal metaphors, as in "Sally is a chameleon", or conceptual metaphors, like "I see it", where 'seeing' indicates 'knowing' (KNOWING IS SEEING).

Another important methodological limitation was the small sample sizes in some included studies (e.g., eight participants in Gunter et al., 2002). Since larger sample sizes correspond to less sampling bias (Borenstein et al., 2009), high-powered studies would provide better effect size estimates for this meta-analysis. However, the advantage of the meta-analysis over a single study is the increased statistical power achieved via aggregating the effect sizes from multiple samples. We, therefore, propose that the magnitude of the group difference reported in our study gives a reliable result.

These limitations will be overcome once more tightly conducted studies are available, allowing for consistent examination of potentially important variables affecting metaphor comprehension in individuals with ASD as compared to individuals with TD.

Implications

The main implication of the findings of our study for future research on metaphor comprehension is that stimuli and tasks need to be created by carefully taking into account a range of characteristics, such as the linguistic properties and response format, whose role in modifying behavioural and neural responses is well-known in the literature (e.g., Bambini et al., 2014; Schmidt & Seger, 2009). Furthermore, studies should consistently examine the role of task properties experimentally to investigate their relationships with the performance among individuals with ASD and those with TD.

Notably, the different tasks may each have advantages, but at the same time they can impede comprehension if inappropriately used with individuals with ASD. For instance, multiple-choice tasks can be desirable from the psychometric perspective due to an easy and precise scoring (Rapp et al., 2018) and high reliability (see for instance the different reliability values of figurative language tasks-multiple choice vs. verbal explanation in Carotenuto et al., 2018). On the other hand, multiple-choice tasks are more susceptible for measurement error due to the possibility of guessing the responses (Kline, 2009), as well as due to tapping more executive functions because of the need to inhibit the incorrect alternatives and select the correct one. Another aspect that should be weighed up when designing metaphor comprehension tasks in ASD research concerns the number of options provided in multiple-choice task. For instance, there is evidence from another pragmatic domain (i.e., scalar implicatures) that presenting two vs. three options might accounted for the presence or absence of group differences between individuals with ASD and individuals with TD (Schaeken, Van Haeren, & Bambini, 2018).

As for verbal explanation tasks, these appear to be more sensitive than other tasks in detecting impairment in metaphor comprehension and allow us to establish with more

confidence whether the metaphors were understood or not. However, it must be pointed out that verbal explanation tasks are not recommended for use with vulnerable groups because of the extra demands they pose on the participants (see Norbury, 2004). Moreover, when using verbal explanation tasks, it is important that experimenters receive adequate training in order to achieve better reliability in scoring responses.

In general, using metaphor tasks created *ad hoc* for the specific purpose of the study is often preferable for researchers, given the multifaceted nature of a metaphor. However, greater advantages would be obtained from the use of validated or/and standardized tests with good psychometric properties. Absence of tests with properties that are consistently controlled for across the studies makes comparison of the results difficult. Only one study (Melogno et al., 2012) used a validated instrument and, despite using the demanding response format of verbal explanation, the effect size yielded was smaller than in the study by Rundblad and Annaz (2010a), which used non-validated verbal explanation task. This anecdotal observation should be investigated in future studies.

Another important suggestion from our findings is that the stimulus materials should be attached to the published papers. Also, providing a detailed description of the stimulus materials is essential to enable interpretation of the findings. In general, we propose that journals develop criteria for reporting metaphor studies in order to make quality appraisal of research for the readers and for future review studies possible.

Furthermore, based on the results of our study indicating limited number of studies using online methods, more high-quality studies on metaphor comprehension in ASD are needed combining off-line and online comprehension methods widely used in psycholinguistic research. Considering the many demands off-line tasks pose (i.e., social and linguistic), it is difficult to pinpoint the real sources of possible difficulties in metaphor comprehension when assessed off-line. On-line tasks such as those employing eye-tracking methodology, priming

paradigms, and computerised tests can therefore add important insight to the knowledge of

metaphor comprehension in ASD by measuring implicit processing (see for example Naigles,

2017, for innovative paradigms and methods to investigate language in ASD that could

beneficially be used in metaphor research as well). For instance, priming paradigms might

offer fine-grained measures of the difficulties experienced by individuals with ASD,

elucidating patterns in response times (Chahboun, Vulchanov, Saldaña, Eshuis, &

Vulchanova, 2017). In addition, behavioural data could profitably be combined with the data

on brain functioning to explain the neurocognitive and neurolinguistic processes underlying

metaphor comprehension in individuals with ASD as compared to TD controls. For instance,

Gold, Faust and Goldstein (2010) employed Event Related Potential (ERP) recordings to

examine difficulties in semantic integration in ASD. The sample in this study, however,

overlapped with the sample in Gold & Faust (2010) and therefore the study was not included

in the meta-analysis. We did not identify any other study with data about the brain response

that met the inclusion criteria for this review.

The main practical implication of our findings is that individuals with ASD need extensive

support to learn metaphor comprehension strategies explicitly and that plans on how to

promote metaphor comprehension should be made. Intervention programs concerning

metaphor comprehension in ASD are very few, but results are promising. For instance,

Mashal and Kasirer (2011) and Melogno, Pinto and Di Filippo (2017) used thinking maps to

enhance the abstraction of semantic features in metaphors. Teachers, special educators, and

speech and language therapists could capitalize on this evidence and develop strategies to

stimulate metaphorical skills. To begin with, the students could be reminded that figurative

language involves the use of words in non-literal ways. Then the students could be

encouraged to use their meta-linguistic skills to consider the overlapping features between the

topic and vehicle of the metaphor (Nippold, 2016), similarly to the approach adopted in

Mashal and Kasirer (2011) and in Melogno et al. (2017). In addition, the students could be asked to collect metaphors from different sources such as advertisements and literature, including the context in which they occur (Nippold, 2016). Teachers may also incorporate metaphors of different degrees of familiarity in minimal or short story contexts and present them both aurally and in print, with and without pictures, eliciting the answers through different response formats (see Nippold, 2016, for more ideas).

Conclusions

This paper reports the systematic review and meta-analysis concerning task properties of the metaphor tasks used in ASD research and the role that they play in determining the differences between groups of individuals with ASD and those with TD in metaphor comprehension. By focusing on the impact of the task properties, this study contributes to the ongoing debate about the potential sources of between-study variation in metaphor comprehension in individuals with ASD and offers novel insight of studies of figurative language in this population.

The included studies used an array of different tasks with a range of properties, whose impact was rarely considered and/or experimentally manipulated. Individuals with ASD in general exhibited more difficulties in metaphor comprehension than their TD counterparts, but this difference is likely to be partially related to the task properties such as the response format. Yet, more research is needed to confirm the relationship between the task properties and between-study variance.

In light of the findings of our study, we argue that future metaphor comprehension studies comparing individuals with ASD to those with TD should carefully take into account task properties such as response format and linguistic characteristics (i.e., metaphor familiarity, syntactic structure of the metaphor, linguistic context, and stimulus modality).

Consideration of task properties is also necessary in order to design appropriate educational programs to improve figurative language competence and ultimately improve communication and academic skills of individuals with ASD.

Author contribution

Kalandadze conceptualized, designed and administered the study; created the coding protocol; established the inclusion and exclusion criteria; collected the data; screened the titles and abstracts; read/screened full-text articles; created the coding scheme and coded the variables; analysed and interpreted the results; drafted the manuscript; Main responsibility for revising and re-submitting the manuscript after peer-review.

Bambini contributed to the study design, especially to the selection of the metaphor task properties and the consideration of metaphor theory aspects; did the double-coding; contributed to analysis and interpretation of the results, especially of the systematic review; contributed to the writing of the manuscript; provided feedback as well as final approval of this paper; contributed to revising of the manuscript.

Næss supervised the research process; contributed to the study design, especially in the meta-analysis part; contributed to the interpretation of the results; contributed to the writing of the manuscript; provided feedback as well as final approval of this paper; contributed to revising of the manuscript.

References

References marked with an asterisk indicate studies included in the systematic review and the meta-analysis

*Adachi, T., Koeda, T., Hirabayashi, S., Maeoka, Y., Shiota, M., Wright, E.C., & Wasa, A. (2004). The metaphor and sarcasm scenario test: a new instrument to help differentiate high functioning pervasive developmental disorder from attention deficit/hyperactivity disorder. *Brain & Development,* 26, 301-306. doi:10.1016/S0387-7604(03)00170-0

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (DSM-V). Arlington, VA: APA

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders:* DSM-IV-TR. Washington, DC: American Psychiatric Association

American Psychiatric Association. *(1994). Diagnostic and statistical manual of mental disorders:* DSM-IV. Washington, DC: American Psychiatric Association

Bambini, V., Bertini, C., Schaeken, W., Stella, A., & Di Russo, F. (2016). Disentangling metaphor from context: An ERP study. *Frontiers in Psychology,* 7, 559. doi:10.3389/fpsyg.2016.00559

Bambini, V., Gentili, C., Ricciardi, E., Bertinetto, P.M., & Pietrini, P. (2011). Decomposing metaphor processing at the cognitive and neural level through functional magnetic resonance imaging. *Brain Research Bulletin,* 86(3–4), 203-216. doi: 10.1016/j.brainresbull.2011.07.015

Bambini, V., Resta, D., & Grimaldi, M. (2014). A dataset of metaphors from the Italian literature: Exploring psycholinguistic variables and the role of context. *PLoS ONE,* 9(9), e105634. doi: 10.1371/journal.pone.0105634

Bishop, D. V. M., & Norbury, C. F. (2005). Executive Functions in Children with Communication Impairments, in Relation to Autistic Symptomatology. 2: Response Inhibition. *Autism*, 9 (1), 29-43. doi.org/10.1177/1362361305049028

Blasko, D.G., & Connine, C.M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 19(2), 295-308. doi:10.1037/0278-7393.19.2.295

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2014). *Comprehensive meta-analysis. Version 3*. Englewood, NJ*:* Biostat, 104

Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein H.R. (2009). *Introduction to meta-analysis*. Chichester, England: John Wiley & Sons.

Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., …Frith, C.D. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language: A positron emission tomography activation study. *Brain,* 117, (Pt 6) 1241-1253. doi.10.1093/brain/117.6.1241

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

Boxhoorn, S., Lopez, E., Schmidt, C., Schulze, D., Hänig, S., & Freitag, C.M. (2018). Attention profiles in autism spectrum disorder and subtypes of attention-deficit/hyperactivity disorder. *European Child & Adolescent Psychiatry*, 27(11):1433-1447. doi:10.1007/s00787-018-1138-8

Brock, J., Norbury, C., Einav, S., & Nation, K. (2008). Do individuals with autism process words in context? Evidence from language-mediated eye-movements. *Cognition,* 108(3)*,* 896-904. doi: 10.1016/j.cognition.2008.06.007

Brynskov, C., Krøjgaard, P., & Eigsti, I. (2016). Language and communication in children with autism: Do research and clinical practice converge in Denmark? *Nordic Psychology,* 68(2), 100–113. doi:10.1080/19012276.2015.1071199

Camp, E. (2009). Metaphor. In L. Cummings (Ed.), *The Pragmatics Encyclopedia*, Oxford, London and New York: Routledge

Cardillo, E.R., Schmidt, G.L., Kranjec, A., & Chatterjee, A. (2010). Stimulus design is an obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavioral Research Methods*, 42(3), 651–664. doi:10.3758/BRM.42.3.651

Cardillo, E.R., Watson, C.E., & Chatterjee, A. (2017). Stimulus needs are a moving target: 240 additional matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavioral Research*, 49(2), 471-483. doi:10.3758/s13428-016-0717-1

Cardillo, E.R., Watson, C.E., Schmidt, G.L., Kranjec, A., & Chatterjee, A. (2012). From novel to familiar: turning the brain for metaphors. *Neuroimage*, 59(4): 3212–3221. doi: 10.1016/j.neuroimage.2011.11.079

Carotenuto, A., Arcara, G., Orefice, G., Cerillo, I., Giannino, V., Rasulo, M., Iodice, R., & Bambini, V. (2018). Communication in Multiple Sclerosis: Pragmatic deficit and its relation with cognition and social cognition, *Archives of Clinical Neuropsychology*, 33(2), 194–205. doi:10.1093/arclin/acx061

Carston, R. (2010). Lexical pragmatics, ad hoc concepts and metaphor: a relevance theory perspective. *Italian Journal of Linguistics*, 22(1), 153–180

Carston, R. (2012). Metaphor and the literal/nonliteral distinction. In K. Allan & K. Jaszczolt (eds.), *Cambridge handbook of pragmatics*. Cambridge: Cambridge University Press pp. 469--492

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

Chakrabarty, M., Sarkar, S., Chatterjee, A., Ghosal, M., Guha, P., & Deogaonkar, M. (2014). Metaphor comprehension deficit in schizophrenia with reference to the hypothesis of abnormal lateralization and right hemisphere dysfunction, *Language Sciences,* 44, 1-14. doi: 10.1016/j.langsci.2014.01.002

Chen, E., Widick, P., & Chatterjee, A. (2008). Functional anatomical organization of predicate metaphor processing. *Brain and Language,* 107(3), 194–202. doi: 10.1016/j.bandl.2008.06.007

Chouinard, B., & Cummine, J. (2016). All the world's a stage: evaluation of two stages of metaphor comprehension in people with autism spectrum disorder. *Research in Autism Spectrum Disorders,* 23, 107-121. doi: 10.1016/j.rasd.2015.12.008

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge Academic

Cooper, H.M., Hedges, L.V., & Valentine, J.C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation

*de Villiers, P.A., de Villiers, J.G., Diaz, S., Cheung, C., Alig, R., & Raditz, V. (2011, November). *Non-literal language and theory of mind in autism spectrum disorders.* Poster presented at the American Speech-language Hearing Association convention, San Diego, CA

Dettmer, S., Simpson, R.L., Smith Myles, B., & Granz, J.B. (2000). The use of visual supports to facilitate transitions of students with autism. *Focus on Autism and Other Developmental Disabilities,* 15(3), 163.169. doi:10.1177/1088357614528799

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics,* 56(2), 455-463. doi:10.1111/j.0006-34IX.2000.00455.x

Frith, U. (1989) A new look at language and communication in autism. *International journal of language and communication disorders,* 24(2), 123-150. doi:10.3109/13682828909011952

Frith, U., & Happé, F. (1994). Autism: beyond "theory of mind", *Cognition*, 50(1–3), 115-132. doi:10.1016/0010-0277(94)90024-8

Fu, R., Gartlehner, G., Grant, M., Shamliyan, T., Sedrakyan, A., Wilt. T.J,…Trikalinos, T.A. (2011). Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. *Journal of Clinical Epidemiology,* 64, 1187-1197. doi: 10.1016/j.jclinepi.2010.08.010

Gagnon, L., Goulet, P., Giroux, F., & Yves, J. (2003). Processing of metaphoric and non-metaphoric alternative meanings of words after right- and left-hemispheric lesion. *Brain and Language,* 87(2), 217–226. doi:10.1016/S0093-934X(03)00057-9

Gentner, D., Bowdle, B.F., Wolff, P., & Boronat, C. (2001). Metaphor is like analogy. In Centner, D, Holyoak, K.J, & Kokinov, B.N. (Eds.), *The analogical mind: Perspectives from cognitive science* (pp.199-253). Cambridge MA, MIT Press

Gernsbacher, M.A, Keysar, B., Robertson, R.R.W., & Werner, N.K. (2001). The role of suppression and enhancement in understanding metaphors. *Journal of Memory and Language, 45(3),* 433-450. doi:10.1006/jmla.2000.2782

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

Gibbs, R. W. (2011). Evaluating conceptual metaphor theory. *Discourse Processes*, 48(8), 529–562. doi:10.1080/0163853X.2011.606103

Gibbs, R.W., and Tendahl, M. (2006). Cognitive effort and effects in metaphor comprehension: Relevance theory and psycholinguistics. *Mind & Language* 21 (3), 379-403. doi: 10.1111/j.1468-0017.2006.00283.x

Giora, R., Gazal, O., Goldstein, I., Fein, O., & Stringaris, A. (2012). Salience and context: interpretation of metaphorical and literal language by young adults diagnosed with Asperger's syndrome, *Metaphor and Symbol,* 27(1), 22-54. doi: 10.1080/10926488.2012.638823

Glucksberg, S. (2001). *Understanding figurative language: From metaphors to idioms*. Oxford: Oxford University Press

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review,* 97(1), 3–18. doi:10.1037/0033-295X.97.1.3

Gluksberg, S. (2003). The Psycholinguistics of metaphor. *Trends in Cognitive Sciences,* 7(2), 92-96. doi:10.1016/S1364-6613(02)00040-2

Gluksberg, S., Gildea, P., & Bookin, H.B. (1982). On understanding nonliteral speech: can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior,* 21(1)*,* 85-98. doi:10.1016/S0022-5371(82)90467-4

*Gold, R., & Faust, M. (2010). Right hemisphere dysfunction and metaphor comprehension in young adults with Asperger syndrome. *Journal of Autism and Developmental Disorders,* 40(7), 800-811. doi:10.1007/s10803-009-0930-1

Gold, R., Faust, M., & Goldstein, A. (2010). Semantic integration during metaphor comprehension in Asperger syndrome. *Brain and Language,* 113(3), 124-134. doi: 10.1016/j.bandl.2010.03.002

Golden, A. (2010). Grasping the point. A study of 15-year-old students' comprehension of metaphorical expression in schoolbooks. In Low, G., Todd, Z., Deignan, A., Cameron, L. (Eds.), *Researching and applying metaphor in the real world*. Amsterdam, The Netherlands, and Philadelphia, PA: John Benjamins

*Gunter, H.L., Ghaziuddin, M., & Ellis, H.D. (2002). Asperger syndrome: Tests of right hemisphere functioning and interhemispheric communication. *Journal of Autism and Developmental Disorders, 32*(4), 263-281. doi:10.1023/A:1016326701439

Happé, F. (1995) Understanding minds and metaphors: Insights from the study of figurative language in autism. *Metaphor and Symbolic Activity*, 10(4), 275-295. doi:10.1207/s15327868ms1004_3

Happé, F. (1999). Autism: cognitive deficit or cognitive style? *Trends in cognitive sciences*, 3(6), 216-222. doi:10.1016/S1364-6613(99)01318-2

Happé, F.G.E. (1993). Communicative competence and theory of mind in autism: a test of relevance theory. *Cognition, 48*(2), 101-1. doi:10.1016/0010-0277(93)90026-R

Harper-Hill, K., Copland, D., & Arnott, W. (2014). Pathways to meaning: written and spoken word priming in children with ASD versus typically developing peers. *Research in Autism Spectrum Disorders, 8*(10), 1351-1363. doi: 10.1016/j.easd.2014.07.004

Hedges, L. & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press

Hedges, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics, 6*(2), 107-128. doi:10.3102%2F10769986006002107

Hermann, I., Haser, V., van Elst, L.T., Ebert, D., Müller-Feldmeth, D., Riedel, A., & Konieczny, L. (2013). Automatic metaphor processing in adults with Asperger syndrome: a metaphor interference effect task. *European Archives of Psychiatry and Clinical Neurosciences*, 263 (Suppl.2): S177-S187. doi:10.1007/s00406-013-0453-9

Higgins, J.P.T., Thompson, S.G., Deeks, J.J., & Altman, D.G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal*, 327(7414), 557-560

Hill, E.L. (2004). Evaluating the theory of executive dysfunction in autism. *Developmental Review*, 24(2), 189–233. doi:10.1016/j.dr.2004.01.001

Huang, S., Oi M., & Taguchi, A. (2015). Comprehension of figurative language in Taiwanese children with autism: the role of theory of mind and receptive vocabulary. *Clinical Linguistics and Phonetics,* 29(8-10), 764-775. doi:10.3109/02699206.2015.1027833

Jacobs, A.M., & Kinder, A. (2017). "The brain is the prisoner of thought": A machine-learning assisted quantitative narrative analysis of literary metaphors for use in neurocognitive poetics. *Metaphor and Symbol*, 32(3), 139-160. doi:10.1080/10926488.2017.1338015

Jodzio, K, Lojek, E., & Bryan, K. (2005). Functional and neuroanatomical analysis of extralinguistic disorders in right hemisphere-damaged patients, *Psychology of Language and Communication,* 9(1), 55-73. doi:10.3389/fpsyg.2013.00863

Kalandadze, T., Norbury, C., Nærland, T., & Næss, B K-A. (2016). Figurative language comprehension in individuals with autism spectrum disorder: a meta-analytic review. *Autism,* 22(2), 99–117. doi:10.1177/1362361316668652

Kamio, Y., & Toichi, M. (2000). Dual access to semantics in autism: is pictorial access superior to verbal access? *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 41(7), 859-867. doi:10.1017/S0021963099006137

*Kasirer, A., & Mashal, N. (2014). Verbal creativity in autism: comprehension and generation of metaphoric language in high-functioning autism spectrum disorder and typical development. *Frontiers in Human Neuroscience*, 8, 615. doi:10.3389/fnhum.2014.00615

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

Kline, A.B. (2009). *Becoming a behavioural science researcher. A guide to producing research that matters.* New York. The Guilford Press

Kwok, E.Y.L., Brown, H.M., Smyth, R.E., & Cardy, J.O. (2015). Meta-analysis of receptive and expressive language skills in autism spectrum disorder. *Research in Autism Spectrum Disorders*, 9, 202–222. doi: 10.1016/j.rasd.2014.10.008

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863

Lakoff, G, & Johnson, M. (1980). Metaphors We Live By. Language. Chicago: University of Chicago Press

Lau, J., Ioannidis, J., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *British Medical Journal,* 333, 597–600. doi:10.1136/bmj.333.7568.597

Lecce S., L. Ronchi, P. Del Sette, L. Bischetti, V. Bambini (2019) *Interpreting physical and mental metaphors: Is Theory of Mind associated with pragmatics in middle childhood?*. Journal of Child Language 46(2):393-407. doi: 10.1017/S030500091800048X

Lee, S.S., & Dapretto, M. (2006). Metaphorical vs. literal word meanings: FMRI evidence against a selective role of the right hemisphere. *NeuroImage,* 29(2), 536-544. doi: 10.1016/j.neuroimage.2005.08.003

Lewis, F.M, Murdoch, B.E., & Woodyatt G.C. (2007). Communicative competence and metalinguistic ability: Performance by children and adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders,* 37(8)*,* 1525–1538. doi:10.1007/s10803-006-0265-0

Liberati, A., Altman, D.G., Tetzlaff, J, Mulrow, C., Gøtzsche, P.C, et al. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. PLoS Med 6(7): e1000100. doi:10.1371/journal.pmed.1000100

*Mashal, N., & Kasirer, A. (2011). Thinking maps enhance metaphoric competence in children with autism and learning disabilities. *Research in Developmental Disabilities,* 32(6), 2045-2054. doi: 10.1016/j.ridd.2011.08.012

Mashal, N., Faust, M., Hendler, T., & Jung-Beeman. M. (2009). An fMRI study of processing novel metaphoric sentences. *Laterality,* 14(1), 30-54. doi: 10.1080/13576500802049433

*Melogno, S., D'Ardia, C., Pinto, M.A., & Levi, G. (2012). Explaining metaphors in high-functioning autism spectrum disorder children: A brief report. *Research in Autism Spectrum Disorders,* 6(2), 683-689. doi: 10.1016/j.rasd.2011.09.005

Melogno, S., Pinto, M.A, & Levi, G. (2012). Metaphor and metonymy in ASD children: A critical review from a developmental perspective. *Research in Autism Spectrum Disorders.* 6, (4), 1289-1296. doi.org/10.1016/j.rasd.2012.04.004

Melogno, S., Pinto, M.A., & Levi, G. (2015). Profile of the linguistic and metalinguistic abilities of a gifted child with autism spectrum disorder: A case study. *Child Language Teaching and Therapy*, 31(1), 113–126. doi:10.1177/0265659014530414clt

Melogno, S., Pinto, M. A., & Di Filippo, G. (2017). Sensory and Physico-Psychological Metaphor Comprehension in Children with ASD: A Preliminary Study on the Outcomes of a Treatment. *Brain sciences*, 7(7), 85. doi:10.3390/brainsci7070085

Moore, M., & Calvert, S. (2014). Brief Report: Vocabulary Acquisition for Children with Autism: Teacher or Computer Instruction. *Journal of Autism and Developmental Disorders,* 30(4), 359. doi:10.1023/A:1005535602064

Naigles, L.R. (ed.) (2017). *Innovative investigations of language in autism spectrum disorder*. De Gruyter, Mouton, American Psychological Association. Washington, DC

Nippold, M. A. (2016). *Later language development: School-age children, adolescents, and young adults (4th ed.)*. Austin, TX: Pro-Ed

Norbury CF (2004) Factors supporting idiom comprehension in children with communication disorders. *Journal of Speech, Language, and Hearing Research* 47(5): 1179–1193.

Norbury, C.F. (2005). The relationship between theory of mind and metaphor: evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology,* 23(3), 383-399. doi:10.1348/026151005X26732

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: A review. *Neuroscience and Biobehavioral Reviews,* 36(2), 836-854. doi: 10.1016/j.neubiorev.2011.11.008

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

Paul R., & Norbury, C. (2012). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating.* St. Louis, MO: Elsevier Health Sciences

Pennington, B.F. & Ozonoff, S. (1996). Executive Functions and Developmental Psychopathology, *Journal of Child Psychology and Psychiatry*, 37(1), 51-87. doi:10/1111/j.1469-7610.1996.tb01380.x

Perlini, C., Bellani, M., Finos, L., Lasalvia, A., Bonetto, C., Scocco, P.,… the GET UP group. (2018). Non-literal language comprehension in a large sample of first episode psychosis patients in adulthood. *Psychiatry Research,* 260, 78–89. doi: 10.1016/j.psychres.2017.11.032

Pexman, P.M., Rostad, K.R., McMorris, C.A., Climie E.A., Stowkowy J., & Glenwright M.R. (2011). Processing of ironic language in children with high-functioning autism spectrum disorder. *Journal of autism and developmental disorders,* 41(8), 1097-1112. doi:10.1007/s10803-010-1131-7

Pinto, M. A., Melogno, S., & Iliceto, P. (2008). *TCM Junior: Test di comprensione delle metafore (Scuola dell'infanzia e scuola primaria).* Roma: Carocci-Faber

Pouscoulous, N. (2011). Metaphor: for adults only? *Belgian Journal of Linguistics,* 25, 51-79. doi:10.1075/bjl.25.04pou

Pouscoulous, N. (2014). "The elevator's buttocks" metaphorical abilities in children, in Matthews, D (ed.) *Pragmatic development in first language acquisition,* John Benjamins Publishing Company, Amsterdam/Philadelphia

Rao, S.M., & Gagie, B. (2006, June/August). Learning through seeing and doing. Visual supports for children with autism. *Teaching Exceptional Children.* Retrieved from https://www.kresa.org/cms/lib/MI01000312/Centricity/Domain/135/LearningThruSeeingAnd Doing.pdf

Rapp, A.M., Felsenheimer, A.K., Langohr, K., & Klupp, M. (2018). The comprehension of familiar and novel metaphoric meanings in schizophrenia: A pilot study. *Frontiers in Psychology,* 8, 2251. doi:10.3389/fpsyg.2017.02251

Reed, P., & McCarthy, J. (2011). Cross-modal attention-switching is impaired in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 42(6), 947-53. doi:10.1007/s10803-011-1324-8

Rinaldi, M.C., Marangolo, P., & Baldassarri, F. (2004). Metaphor comprehension in right brain-damaged patients with visuo-verbal and verbal material: a dissociation (re)considered, *Cortex,* 40(3), 479-490. doi:10.1016/S0010-9452(08)70141-2

Rossetti, I., Brambilla, P., & Papagno, C. (2018). Metaphor comprehension in schizophrenic patients. *Frontiers in Psychology,* 9, 670. doi:10.3389/fpsyg.2018.00670

*Rundblad, G., & Annaz, D. (2010a). The atypical development of metaphor and metonymy comprehension in children with autism. *Autism*, 14(1), 29–46. doi:10.1177/1362361309340667

Rundblad, G., & Annaz, D. (2010b). Development of metaphor and metonymy comprehension: receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology,* 28(3), 547-563. doi:10.1348/026151009X454373

Schaeken, W., Van Haeren, M., & Bambini, V. (2018). The understanding of scalar implicatures in children with Autism Spectrum Disorder: Dichotomized responses to violations of informativeness. *Frontiers in Psychology*, 9, 1266. doi:10.3389/fpsyg.2018.01266

Schmidt, WL., & Seger, CA. (2009). Neural correlates of metaphor processing: The roles of figurativeness, familiarity and difficulty, *Brain and Cognition,* 7(3), 375-386. doi.org/10.1016/j.bandc.2009.06.001

Simohsohn, U., Nelson, L.D., & Simmons, J.P. (2014). *p*-Curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science,* 9(6), 666–681. doi:10.1177/1745691614553988

Siqueira M., Marques D.F., & Gibbs, R.W. Jr. (2016). Metaphor-related figurative language comprehension in clinical populations: a critical review. *SCRIPTA, Belo Horizonte,* 20(40), 36-60

Steen, G., Dorst, A., & Hermann, J. (2010). Metaphor in usage. *Cognitive Linguistics*, 21(4), 765-796. doi:10.1515/cogl.2010.024

Sperber, D and Wilson, D. (2008). A deflationary account of metaphors. In R. Gibbs (ed.), *The Cambridge Handbook of Metaphor and Thought,* 84-105. Cambridge: Cambridge University Press; Reprinted in Wilson, D. & Sperber, D. (2012). *Meaning and Relevance*, pp. 97-122. Cambridge: Cambridge University Press

Tager-Flusberg, H. (1992). Autistic children's talk about psychological states: Deficits in the early acquisition of a theory of mind. *Child Development,* 63(1), 161-172. doi:10.1111/j.1467-8624.1992.tb03604.x

Tager-Flusberg, H., Calkings, S., Nolin, T., Baumberger, T., Anderson, M., & Chadwick-Dias, D. (1990). A longitudinal study of language acquisition in autistic and Down's syndrome children. *Journal of Autism and Developmental Disorders,* 20(1), 1-21. doi:10.1007/BF02206853

Thibodeau, P.H., Sikos, L., & Durgin, F.H. (2018). Are subjective ratings of metaphors a red herring? The big two dimensions of metaphoric sentences. *Behavior Research Methods,* 50(2), 759-772. doi:10.3758/s13428-017-0903-9

Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools*, 44(5), 423–432. doi:10.1002/pits.20234

Utsumi, A., & Sakamoto, M. (2007). Computational evidence for two-stage categorization as a process of adjective metaphor comprehension. *In Proceedings of the 2nd European Cognitive Science Conference (EuroCogSci2007)*:77–82

Utsumi, A., & Sakamoto, M. (2011). Indirect categorization as a process of predicative metaphor comprehension. *Metaphor and Symbol*, 26(4), 299–313. doi:10.1080/10926488.2011.609120

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

Varga, E., Schnell, Z., Tényi, T., Németh, N., Simon, M., Hajnal, A., ... Herold, R. (2014). Compensatory effect of general cognitive skills on non-literal language processing in schizophrenia: a preliminary study. *Journal of Neurolinguistics,* 29, 1–16. doi: 10.1016/j.jneuroling.2014.01.001

Vermeulen, P. (2014). Context blindness in autism spectrum disorder: Not using the forest to see the trees as trees. *Focus on Autism and Other Developmental Disabilities,* 30(3), 1-11. doi:10.1177/1088357614528799

Vosniadou, S., Ortony, A., Reynolds, R.E., & Wilson, P.T. (1984). Sources of difficulty in the young child's understanding of metaphorical language. *Child Development,* 55(4), 1588-1606. doi:10.2307/1130028

Vulchanova M, Saldaña D, Chahboun S, et al. (2015) Figurative language processing in atypical populations: the ASD perspective. *Frontiers in Human Neuroscience,* 9: 24. doi: 10.3389/fnhum.2015.00024

Vulchanova, M., Milburn, E., Vulchanov, V, Baggio G. (2019). Boon or Burden? The Role of Compositional Meaning in Figurative Language Processing and Acquisition. Journal of logic, language and information, 1-29. doi.org/10.1007/s10849-019-09282-7

Wilson, D. (2011). Parallels and differences in the treatment of metaphor in relevance theory and cognitive linguistics. Intercultural Pragmatics. 8(2):177–196. doi: 10.2478/v10148-011-0025-1

Wilson, D., & Carston, R. (2006). Metaphor, Relevance and the 'Emergent Property'. *Mind & Language,* 21(3)*, * (404-433). doi:10.1111/j.1468-0017.2006.00284

Winner, E. (1988). *The point of words: Children's understanding of metaphor and irony*. USA: Harvard University Press.

Winner, E., Rosenstiel, A. K., & Gardner, H. (1976). The development of metaphoric understanding. *Developmental Psychology*, 12(4), 289–97

World Health Organization. (1992). The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. Geneva: World Health Organization. http://www.who.int/iris/handle/10665/37958

*Zheng, Q., Jia Z., & Liang, D. (2015). Metaphor and metonymy comprehension in Chinese-speaking children with high-functioning autism. *Research in Autism Spectrum Disorders*, 10, 51-58. doi: 10.1016/j.rasd.2014.11.007

Table 1. Examples of metaphor task properties taken from the included studies

	Task property	Options	Example
Linguistic characteristics of the stimuli	Familiarity	Conventional	'There is a *flood* outside the museum' (Rundblad & Annaz, 2010a)
		Novel	'The house has a *hat'* (Melogno et al., 2012)
	Linguistic structure	Nominal	'The moon is a *light bulb'* (Melogno et al., 2012)
		Predicate	'When Taro plays soccer, no one at his school *comes close to him'* (Adachi et al., 2004)
		Adjective-Noun pair	*'Sharp tongue'* (Kasirer & Mashal, 2014)
	Linguistic context	Minimal	'The *sky's scarf* is colored' (Melogno et al., 2012)
		Discourse	'Stuart works at a museum. The museum is in the middle of town near a big river. It is a small museum and not so many people come to the museum. Stuart's boss wants more people to come to the museum. So Stuart prepares a very special exhibition. Stuart's boss tells lots and lots of people about Stuart's exhibition. It is Monday morning and Stuart is at home. Suddenly, the phone rings; it is Stuart's boss. Stuart's boss says: 'You did it Stuart! There is a *flood* outside the museum'. Stuart runs to the museum to look. What does Stuart see?' (Rundblad & Annaz, 2010a)
	Modality	Auditory	*'Some roads are ribbons'* (Chouinard & Cummine, 2016). Participants heard a sentence
		Written	'When Taro plays soccer, no one at his school *migi ni deru* (*comes close to him*)' (Adachi et al., 2004). Children were asked to read each question silently
		Multimodal	A story for '*KNOWING IS SEEING'* conceptual metaphor: 'Kristin is trying to make cookies. She doesn't know how to make them. Her cookie dough looks wrong. Her mom teachers her how to make the dough. Kristin says, *"Now I view it!"*'* (Olofson et al., 2014). After hearing (*auditory*) the metaphoric utterance participants were asked which picture (*pictorial*) displayed the meaning of the target utterance. At the same time, a question mark was displayed on the screen.
Response format		Multiple choice	'When Taro plays soccer, no one at his school *migi ni deru (comes close to him)'* (Adachi et al., 2004) Choices: 'Taro… Is the best soccer player in the school Is the worst soccer player in the school Is sitting to the right of all the students Taro thought he was going to play soccer Don't know'

		Children were asked to read each question silently
	Verbal explanation	Instructions to the participants: "When I say, *'Is the sun a ball'*? you will tell me what you think that means. Or when I say, *'Does the sun have arms*?'*, then you will say what this means to you. Right, now it's your turn". 'The *sky's scarf* is colored' (Melogno et al., 2012)
	Multiple-choice combined with verbal justification	Ten metaphors are incorporated into short sentences. The sentences are printed on a card in the Written Metaphor test. Each sentence is followed by three possible responses: the correct metaphorical (the target), the literal (a concrete type of incorrect choice), and the inappropriate meaning (another type of incorrect choice). The participant is asked to listen to the metaphorical sentence and then to point to the one that explains it. After making the choice, participant is requested to give his/her own interpretation of the metaphor (Gunter et al., 2002)
	Meaningful-ness decision	Participants were instructed to read the words, and indicate as rapidly and accurately as possible whether the target expression was meaningful by lifting and moving the right index finger from the middle position to the right or left keys (Gold & Faust, 2010)

Note. Original items extracted from the studies are enclosed within single quotation marks with metaphoric expressions italicized. The instruction directly cited is enclosed within double quotation marks.

* The expression "Now I view it" is used as a novel version of the conceptual mapping KNOWING IS SEEING, as opposed to the lexicalized version "Now I see it".

Table 2. Properties of the tasks employed in the included studies

Study	Response format	Linguistic characteristics 1) **Metaphor familiarity** 2) **Syntactic structure** 3) **Linguistic context** 4) **Modality**	Effect size (*g*)	95% CI	Language
Adachi et al. (2004)	Multiple-choice	1) Not specified (The words and sentences were selected from standard textbooks) 2) Mixed/not specified 3) Short stories 4) Written	-0.51	[-0.81, -0.21]	Japanese
Chounard & Cummine (2016)	Meaningful-ness decision	1) Not specified 2) Nominal 3) Sentence 4) Auditory, computer-based	-0.45	[-1.22, 0.31]	English
Gunter et al. (2002)	Multiple-choice (written) combined with verbal justification	1) Conventional 2) Not specified 3) Sentence 4) Auditory	-1.14 (combi-ned)	[-2.17, -0.11]	English
Gunter et al. (2002) (Same study as above)	Multiple-choice (picture) combined with verbal justification	1) Conventional 2) Not specified 3) Sentence 4) Auditory	See above		English
Gunter et al. (2002) (Same study as above)	Meaningful-ness decision	1) Novel (unusual) 2) Mostly nominal 3) Sentence 4) Not specified	See above		English
Gold & Faust (2010)	Meaningful-ness decision	1) Novel; conventional 2) Word pairs 3) No context 4) Written, computer-based	-0.52	[-1.02, -0.01]	Hebrew

Hermann et al. (2013)	Meaningful-ness decision	1) Novel 2) Nominal 3) No context 4) Written, computer-based	-0.39	[-1.00, 0,23]	German
Huang et al. (2015)	Multiple-choice	1) Not specified 2) Mixed/Not specified 3) Short stories 4) Written	0.52	[-0.92, -0.13]	Taiwanese
Kasirer & Mashal (2014)	Multiple-choice	1) Novel; conventional 2) Word pairs 3) No context 4) Not specified	0.42	[-1.09, 0.26]	Hebrew
Kasirer & Mashal (2016)	Multiple-choice	1) Novel; conventional 2) Word pairs 3) No context 4) Not specified *The same task as in Kasirer & Mashal (2014)*	-0.46	[-0.92, -0.02]	Hebrew
Mashal & Kasirer (2011)	Multiple-choice	1) Novel; conventional 2) Word pairs 3) No context 4) Not specified *The same task as in Kasirer & Mashal (2014) and Kasirer & Mashal (2016)*	-0.76	[-1.40, -0.13]	Hebrew
Melogno et al. (2012)	Verbal explanation	1) Novel 2) Sentences 3) 12 metaphors included in sentences, and 13 metaphors contextualized in four stories 4) Not specified *Junior Metaphor Comprehension Test. Designed for the specific age-range and validated for use with a pediatric population*	-0.61	[-1.18, -0.43]	Italian
Olofson et al. (2014)	Multiple-choice	1) Primary conceptual novel and lexicalized 2) Sentence with metaphorical verbs or adjectives. Ex: 'Now I view it!' or 'Now I see it'	-0.91	[-1.69, -0.12]	English

|					
		3) Two five-sentence stories 4) Auditory, computer-based			
Rundblad & Annaz (2010a)	Verbal/ Open question	1) Conventional 2) Sentences 3) Short picture stories 4) Auditory	-2.20	[-3.14, -1.27]	English
Zheng et al. (2015)	Multiple-choice	1) Conventional; novel 2) Nominal 3) Short stories and pictures 4) Written and illustrations	-0.71	[-1.43, 0.02]	Chinese
De Villiers et al. (2011)	Multiple-choice combined with verbal response	1) Not specified 2) Word pairs 3) Short picture stories 4) Not specified	-0.84	[-1.37, -0.31]	English

Figure 1
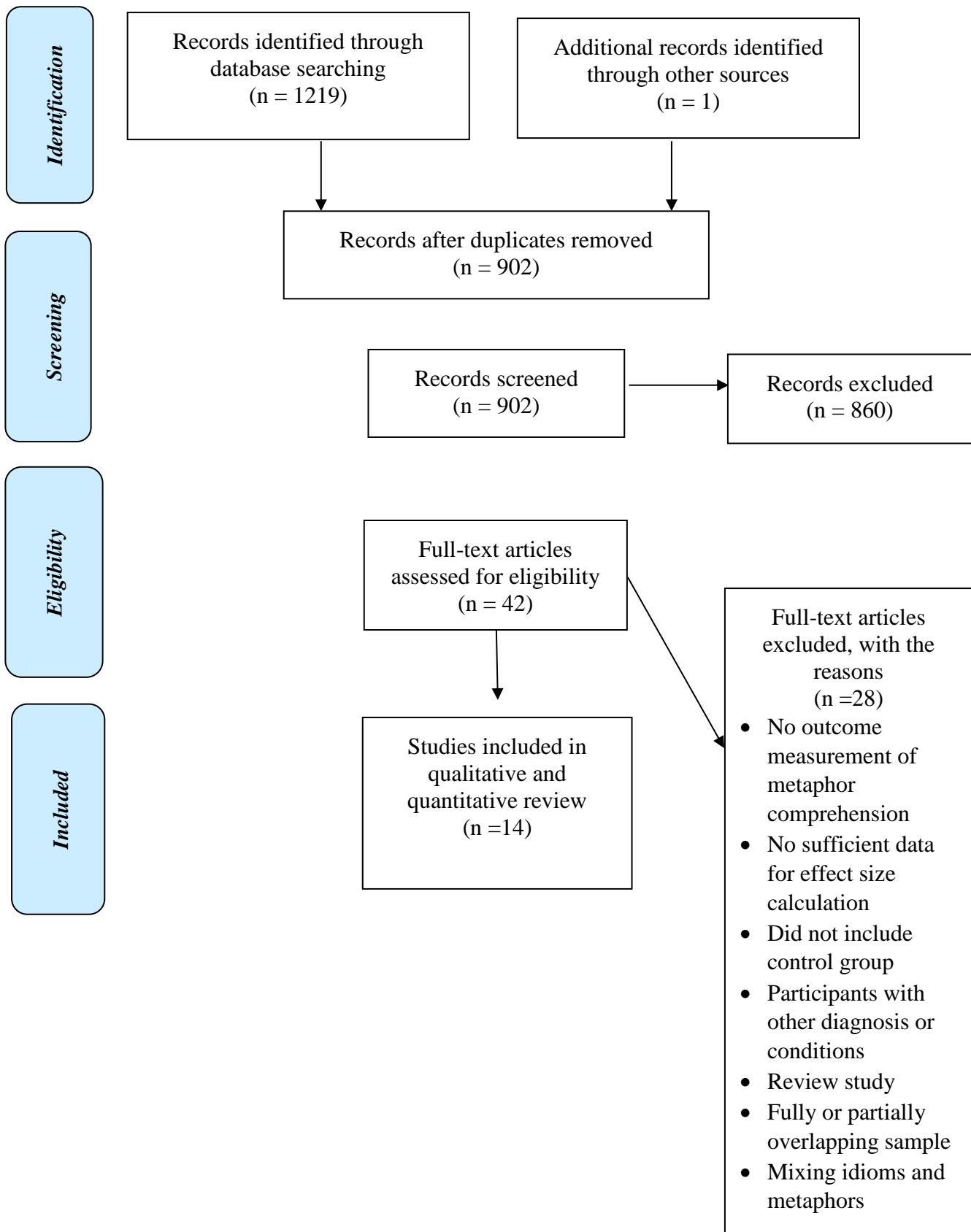
Figure 1. Flow chart for the screening and inclusion of studies according to the PRISMA statement.
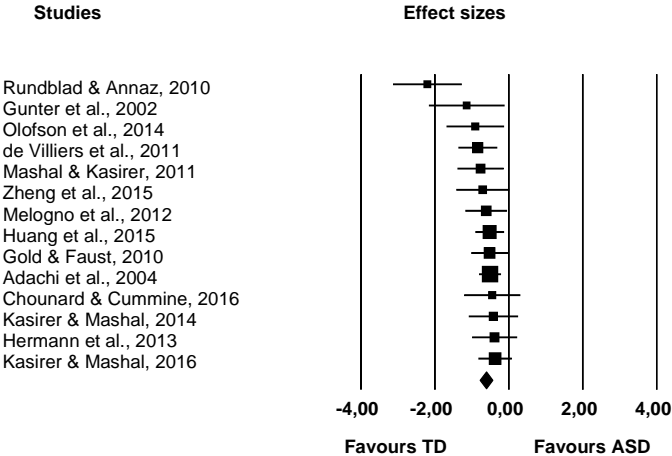
Figure 2

Figure 2. Overall average effect size for metaphor comprehension and the effect size with confidence intervals for each study comparing individuals with ASD and TD

Appendix I:

Metaphor task

<u>**Metaforoppgave**</u>

<u>Til testleder:</u>

Det er nødvending med lydopptak av hele oppgaven. Du skal starte med å si dato og sted for undersøkelsen slik at opptaker får et eget nummer.

Kort om oppgaven: denne flervalgsoppgaven består av metaforiske og bokstavelige utsagn. Det følger tre svaralternativer etter hver oppgave.

Viktig: Du skal KUN kommentere prøveoppgaven. Det er imidlertid tillatt å oppmuntre deltakeren i forhold til god innsats. Når deltakerens svar er ufullstendig eller uklart er det tillatt å stille oppfølgingsspørsmål kun for å få mer informasjon (**dette er viktig**), men ikke for å rette opp et evt. feilaktig svar. Du kan bruke uttrykk, som, for eksempel:

   *-Kan du fortelle mer? eller **Hva mener du?***

Både instruksjoner og oppgaver skal gjentas så ofte som deltakeren ønsker det.

Si til deltaker:

-Nå *skal jeg først lese to setninger, så skal du velge et av tre svaralternativene som du synes passer best til BEGGE setningene (**Det er altså viktig at deltakeren forstår at han/hun skal tenke på hele oppgaven og ikke bare den siste setningen når de skal velge et svaralternativ**)*

--Etter hver oppgave skal jeg spørre deg om å fortelle hvorfor du synes valget ditt passet best til setningene

- Si ifra hvis jeg skal lese enten setninger eller svaralternativene en gang til.

Nå prøver vi en oppgave. Hør nøye etter:

Hanne er ofte på besøk hos Kari. Kari er varm

1) Kari er kjærlig

2) Kari har høy kroppstemperatur

3) Kari er sulten

1. **Skuespillerinnen gikk nettopp inn i teatret. Hun er en stjerne**

   1) Skuespillerinnen er et himmellegeme

   2) Skuespillerinnen er berømt

   3) Skuespillerinnen er tørst

2. **Daniel lytter etter lyder. Det er en storm**

   1) Det er en flink skoleelev

   2) Det er en kraftig vind

   3) Det er et bortskjemt enebarn

3. **Anne ser seg selv i speilet. Frisøren hennes er en tryllekunstner**

   1) Frisøren hennes er ansatt på sirkus

   2) Frisøren hennes er treg

3) Frisøren hennes er flink

4. Elisabeth tenker ikke før hun sier noe. Ordene hennes er sverd

1) Ordene hennes er et redskap

2) Ordene hennes er sårende

3) Ordene hennes er morsomme

5. Per prater ofte med norsklæreren. Norsklæreren er et leksikon

1) Norsklæreren vet alt

2) Norsklæreren er ung

3) Norsklæreren er et oppslagsverk

6. Elena har vært gift med Jens i femten år. Jens er en perle

1) Jens er en musling

2) Jens er en farlig mann

3) Jens er en god mann

7. Peter tenker at han ser en hund. Det er en rev

1) Peter ser et rovdyr

2) Peter ser en student

3) Peter ser en musikklærer

8. Nina ser en figur. Figuren er en engel

1) Nina ser en vegg

2) Nina ser vinger

3) Nina ser en bok

9. Hanne finner sin favoritt smykke i butikken. Det er en perle

   1) Hanne finner en musling

   2) Hanne finner salat

   3) Hanne finner en traktor

10. Peter gir aldri opp. Han er en rev

   1) Peter er et rovdyr

   2) Peter er lur

   3) Peter er musikklærer

11. Steinar ser seg selv i speilet. Han er en løve

   1) Steinar har langt hår

   2) Steinar er ansatt i barnehagen

   3) Steinar er et kattedyr

12. Willhelm får alltid mye ros. Han er en gepard

   1) Wilhelm er rask

   2) Wilhelm er et rovpattedyr

   3) Wilhelm er en tenåring

13. Liv hører et drønn. Det er en vulkan

   1) Det er et ildsprutende fjell

   2) Det er en bok

   3) Det er en butikk

14. Nico hører kurring. Det er en due

    1) Nico hører en mann med frakk

    2) Nico hører en fugl med klatreøetter

    3) Nico hører en dame med hatt

15. Lea får akkurat den gaven hun har ønsket seg lenge. Det er en katt

1) Lea får en bok

2) Lea får en blyant

3) Lea får et kjæledyr

16. Den unge skiløperen er stadig på TV. Han er en hane

    1) Den unge skiløperen er en hønsefugl

    2) Den unge skiløperen er stolt

    3) Den unge skiløperen er lat

17. Magnus tror han ser en kniv. Det er et sverd

    1) Magnus ser et fjell

    2) Magnus ser en hytte

    3) Magnus ser et redskap

18. Katten fanger et lite dyr. Det er en mus

    1) Katten fanger en gnager

    2) Katten fanger et papir

    3) Katten fanger en bok

19. Wilhelm hører knurring. Det er en gepard

1) Wilhelm hører en sang

2) Wilhelm hører en mann

3) Wilhelm hører et rovpattedyr

20. Nina snakker i telefonen. Stemmen hennes er en sirene

    1) Stemmen hennes er høy

    2) Stemmen hennes er et varslingsinstrument

    3) Stemmen hennes er søt

21. Lykke peker alltid på det samme bildet i boken. Det er en gris

    1) Lykke peker på et bondegårdsdyr

    2) Lykke peker på et tog

    3) Lykke peker på et tre

22. Gutten ser at noen sover i gresset. Det er en hare

    1) Gutten ser et dyr

    2) Gutten ser en penn

    3) Gutten ser en lampe

23. Berit hører lyder. Det er en hakkespett

    1) Det er en spettefugl

    2) Det er en seng

    3) Det er en dyne

24. Ingvild får vite hva som hadde skjedd med Jens. Nyheten er et jordskjelv

    1) Nyheten er et sjokk

    2) Nyheten er en bølgebevegelse i bakken

3) Nyheten er hyggelig

25. Elisabeth slutter å prate med Nina. Nina er en storm

1) Nina er rasende

2) Nina er en kraftig vind

3) Nina er en paraply

26. Pappa hadde ventet på meg hele kvelden. Da jeg kom hjem, var pappa en vulkan

1) Pappa var sint

2) Pappa var et ildsprutende fjell

3) Pappa var mett

27. Sofia åpner bursdagsgaven hun fikk av mamma. Det er et maleri

1) Bursdagsgaven er et bilde

2) Bursdagsgaven er en dukke

3) Bursdagsgaven er en bok

28. Mannen underholder barna. Han er en tryllekunstner

1) Mannen er ansatt på sirkus

2) Mannen er huseier

3) Mannen er leietaker

29. Morten prater sjelden med sin nærmeste nabo. Naboen er kald

1) Naboen er følelsesløs

2) Naboen er smart

3) Naboen fryser

30. Peter ser at noe beveger seg i skogen. Det er en bjørn

1) Peter ser et stort dyr

2) Peter ser en ung branmann

3) Peter ser et gammelt tre

31. Russland får ny president. Han er en due

1) Russlands nye president er en fredsskapende person

2) Russlands nye president er en svømmer

3) Russlands nye president er en fugl med klatreføtter

32. Aleksander ser en stor planteeter. Det er en elefant

1) Aleksander ser et savannedyr

2) Aleksander ser en blyant

3) Aleksander ser en sol

33. Magnus er tydeligvis fortsatt på rommet. Han er en mus

1) Magnus er en stille person

2) Magnus er en gnager

3) Magnus er glad

34. Oscar ser at noen angriper kalven. Det er en løve

1) Oscar ser et kattedyr

2) Oscar ser en gutt

3) Oscar ser en sanger

35. Trygve ser en prikk på himmelen. Det er en stjerne

1) Trygve ser en kopp

2) Trygve ser et himmellegeme

3) Trygve ser en hest

36. Jasper hører at noe rører seg i buret. Det er en hane

1) Det er en vinduskarm

2) Det er en hønsefugl

3) Det er en husnøkkel

37. Kjæresten til Lars heter Nina. Hun er en engel

1) Nina er snill

2) Nina har vinger

3) Nina er gartner

38. Peter fylte leiligheten sin med møbler. Han er en bjørn

1) Peter er sterk

2) Peter er et stort dyr

3) Peter er barnehagelærer

39. Anna får en hund av pappa. Hunden er en elefant

1) Hunden er stor

2) Hunden er blind

3) Hunden er et savannedyr

40. Ludvig er en tynn liten gutt. Han er en hare

1) Ludvig er redd

2) Ludvig er et dyr

3) Ludvig er fornøyd

41. Katja er en ny elev i klassen min. Hun er et maleri

1) Katja er vakker

2) Katja er et tre

3) Katja er et bilde

42. Ingrid har vært ute lenge. Hun er kald

1) Ingrid fryser

2) Ingrid sover

3) Ingrid spiser

43. Stig sitter og spiser grøt. Han er en gris

1) Stig er et bondegårdsdyr

2) Stig er lærer

3) Stig søler mye

44. Lyden skremmer Marit. Det er en sirene

1) Det er et varslingsinstrument

2) Det er en bordlampe

3) Det er et veggdyr

45. Plutselig rister huset. Det er et jordskjelv

1) Det er en bølgebevegelse i bakken

2) Det er en interessant film

3) Det er en god nyhet

46. Lea setter jogaundervisningen høyt på timeplanen. Lea er en katt

1) Lea er et kjæledyr

2) Lea er smidig

3) Lea er en paraply

47. Berit legger merke til en tykk bok i hyllen. Det er et leksikon

1) Berit legger merke til en sofa

2) Berit legger merke til et oppslagsverk

3) Berit legger merke til en paraply

48. Snekkeren arbeider ved siden av meg. Han er en hakkespett

1) Snekkeren er bråkete

2) Snekkeren er ung

3) Snekkeren er en spettefugl

Appendix II:

Information letter and request to participation

II

Forespørsel om deltakelse i forskningsprosjektet

"Språkforståelse hos barn og ungdommer med autismespekterforstyrrelser(ASF)/Asperger syndrom og hos barn og ungdommer med typisk utvikling"

Bakgrunn og formål med studien

Doktorgradsstudien «*Språkforståelse hos barn og ungdommer med autismespekterforstyrrelser (ASF)/Asperger Syndrom og hos barn og ungdommer med **typisk utvikling**"* er tilknyttet Institutt for Spesialpedagogikk og forskergruppen Child Language and Learning ved Universitetet i Oslo. Studien gjennomføres i samarbeid med Nasjonalt kompetansesenter for ADHD, autisme, Tourettes syndrom og hypersomnier. En annen viktig samarbeidspartner er Språktilegnelses- og Språkprosesseringslaben ved Norges teknisk-naturvitenskapelige Universitet. Vitenskapelig ansvarlige er førsteamanuensis Kari-Anne Næss, forsker/psykolog, ph.d, Terje Nærland, prof. Mila Vulchanova og forsker, ph.d Ingrid Lossius Falkum.

Studien fokuserer spesielt på forståelse av metaforiske uttrykk. Metafor er et ord eller et uttrykk som brukes i overført betydning. Vi bruker metaforer både når vi snakker og når vi skriver, ofte uten at vi er bevisst på dem. Dessuten møter vi ofte metaforer i skolebøker, og de brukes også hyppig av lærere i undervisningssammenheng. Med andre ord spiller forståelse av metaforer en viktig rolle i kommunikasjon med andre mennesker og også for kunnskapstilegnelse i skolen.

Formålet med denne studien er å skaffe mer kunnskap om språkforståelse hos barn og ungdommer med og uten autismespekterforstyrrelser for å kunne tilrettelegge for en god språkstimulering på skolen, noe som igjen kan ha avgjørende betydning for barns og unges muligheter til å forstå språk. God språkforståelse kan bidra til trivsel og bedre livskvalitet for disse personene.

Hva innebærer deltakelse i studien?

Deltakelse i studien innebærer følgende:

1) Barnet/ungdommen vil gjøre ulike språkoppgaver sammen med undertegnede. Oppgavene vil bli presentert enten på dataskjerm, på papir eller som praktiske «gjøreoppgaver». Et eksempel på en oppgave er f.eks. at barnet/ungdommen skal enten peke eller si nummeret på det bildet som best representerer mening til det ordet han/hun hørte. Det skal gjøres lydopptak kun en av språkoppgavene. I tillegg til språkoppgaver vil det gjennomføres en oppgave som omhandler generell strategibruk. Alle oppgavene vil tilsammen ta ca. 120 minutter. Oppgavene kan enten gjøres samme dag eller fordeles over to dager for å unngå at barnet/ungdommen blir sliten. Det skal sørges for at testsituasjonen vil være lystbetont og gøy. Barnet/ungdommen vil få tilpasset informasjon om hva han/hun skal gjøre før oppgavene presenteres.

Datainnsamlingen utføres av undertegnede og vil foregå i et stille rom på skolen. Hvis ønskelig, kan eventuelt forelder eller lærer/assistent være til stede i arbeidsøktene. Innsamlede data registreres på datamaskinen og som notater.

2) Som en del av datainnsamlingen, vil det bli sendt ut to spørreskjemaer til foreldrene. Det ene spørreskjemaet består av 70 utsagn og det andre spørreskjemaet består av 65 utsagn. Foreldrene vil bli bedt om å vurdere hvor ofte de har observert et bestemt utsagn hos barnet/ungdommen.

Hvem blir invitert til å delta i studien?

Det skal rekrutteres barn og ungdommer innenfor to ulike grupper (10-16 år):

1) Barn/ungdom med Autismespekterforstyrrelser (ASD), f. eks Asperger syndrom
2) Barn/ungdom med typisk utvikling

Alle deltakere skal ha norsk som sitt første språk. Utvalget vil rekrutteres via skoler, Autismeforeningen og andre relevante instanser. Disse instansene videresender samtykkeskrivet til foreldrene. Identiteten til dem som forespørres vil ikke være kjent undertegnede før foreldrene eventuelt har samtykket til deltakelse.

Hva skjer med informasjonen vi samler inn?

Alle personopplysninger og data fra studien vil bli behandlet konfidensielt. Personindentifiserende opplysninger erstattes med en kode. Personopplysninger vil bli lagret adskilt fra øvrige data for å ivareta konfidensialitet og det er kun undertegnede som har adgang til navneliste/andre personopplysninger. Deltakerne vil heller ikke kunne gjenkjennes i publikasjoner. Lydopptaket skal oppbevares sikkert. Studien skal etter planen avsluttes 30/06/17 og alle innsamlede personopplysninger inkludert lydopptak vil da bli destruert.

Deltakelse er frivillig

Det er frivillig å delta i studien. Hvis dere senere skulle ombestemme dere, er det mulig å trekke seg fra deltakelse. Dere behøver eventuelt ikke å oppgi noen grunn for at dere trekker dere. Dersom du velger at barnet ditt skal trekke seg underveis, vil alle innsamlede opplysninger inkludert evt lydopptak bli slettet.

Studien er godkjent av Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

Dersom du ønsker at barnet ditt skal delta eller har spørsmål til studien, ta kontakt med doktorgradsstipendiat Tamar Kalandadze på tlf: 93606087 eller e-post: tamar.kalandadze@isp.uio.no

Hilsen
Tamar Kalandadze

Doktorgradsstipendiat
Institutt for Spesialpedagogikk
Universitetet i Oslo

Adresse:
Boks 1072 Blindern
0316 Oslo

Samtykke til deltakelse i studien

Hvis dere ønsker å delta i studien «Språkforståelse hos barn og ungdommer med autismespekterforstyrrelser (ASF)/Asperger syndrom og hos barn og ungdommer med typisk utvikling», ber vi dere signere samtykkeerklæringen. Signering vil innebære at du bekrefter å ha mottatt informasjon om studien og du er villig til at barnet ditt deltar.

Signert av (f.eks. mor, far):

Dato:

Barnets navn:

Ditt telefonnummer:

Din mailadr.:

Appendix III:

Ethical approval from the Norwegian Social Science Data Services (NSD)

III

Tamar Kalandadze
Institutt for spesialpedagogikk Universitetet i Oslo
Postboks 1140 Blindern
0318 OSLO

Vår dato: 27.11.2015   Vår ref: 45346 / 3 / MSI   Deres dato: Deres ref:

TILBAKEMELDING PÅ MELDING OM BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 26.10.2015. Meldingen gjelder prosjektet:

45346	*Forståelse av metaforer hos barn og ungdommer*
Behandlingsansvarlig	*Universitetet i Oslo, ved institusjonens øverste leder*
Daglig ansvarlig	*Tamar Kalandadze*

Personvernombudet har vurdert prosjektet, og finner at behandlingen av personopplysninger vil være regulert av § 7-27 i personopplysningsforskriften. Personvernombudet tilrår at prosjektet gjennomføres.

Personvernombudets tilråding forutsetter at prosjektet gjennomføres i tråd med opplysningene gitt i meldeskjemaet, korrespondanse med ombudet, ombudets kommentarer samt personopplysningsloven og helseregisterloven med forskrifter. Behandlingen av personopplysninger kan settes i gang.

Det gjøres oppmerksom på at det skal gis ny melding dersom behandlingen endres i forhold til de opplysninger som ligger til grunn for personvernombudets vurdering. Endringsmeldinger gis via et eget skjema, http://www.nsd.uib.no/personvern/meldeplikt/skjema.html. Det skal også gis melding etter tre år dersom prosjektet fortsatt pågår. Meldinger skal skje skriftlig til ombudet.

Personvernombudet har lagt ut opplysninger om prosjektet i en offentlig database, http://pvo.nsd.no/prosjekt.

Personvernombudet vil ved prosjektets avslutning, 30.06.2017, rette en henvendelse angående status for behandlingen av personopplysninger.

Vennlig hilsen

Katrine Utaaker Segadal

Marte Byrkjeland

Kontaktperson: Marte Byrkjeland tlf: 55 58 36 01
Vedlegg: Prosjektvurdering

Dokumentet er elektronisk produsert og godkjent ved NSDs rutiner for elektronisk godkjenning.

Personvernombudet for forskning

Prosjektvurdering - Kommentar

Formålet er å undersøke forståelse av metaforer hos barn og ungdommer med typisk utvikling og hos barn og ungdommer med Autismespekterforstyrrelser (ASF).

Utvalget består av barn i aldersgruppen 10-12 år både med Autismespekterforstyrrelser (ASF) og med typisk utvikling.

Utvalget (v/foresatte) informeres skriftlig om prosjektet og samtykker til deltakelse. Informasjonsskrivet er godt utformet, men det bør informeres om hvor og i hvilken sammenheng datainnsamlingen gjøres. Det må legges til rette for elever som ikke ønsker å delta, dersom data samles inn i undervisningssammenheng. Videre bør det da planlegges et alternativt opplegg for de som ikke deltar.

Merk at når barn skal delta aktivt, er deltagelsen alltid frivillig for barnet, selv om de foresatte samtykker. Barnet bør få alderstilpasset informasjon om prosjektet, og det må sørges for at de forstår at deltakelse er frivillig og at de når som helst kan trekke seg dersom de ønsker det.

Personvernombudet vurderer at det vil behandles sensitive personopplysninger om helseforhold, på grunn av utvalgskriteriene.

Personvernombudet legger til grunn at forsker etterfølger Universitetet i Oslo sine interne rutiner for datasikkerhet.

Forventet prosjektslutt er 30.06.2017. Ifølge prosjektmeldingen skal innsamlede opplysninger da anonymiseres. Anonymisering innebærer å bearbeide datamaterialet slik at ingen enkeltpersoner kan gjenkjennes. Det gjøres ved å:
- slette direkte personopplysninger (som navn/koblingsnøkkel)
- slette/omskrive indirekte personopplysninger (identifiserende sammenstilling av bakgrunnsopplysninger som f.eks. bosted/arbeidssted, alder og kjønn).