Article:
Davies, CN orcid.org/0000-0001-9347-7905, Andres-Roqueta, C and Norbury, CF (2016) Referring expressions and structural language abilities in children with Specific Language Impairment: A pragmatic tolerance account. Journal of Experimental Child Psychology, 144. pp. 98-113. ISSN 0022-0965

https://doi.org/10.1016/j.jecp.2015.11.011

This paper was accepted on 19/11/2015 and will appear as:

Davies, C., Andrés-Roqueta, C. & Norbury, C.F. (2016) Referring expressions and structural language abilities in children with Specific Language Impairment: A pragmatic tolerance account. *Journal of Experimental Child Psychology.*

Abstract

Specific language impairment (SLI) has traditionally been characterised as a deficit of structural language (specifically grammar), with relative strengths in pragmatics. In this study, comprehensive assessment of production, comprehension and metalinguistic judgment of referring expressions reveals that children with SLI have weaknesses in both structural and pragmatic language skills relative to age-matched peers. Correlational analyses highlight a relationship between their performance on the experimental tasks and their structural language ability. Despite their poor performance on the production and comprehension tasks, children with SLI were able to recognise pragmatically under-informative reference relative to other types of utterance, though imposed a less severe penalty on such expressions than their typically-developing peers; a pattern which supports the pragmatic tolerance account. Our novel methodology (which probes structural abilities from both the speaker's and hearer's perspectives as well as metalinguistic and pragmatic skills in the same sample) challenges the assumption that pragmatic errors stem from deficits in social cognition and instead supports recent findings suggesting that when the impact of structural language is isolated, pragmatic deficits may be resolved.

Reference is a central communicative skill in which speakers must identify the entity they wish to talk about and recognise how to describe the referent in a way that will unambiguously identify the intended referent for the listener. Thus, successful reference requires integrating cognitive, linguistic and social-pragmatic skills in order to scan the visual scene, identify distinctive features of the target relative to other potential referents, take account of previous and concurrent linguistic and environmental context, understand the listener's perspective and knowledge state, provide sufficient detail to disambiguate similar items, and then produce a felicitous referring expression (Ariel, 1990; Chafe, 1976; Clark & Bangerter, 2004; Clark, Schreuder, & Buttrick, 1983; Clark & Wilkes-Gibbs, 1986). For example, an optimally-informative speaker should refer to 'the small apple' when there are two apples of different sizes in the context; 'apple' alone is under-informative in the given context and creates ambiguity and potential for communicative breakdown.

Although the development of reference in typically developing children has been researched extensively since the 1960s (for reviews see Dickson, 1982 and Graf & Davies, 2014), the development of referential communication in clinical populations in which component skills such as language or social-pragmatics may be compromised has been under-researched. This is particularly so for children with specific language impairments (SLI), although such children may provide crucial evidence about the underlying skills associated with successful reference. SLI is a common developmental condition, diagnosed when children fail to develop aspects of structural language, e.g. phonology, vocabulary and syntax, despite meeting other developmental milestones (Norbury & Paul, 2013). Traditionally, children with SLI were considered to have strengths in pragmatic language skills, or the use of language for social communication purposes, relative to their structural language abilities. In general, deficits in social-cognition are not thought to characterise SLI (van der Lely, 1997) and their conversational skills are typically appropriate to context (Bishop & Adams, 1991). However, more recent investigations have highlighted delays in social-cognitive understanding (Andrés-Roqueta, Adrian, Clemente, & Katsos,

2013), lower ratings of pragmatic language competence (Norbury, Nash, Baird, & Bishop, 2004), and deficits in pragmatic tasks that require analysis of linguistic context, for example, inferencing and ambiguity resolution (Norbury, 2005a,b). In general, the pragmatic skills of children with SLI are reported to be commensurate with their overall level of linguistic competence (cf. Norbury, Nash, Baird, & Bishop, 2004). This highlights the importance of structural language skills in both solving pragmatic language tasks, and potentially in developing social-pragmatic awareness. In addition, investigations into the pragmatic language skills of children with SLI may elucidate sources of pragmatic breakdown in other developmental populations, such as autism spectrum disorder.

The current study thus aims to examine reference skills in SLI using a novel and comprehensive paradigm probing production, comprehension and metalinguistic abilities and to explore the potential origins of pragmatic deficit in reference by examining the relationship between our novel tasks and standardised language tests. Reference is a potentially fruitful arena to test the pragmatic skills of children with SLI, as the pragmatic element of this task draws more on an understanding of the listener's knowledge state, rather than an appreciation of linguistic context. Therefore, reference taps a strength in pragmatic understanding as distinct from structural language abilities in this population. This study also extends the literature by testing pragmatic skills in non-English speaking children with SLI.

Typical and Atypical Development of Reference

Children typically develop adult-like abilities to spontaneously produce unambiguous referring expressions in simple contexts by approximately six years of age (Dickson, 1982; Girbau, 2001; Matthews, Lieven, & Tomasello, 2007). It has been suggested that their production of under-informative referring expressions prior to this age is due to a lack of awareness of other potential targets in the environment (Whitehurst, 1976) or because they don't realise that to refer means to implicitly describe differences that will eliminate potential ambiguity (Whitehurst & Sonnenschein, 1981). Likewise, in comprehension, children are

able to use the presence of modification to make early inferences about an intended target by five years of age (Huang & Snedeker, 2009), and will seek clarification of under-informative expressions at around the same age (Nadig & Sedivy, 2002).

Off-line studies have shown that the necessary metalinguistic skills for judging the adequacy of referring expressions are evident by age four (Nilsen & Graham, 2012; Morisseau, Davies, & Matthews, 2013). These studies elicited such abilities at earlier ages than in previous studies (e.g. Davies & Katsos, 2010; Robinson & Robinson, 1982) by using implicit rather than explicit measures of inadequacy detection. The Pragmatic Tolerance account (Katsos & Bishop, 2011) sheds light on this apparent disconnect through its discovery of methodological effects on pragmatic sensitivities. In studies of the development of scalar implicature, Katsos & Bishop (2011) found that although five- to six-year-old children appeared to be insensitive to violations of informativeness when they were required to straightforwardly accept or reject utterances, the same population performed well in picture-matching and action-based tasks. Furthermore, they penalised under-informative items when using a Likert scale of acceptability, suggesting that they are indeed sensitive to these types of pragmatic violation. These findings are taken as evidence that children do not lack overall competence with informativeness as comprehenders, but that their low performance on binary judgment tasks should be attributed to their tolerant attitude towards violations of pragmatic meaning when engaged in a forced choice judgment.

Reference is vulnerable in neurodevelopmental disorders, for example, children with autism spectrum disorder (ASD) show inappropriate use of linguistic markers of accessibility such as full nouns in place of pronouns (Arnold, Bennetto, & Diehl, 2009) and definite noun phrases in place of definite reference when introducing characters in narrative (Tager-Flusberg, 1995). These and other pragmatic impairments have been attributed to concurrent deficits in social cognition (Baron-Cohen, 2000; Surian, Baron-Cohen, & van der Lely, 1996) and/or weak central coherence (Happé, 1997). However, identifying potential sources of pragmatic deficit is impeded by the variable cognitive and language profiles that characterise

ASD (Kjelgaard & Tager-Flusberg, 2001; Norbury 2005a), meaning that the source of error is far from clear (Bishop, 1989).

The fact that many children with SLI are challenged by pragmatic aspects of communication suggests structural language abilities as a potential contributor to pragmatic difficulty in ASD, at least for those children with ASD who present with concomitant language impairment (cf. Norbury, 2013). For instance, children with SLI are less skilled than age-matched typically-developing (TD) peers at using contextual information to resolve lexical ambiguities and to understand metaphors due to their low vocabulary skills (Norbury, 2005a,b). Similarly, Katsos, Andrés-Roqueta, Estevan, & Cummins (2011) reported that children with SLI were less likely than age-matched TD peers to reject under-informative quantifiers such as 'some of the bananas are in the boxes' when all of the bananas were in the boxes. Again, pragmatic language abilities in this study were in line with their overall low structural language abilities. These findings suggest that difficulties using linguistic context in SLI reflects inadequate semantic or syntactic knowledge, rather than an additional pragmatic deficit. Since reference involves the interplay of language competence with an understanding of others' epistemic states, it is a useful test-bed for investigating these two spheres of ability.

There are few experimental studies of referential abilities in SLI, and those that exist provide conflicting results. Bishop and Adams (1991) reported that children with SLI (aged 8-12 years) achieved lower scores on a production measure of referential communicative adequacy relative to age-matched TD peers. Error patterns varied with age; 8-9 year-olds gave fewer discriminating features than peers, while the 10-12 year-olds provided more redundant information. In contrast, Johnston, Smith, and Box (1997) found no differences between 4-year-old children with SLI and age-matched TD peers and Reuterskiöld-Wagner, Nettelbladt, and Sahlén (2001) reported similar performance in 7 year-old children with SLI. However, qualitative analyses by Johnston et al (1997) revealed that children with SLI were less likely to refer to the targets as a set with attributes in common (e.g. 'the green ones'),

adopting instead a more piecemeal referential strategy (e.g. “the green one and the green one”). The current study aims to clarify this rather mixed picture of referential abilities in SLI by investigating production, comprehension and judgement of reference in the same children.

Relative to production and comprehension tasks, judgements of referring expressions arguably rely less on linguistic computation and more on pragmatic awareness of conversational adequacy, and may therefore reveal relative communicative strengths in children with SLI. Surian et al. (1996) investigated the detection of maxim violation by children with SLI and those with ASD matched for age and verbal ability. Both clinical groups were less sensitive to violations of the maxim of quantity relative to age-matched peers, accepting utterances that contained redundant information. Due to the wide variation in language ability in the ASD sample, clear conclusions cannot be made regarding the source of the pragmatic impairment in this study. It does however provide preliminary evidence of pragmatic difficulty in SLI in judging the optimal amount of information required to make referential choices. Similarly, in a study investigating awareness of inadequate messages, Meline & Brackin (1987) found that language-impaired children predominantly blamed the listener rather than the speaker for under-informative messages, whereas age-matched TD peers correctly blamed the speaker.

Identifying the Source of Error in Referential Communication

Findings to date suggest that children with SLI produce more inadequate referential expressions which may be both under- and over-informative, and are less likely to penalise or repair non-optimal utterances relative to age-matched peers. Identifying the source of these errors is challenged by the range of referential tasks employed and the variable ages of participants across studies. To address these issues, we explicitly compared production, comprehension and judgement of referential terms in the same participants. Furthermore, we examined the relationships between aspects of reference and performance on

standardised measures of language ability. This approach aims to clarify whether the locus of impaired referential ability in SLI lies in production/comprehension skills (i.e. due to a structural linguistic deficit), in problems recognising the need for sufficient information (i.e. due to a pragmatic deficit), or some combination of the two.

We hypothesized that children with SLI would perform more poorly than TD peers on the production task and that their lower performance would be related to expressive language. With regard to comprehension, we predicted relatively few errors overall given the simple adjective/noun combinations used in the experiment, but nevertheless, children with SLI were expected make more errors associated with level of receptive language. In contrast, we predicted that children with SLI would perform comparably to TD peers on a judgement task that should more directly tap pragmatic abilities as distinct from linguistic competence.

Method

Participants and standardised tests

Eighteen children with SLI aged 5;0 to 10;11 (11 males) and eighteen TD children (11 males) matched for chronological age and non-verbal reasoning abilities were recruited from the north-east region of Spain. Ethical permission was obtained from the Spanish Ministry of Culture, Education and Sports. Informed, written consent was obtained from parents and verbal assent was obtained from eligible children prior to assessment.

Children with SLI met the following criteria: (i) diagnosis and on-going clinical involvement from a speech-language therapist; (ii) no sensorineural hearing loss, intellectual impairment, physical disability, or diagnosis of ASD; (iii) Spanish as first language; (iv) non-verbal abilities within 1SD of normative mean on the Raven's Coloured Progressive Matrices (Raven, Raven and Court, 1998). Participants were also required to score -1SD or more on one or both of the standardized language assessments we ran, i.e. the *Comprensión de*

Estructuras Gramaticales (CEG; Mendoza, Carballo, Muñoz, and Fresneda, 2005); the Sentence Memory Recall subtest from the *Evaluación del Lenguaje Infantil* (ELI; Saborit and Julian, 2005). We also tested the children on the vocabulary sub-tests of the WISC-IV (Wechsler, 2003) in order to more comprehensively measure participants' language ability, and to explore any relationship to their referential abilities.

In addition to measuring participants' language levels, we also assessed their false-belief abilities as a proxy measure of social-cognitive skill. Two established tasks for evaluating first-order false beliefs were employed: 1) the Change of Location task (Wimmer & Perner, 1983), in which children were asked to predict a third person's behaviour based on her false belief (*where will Sally look for the ball?*). Control questions were asked to check that the participant remembered and understood the story, and pictures were provided to help the children's understanding. *2)* the Unexpected Contents task (Perner, Leekam, & Wimmer, 1987), in which children are asked to predict the behaviour of a third person based on the understanding of a false belief. As in the change of location task, control questions were asked to check that the participant remembered and understood the story, and props (Smarties tube and a pencil) were used to help the understanding. In both tasks, control questions and critical false-belief questions had to be answered correctly to pass the tasks. Participants were coded as 'passers' when they passed both false belief tasks and coded as 'failers' when they failed one or both false belief tasks. The rationale for using these proxy measures was twofold: 1) to test whether the LI group showed impairments in one aspect of social cognition relative to their TD peers; 2) to provide a potential explanation for any impaired performance in produced informativeness (i.e. an inability to take an addressee's perspective). Further, since there is evidence to suggest that children with SLI can pass these tests of false belief by seven years of age (Andrés-Roqueta et al., 2013; Farmer, 2000), they should be within the reach of the children at the mean age of our clinical sample.

TD children were recruited from the same schools as the children with SLI and pairwise matched by gender and age (within three months). Additionally, they had no

intellectual impairment or highly gifted background and spoke Spanish as their first language. Table 1 reports descriptive statistics for the group selection measures.

	SLI (n=18)		**TD (n=18)**				
	Mean **Range**	**(SD)**	**Mean** **Range**	**(SD)**	***t***	***p-value***	***Cohen's d***
Descriptive measures							
age (y;m)	7;4 5;1 – 10;9	(1;10)	7;6 5;0 – 10;9	(1;9)	-.46	ns (*p*=.65)	0.15
Non-verbal reasoning							
Raven matrices	21.94 14-34	(6.14)	25.05 15-34	(6.74)	-1.45	ns (*p*=.16)	0.48
Linguistic measures							
receptive grammar	51.27 36-68	(8.67)	66.16 39-78	(8.73)	-5.12	<.001	1.71
expressive grammar	6.05 4-9	(1.55)	8.27 4-9	(1.17)	-4.83	<.001	1.62
vocabulary	17.05 2-40	(8.94)	25.44 8-41	(8.89)	-2.82	<.01	0.94
Theory of Mind measures							
false belief (passers/failers)	10/8		12/6		Chi-squared $\chi(1) = 0.468$	ns (*p*=.49)	

Table 1. Scores on background measures. All scores reported as raw scores.

Experimental Procedure

Children were assessed in school over three sessions. Language and cognitive skills were assessed in the first session, the production task was completed in the second, and the comprehension and judgement tasks in the third, with an average of two weeks between assessment sessions. The experimental tasks were developed from a similar paradigm used by Davies and Katsos (2010), which had successfully provided a comprehensive measurement of referential ability in TD children and adults. The stimuli were created especially for this study, were equally visually salient, and all objects were familiar to children.

Production task. Participants sat facing a computer screen, with one experimenter (E1) sitting to their left and another (E2) on their right. Greyscale drawings of four everyday objects were displayed in each quadrant (Figure 1). Participants received a booklet containing the same on-screen displays with the target highlighted in red (visible only to the participant), and were instructed to ask E2 to click the image on the computer screen that matched the target image in their book. Four practice items were followed by 32 experimental items. Utterances were audio recorded for transcribing and coding off-line.

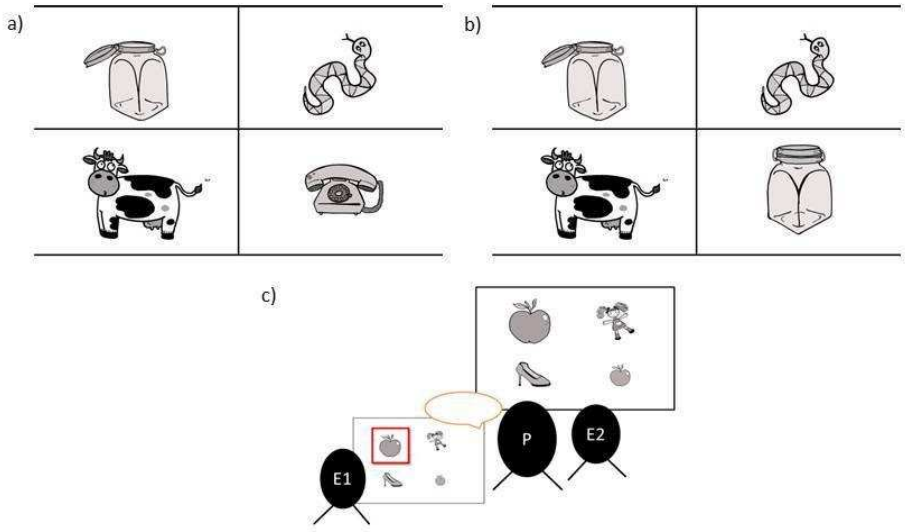

Figure 1. Example arrays for production and comprehension/judgement task; a) no-contrast condition, b) contrast condition, c) experimental set-up for production task (upper display

represents computer screen; lower display represents picture booklet visible only to the participant (P) and E1).

Comprehension and judgement tasks. This task was designed to test (i) participants' comprehension of utterances containing simple referring expressions and (ii) their metalinguistic judgment of felicitous and infelicitous referring expressions. The latter component tests whether participants can reliably integrate information from the visual and discourse context to penalise expressions which fall short of pragmatic expectations to provide enough information for unique identification of a referent (Grice 1975/1989), e.g. 'the jar' in a display containing two jars. Participants viewed displays of four everyday objects whilst listening to pre-recorded sentences. They were instructed to click on the item on the computer screen mentioned in the sentence. The examiner explained that "sometimes the computer uses good Spanish and sometimes it uses bad Spanish. Let's tell it how good its Spanish is to help it learn". The participant then rated utterances on a scale of 1 to 5, with the best utterance indicated by a happy face and the label 'muy buena' (very good), the worst by an unhappy face and the label 'muy mala' (very bad), and the intermediate keys marked with the labels 'buena', 'regular', 'mala' and their corresponding facial expressions. There were five practice items, followed by 64 experimental trials separated into five blocks.

Items and Experimental Design

For the production task, 32 arrays were created in two versions: a no-contrast display, which contained only one referent of a noun category (e.g. a jar, a snake, a cow and a telephone) and a contrast display, which contained two referents of the same noun category, differing in one dimension (e.g. an open jar, a snake, a cow and a closed jar; see Figure 1a and 1b), thus requiring modification to disambiguate the target referent. Each trial comprised four images: a target image and three distracters. These 32 items were split into

two equal groups; half of the participants saw the first group of items in a no-contrast display and the second group of items in the contrast display while the other half participants saw the item groups in the reverse order. Presentation order was counterbalanced across participants. The order of items within each group was pseudorandomised and presented using Microsoft PowerPoint.

In the comprehension task, images were displayed in E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) and were accompanied by pre-recorded utterances spoken by a female native speaker of Spanish. We used a 2(contrast-set) x 2(adjectival modification) repeated measures design, creating four conditions: over-informative (a no contrast set with modification), under-informative (a contrast set, no modification), optimal-1 (no contrast set, no modification), and optimal-2 (a contrast set with modification). There were 32 critical items, eight in each condition, plus 32 filler items. The filler items were descriptions such as 'click on the animal / furniture / thing you find in the kitchen'. Items were pseudorandomly presented, the position of the target referent and the contrasting referent was rotated between items in all conditions, and every target item appeared in only one of the four conditions between participants. Following selection of target item, participants were asked to rate the utterance on a 5-point scale.

Coding Procedures

There are two ways of being under-informative in this paradigm: omitting a necessary adjective, e.g. 'the apple' in a context containing a large and a small apple, or producing a globally ambiguous RE, e.g. 'the clothes' when the target referent is a shoe. Accordingly, we used two complementary schemes to code the expressions produced by the participants (see appendix for details). The first coding scheme ('pragmatic') allowed examination of the pragmatic informativeness of utterances by asking whether the referring expression (i) provided the listener with enough information to uniquely identify the referent, and (ii) provided only the essential information (i.e. no redundant information). Expressions were

coded as under-informative (e.g. 'the apple' in a contrast context containing one big and one small apple), optimal ('the big apple' in the same display) or over-informative ('big apple' in a display containing only one apple). Referring expressions that deviated slightly from the anticipated noun, but would lead a listener to resolve reference just as easily (e.g. 'box' for present) were included in this analysis. Production data were scored by two raters, with inter-rater reliability of r=0.98, p<.001 for under-informative codes, r=0.99, p<.001 for optimally-informative codes, and r=0.95, p<.001 for over-informative codes.

The second coding scheme ('lexical') identified lexical inadequacies which threatened reference resolution, and thus may be more sensitive to language impairment. Two points were awarded for accurate responses (e.g. 'the open jar' for the target item shown in Figure 1b), one point for plausible expressions that deviated slightly from the anticipated response (e.g. 'the open bottle' for the same item), and no points for unacceptable/uninterpretable expressions (e.g. 'the open box' for the same item). Inter-rater reliability was r=0.99, p<.001. See appendix for further details of how various forms of referring expression were scored.

Results

Background Measures

As reported in table 1, there were no differences between groups on age, non-verbal IQ, or false-belief comprehension. As expected, there were significant group differences on all language measures. Thus our clinical group had the expected profile of children with SLI, having pronounced structural language difficulties but relatively better socio-cognitive understanding. Although the groups did not differ with respect to age, there was a wide age range in both groups; thus we controlled for age in all statistical analyses.

Production Task

Pragmatic coding scheme. 1152 responses were collected (576 from each group). In the pragmatic coding scheme, four inaccurate responses from the TD group were excluded (e.g. 'box' for an item which contained two cups, a teapot and a sweet). Twenty-seven responses from children with SLI were excluded due to no response, pointing response, lexical error, incorrect target, or an adjective-only response provided. The number of rejected trials differed significantly between groups, $t(34) = -3.86$, $p < .001$, $d = 1.32$.

As response types are non-independent observations, the analysis focused on rates of optimal reference. Using the pragmatic coding scheme, the mean frequencies of optimal referring expressions are presented in Figure 2. Results were analysed in a 2 (group: SLI, TD) x 2 (contrast condition: no contrast, contrast) repeated measures ANCOVA. All effects are reported as significant at $p < .05$.

There was a significant main effect of group on mean rate of informativeness, $F(1,33) = 11.32$,, $p = .002$, $\eta^2 p = .25$, with children in the TD group producing more informative referring expressions (maximum 16) compared with children in the SLI group (TD: $M = 13.5$, $SE = .49$; SLI: $M = 11.2$, $SE = .49$). There was no significant interaction between group and the co-variate (age); $F(1, 33) = 2.15$, $p = .15$. The main effect of contrast condition was not significant, $F(1,33) = 3.68$, $p = .06$, $\eta^2 p = .12$, though there was a tendency for more informative referring expressions to be produced in the no contrast condition relative to the contrast condition (no contrast: $M = 13.2$, $SE = .41$; contrast: $M = 11.5$, $SE = .63$). The interaction between group and contrast condition was marginal, $F(1,33) = 3.35$, $p = .076$, $\eta^2 p = .10$. This suggests that whilst both groups provided fewer optimal expressions in the contrast condition, this pattern tended to be more marked in the SLI group, as these children were less likely to provide a modified noun.

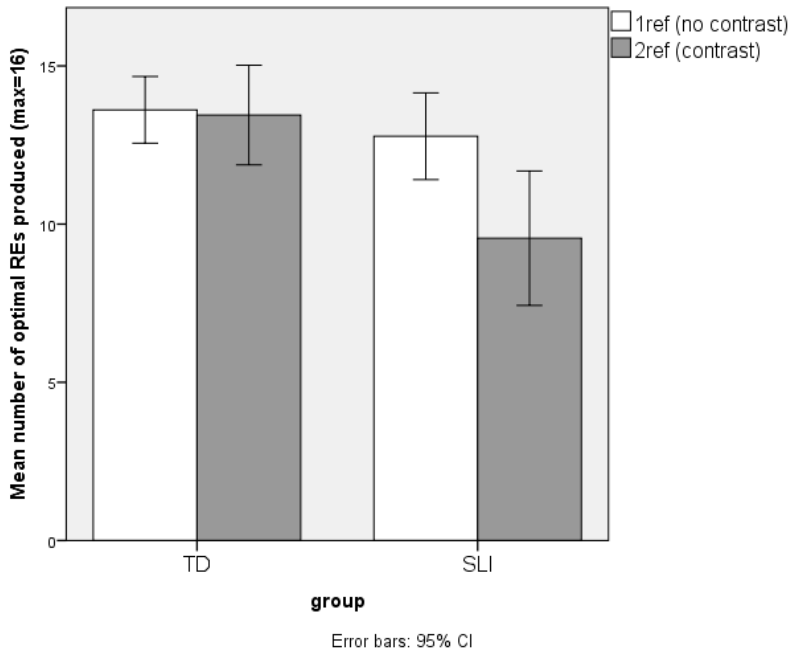

Figure 2. Production results: a) Mean frequency of optimally informative referring expressions by group and contrast condition (pragmatic coding method).

Children tended not to produce under-informative referring expressions; however, an independent samples *t*-test revealed that children with SLI produced significantly more under-informative responses (*M*=5.6, *SD*=4.1), relative to TD peers (*M*=2.1, *SD*=2.9), *t*(31) = -2.93, *d*=-1.06.

Lexical coding scheme. The lexical coding system evaluated the semantic accuracy of all responses and highlighted increasing challenges for children with SLI. . A repeated measures ANCOVA confirmed a main effect of contrast condition $F(1, 33) = 12.34$, $p=.001$, $\eta^2 = .27$, with greater accuracy in the no-contrast condition. There was also a main effect of group $F(1, 33) = 14.38$, $p=.001$, $\eta^2 = .30$, and a significant group x condition interaction: $F(1, 33) = 4.90$, $p=.034$, $\eta^2 = .13$. The interaction between group and the co-variate (age) was not significant, $F(1, 33) = 3.12$, $p = .086$. Children with SLI produced fewer accurate utterances overall (*M*=45.7, *SE*=2.0) relative to TD peers (*M*=55.1, *SE*=1.5), *d*=1.34. In the contrast condition, children with SLI achieved significantly lower scores (*M*=17.4, *SE*=1.8) relative to TD peers (*M*=24.6, *SE*=1.4). While this pattern is similar in the

no-contrast condition, (SLI: *M*=28.3, *SE*=0.5; (TD: *M*=30.5, *SE*=0.4), the difference between groups was attenuated, as illustrated in Figure 3. In the contrast condition, the pragmatic infelicity is marked by under-informativeness, as multiple referents require modification of a target. In the no-contrast condition, the corresponding pragmatic infelicity is over-informativeness, since no modification is required. Taken together, the coding systems reveal that the children with SLI tend to be under-informative, both pragmatically and lexically.

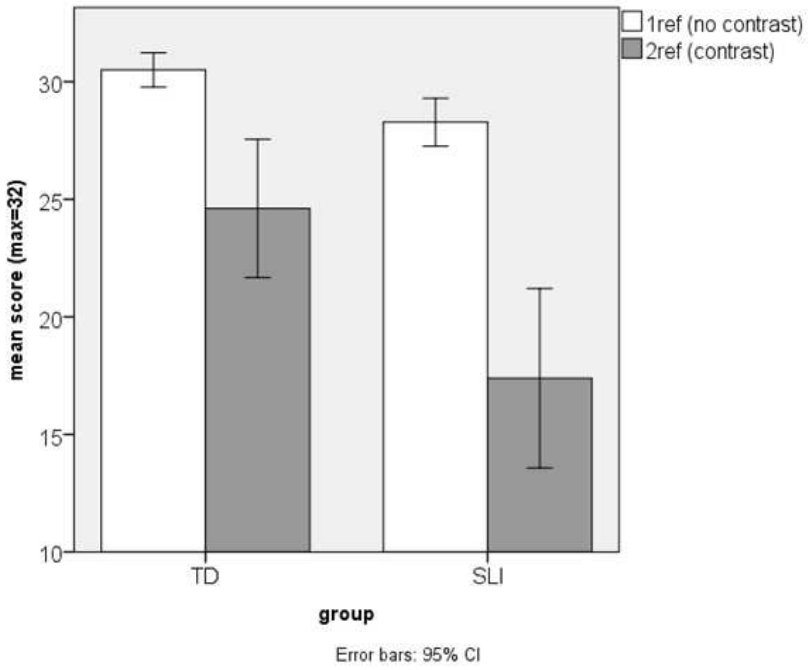

Figure 3. Mean scores on the lexical coding scheme

Comprehension Tasks

Error rates for 24 critical items (over-informative, optimal-1, and optimal-2 conditions were generally low; across 432 trials there was only one error in the entire TD dataset versus 19 in the SLI dataset (under-informative items were not included since there is no single correct target in this condition). Nevertheless, an independent samples t-test revealed that children with SLI were more likely to make errors in target accuracy (*M*=1.06, *SE*= .26) than TD peers (*M*=0.06, *SE*= .06), *t*(17)= -3.67, *p*<.005, *d*=-1.78. Error analysis did not

reveal a consistent response pattern; 11 of the 19 errors made by the SLI group were same-category errors, e.g. clicking on the long sock in response to an instruction to click on the short sock, with the remaining 8 being different-category errors, e.g. clicking on the lemon in response to an instruction to click on the cushion. Error rates on the comprehension task were not associated with any language or social-cognitive measures, probably due to the low error rates found in the data.

Pragmatic judgement ratings are presented in Figure 4. These data did not meet parametric assumptions and were therefore analysed using a Mann-Whitney test, with age uncontrolled. TD children correctly penalised under-informative items more frequently (*M*=3.23, *SE*=.29; maximum penalty=5) than peers with SLI (*M*=2.11, *SE*=.26), *U*= 81.5, *p*<.05, *r*=.42. No significant between-group differences were found for any other condition (all *U*s >147, all *p*-values > .5). Notably, ratings of under-informative utterances were significantly correlated with age within the TD group, *r*(18) =.66, *p*= .003, but there was no association between age and pragmatic ratings within the SLI group, *r(*18) = .297, *p* = .23.

Despite the SLI group's apparent insensitivity in the judgement task, within-group analyses using Wilcoxon matched-pairs tests revealed that *both* groups in fact penalised under-informative utterances significantly more than optimal or over-informative utterances (see table 2).

	SLI		TD	
Comparison conditions	*Z*	*p*	*Z*	*p*
under vs. optimal-1	-2.49	<.05	-3.52	<.001
under vs. optimal-2	-2.94	<.005	-3.52	<.001
under vs. over	-2.60	<.01	-3.52	<.001

Table 2. Pairwise comparisons for within-group ratings of under-informative utterances vs. vs. optimal-1, optimal-2 and over-informative utterances.

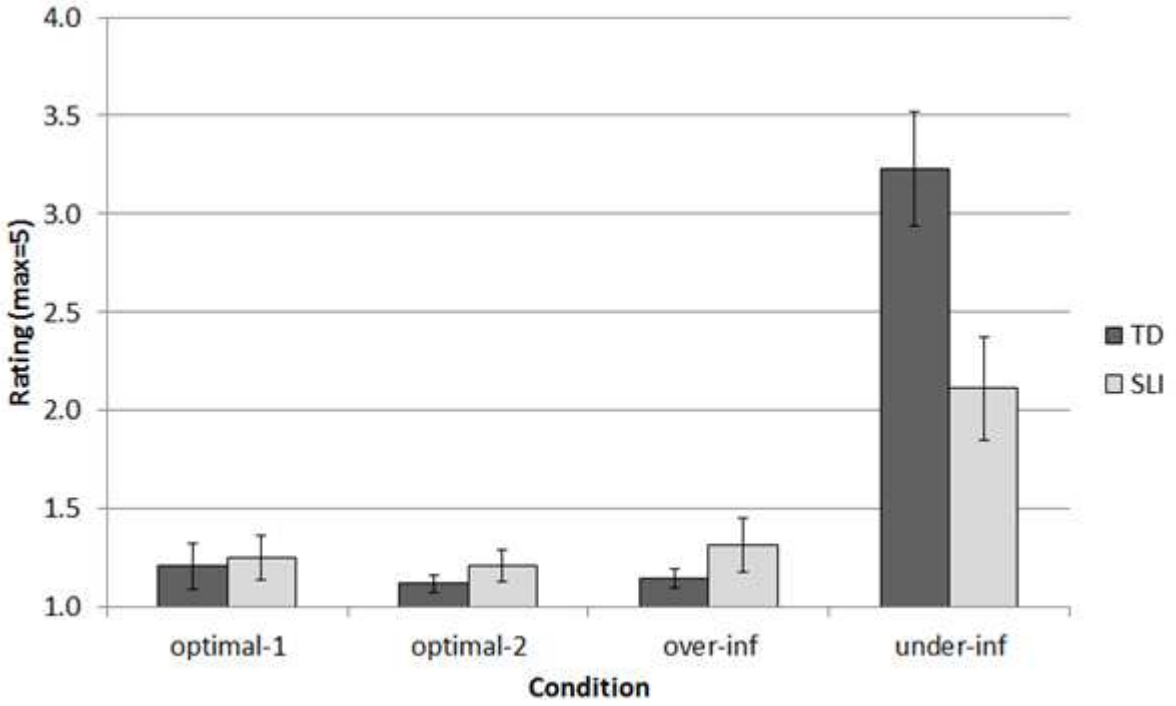

Figure 4. Mean ratings in the judgement task. Error bars show standard error of the mean.

Thus, although the SLI group were less likely to penalise under-informative items relative to their TD peers, they were in fact sensitive to the pragmatic infelicity, as revealed by their higher penalties for this condition relative to the comparison conditions. Children with SLI notice the infelicity, but they did not penalise it as heavily as the TD group.

Associations between Referential Communication and Structural Language Skill

Partial correlations between rates of optimally-informative utterances in the production task, ratings of optimally-informative utterances in the pragmatic judgement task, and the standardised language measures were analysed, using age as a covariate. Given the very low error rates in the comprehension task, it was not possible to explore these associations in the comprehension task. Across both groups, there were strong and significant correlations between production of optimal utterances in the contrast condition and all language measures (receptive grammar, $r = .74$, $p < .001$; vocabulary, $r = .44$, $p = .008$; sentence recall, $r = .42$, $p = .013$). However, with the exception of sentence recall ($r = -.06$), these associations tended to be stronger in the SLI group (receptive grammar, $r = .598$,

$p < .05$; vocabulary, $r = .45$, $p = .07$); in the TD group, only the association between receptive grammar and production of optimal utterances remained significant, $r = .78$, $p < 001$ (cf. vocabulary, $r = .07$; sentence recall, $r = -.06$).

Across both groups, performance on the pragmatic judgement task was found to significantly positively correlate with both of the language measures that were found to be significant above (receptive grammar, $r = .53$, $p = .03$; vocabulary, $r = .75$, $p < .001$). Again, these associations were biased towards the SLI group with judgments correlating strongly and significantly with the receptive grammar measure ($r = .53$, $p < .05$) and vocabulary ($r = .75$, $p < .001$), though not with sentence recall ($r = .29$, $p = .26$). None of the correlations were significant within the TD group, all rs < .315, ps > .22 (see Table 3).

We also tested for significant differences between group correlations using the Fisher's r-to-Z transformation. However, there were no significant differences between any of the correlations in the SLI group vs. the TD group (all ps >.1), suggesting that the relationship between performance on our tasks and performance on the standardised tests was similar in both groups.

	receptive grammar	wisc vocabulary	sentence recall
Optimality in production	.60* (.78*)	.44, p=.07 (.07)	.28 (-.06)
Judgments of under-informativeness	.53* (.41)	.75* (.32)	.29 (.26)

Table 3. Partial correlations (with age as a covariate) between rates of optimally informative utterances in the production task (contrast condition only), pragmatic judgments of under-informativeness, and standardised language measures. r values for the SLI group are shown, with TD correlations in parentheses. Correlations marked with an asterisk were significant at $p < .05$.

Discussion

This study measured referential communication abilities in a sample of children with SLI in production, comprehension and pragmatic judgements in order to explore associations between their referential and language abilities. Our results indicate that children with SLI exhibited deficits on all three tasks. Relative to their typically-developing peers, they produced more under-informative referring expressions, made more errors in identifying the intended target, and penalised under-informative utterances less frequently than TD peers, although importantly, they demonstrated some sensitivity to detecting such utterances. They also made more errors in choice of lexical expression, suggesting that their low performance is symptomatic of a more widespread linguistic deficit. This observation is supported by correlational analyses, which demonstrated significant correlations in the SLI group between performance on the production task and language ability (receptive grammar; vocabulary), and performance on the judgment task and the same measures of language ability.

As predicted, when compared with age-matched TD peers, children with SLI produced more under-informative expressions that did not allow unique identification of a target referent and they made more errors resolving reference in comprehension. Given that children with SLI were as likely as their TD peers to pass explicit tests of false belief reasoning, it seems unlikely that challenges with reference reflect a failure to take account of the listener's perspective. Instead, correlational analyses support the suggestion that these tasks relied on structural language abilities. These findings are consistent with Bishop and Adams (1991), who also reported that children with SLI performed more poorly on a referential communication task, were more likely to produce under-informative utterances, and that performance was associated with measures of receptive grammar. However, Bishop and Adams (1991) argued that the cognitive demands of the task, particularly the need to scan a complex visual array and keep in mind distinguishing features of potentially ambiguous referents may exceed the processing capacity of children with SLI. Our

experiment was not designed to test this possibility; future research may usefully employ eye-tracking paradigms to highlight the extent to which children attend to different aspects of the visual array and adjust their verbal utterances accordingly.

We anticipated that the judgement task would be a more directly pragmatic task, and that children with SLI would perform more similarly to TD children. In fact, they gave lower penalties for under-informative expressions than their TD counterparts. However, they did succeed in penalising these constructions appropriately relative to both optimally informative utterances and the less serious violation of over-informativeness. Interestingly, it is in this most pragmatically demanding task in which the children with SLI fared best, suggesting some relative strength in pragmatic awareness. We conclude that children with SLI differ from TD peers in their *attitudes* towards pragmatic infelicities, not in their abilities to detect them. Notably, while the ability to penalise under-informative utterances improved with age in the TD group, this was not the case for children with SLI. This further suggests that increasing language ability is an important driver of change in pragmatic competence.

The performance of the SLI group in this regard is in line with Pragmatic Tolerance, the developmental account of nonstandard sensitivity to informativeness violations (Davies & Katsos, 2010; Katsos & Bishop, 2011). On this view, typically-developing young children can be competent with informativeness (as evidenced by target-like rating using a gradable scale), while at the same time tolerant of pragmatic infelicity, for example, incorrectly accepting under-informative utterances when using a binary scale. Using a more sensitive gradable scale, we found lower penalties for under-informative utterances in the SLI group relative to their TD peers, but also found sensitivity *within* the SLI group to the same infelicity relative to felicitous comparison items. This suggests that children with SLI behave like younger TD children in not prioritising subtle expectations for optimal amounts of information, provided that the referring expression is semantically true. In other words, faced with limited language, mentioning the head noun may be sufficient, rendering the modifier optional. Notably, a similar pattern of errors was evident in the production task, in which

errors were almost always under- (as opposed to over-) informative. This suggests that the pragmatic demands of the task exceed linguistic competence in our SLI group. Lack of modification was the most common error type, with the head noun generally preserved. This pattern of findings accords with accumulating evidence that while children with SLI experience pragmatic immaturities relative to age-matched peers, their pragmatic abilities are more in-line with their linguistic capacities (cf. Norbury, Nash, Baird & Bishop, 2004).

An alternative hypothesis is that referential problems in SLI may be caused by social cognitive deficits leading to problems with pragmatics, although such deficits may only affect a proportion of children with potential 'pragmatic language impairments' (or now, 'social communication disorder'; Norbury, 2014). We would argue against this explanation for three reasons: first, performance on the experimental tasks was significantly correlated with measures of structural language skill, particularly in the SLI group. Second, our groups did not differ in the ability to pass explicit tests of social-cognitive reasoning as measured by two classic Theory of Mind tasks. Although false-belief constitutes only one aspect of social cognition, it suggests that our SLI group do not have pronounced deficits in understanding other minds. While a significant minority of children with SLI failed the false belief task, these children also had more severe language deficits, making it difficult to disentangle the effects of language and social reasoning on referential communication skill. Third, we guarded against a possible subset of children with pragmatic language impairment by requiring that our SLI participants did not have a diagnosis of ASD or ratings of significant pragmatic impairment by parents and teachers. Of course, further investigation of pragmatic and socio-cognitive competence is necessary to definitively rule out such an explanation, but in our sample, structural language deficit appears to be a more likely contributor to referential inadequacy.

The pragmatic judgement task is demanding in that it requires the listener-rater to compare the need for information (based on the visual context) with the actual utterance encountered, and then recognise that something is missing. Under a general cognitive

account of SLI (e.g. Leonard's Surface Hypothesis, 1998), children with SLI may be less able to integrate simultaneously multiple pieces of information or retain rapidly incoming information for sufficient periods to make the necessary comparison between expected and experienced input (Bishop & Adams, 1991). The SLI group's reasonably good performance on the pragmatic judgement task suggests that their integration skills were not entirely compromised, despite the fact that they performed more poorly on the comprehension task. These results may seem contradictory, as one might expect adequate comprehension to be a prerequisite for passing the judgement task. However, error rates in comprehension were low: 7 of the 18 participants with SLI made no errors at all, 6 only made one error, 2 made two errors and 3 made three errors. On this simple task, SLI participants did perform more poorly than the TD peers, but understood the utterances well enough to resolve reference appropriately on the majority of trials and crucially, well enough to make informed pragmatic judgements.

In conclusion, our results support the view that structural language is a key factor in referential ability, and it is such syntactic and lexical impairments rather than a pragmatic deficit which challenge children with SLI in this domain. These findings have important implications for research with other clinical populations. For example, investigators have tended to link pragmatic deficits in ASD to impaired social cognition (e.g. Tager-Flusberg, 2000), though linguistic ability varies enormously within ASD (cf. Kjelgaard & Tager-Flusberg, 2001; Loucas et al. 2008). Our findings suggest that language may be an important predictor of pragmatic performance in ASD; we recommend that future research in ASD investigates the impact of individual differences in structural language competence on ostensibly pragmatic, or social-communication, tasks (e.g. Norbury, 2005a, b). In line with this suggestion, an increasing number of studies have highlighted that individuals with ASD and age appropriate language scores perform comparably to age-matched TD peers on pragmatic tasks (Norbury, 2005a,b; Pijnacker et al, 2009; Chevallier et al, 2010). Comparing rates of over- and under-informative utterances in different clinical groups will yield novel

insights into the underlying skills that support communicative competence. An important avenue for future research will be to determine whether improving structural language skills will benefit pragmatic language development.

Acknowledgements

This work was part-funded by a research travel grant from the ESF's EURO-XPRAG Research Network Program. We are grateful to the Fundacio Caixa-Castelló (grant number P1–1B2010–16), the Conselleria de Educación, Cultura y Deporte of Generalitat Valenciana (grant number GV/2015/092) and the Spanish Ministerio de Ciencia e Innovacion (grant number EDU2010–21791) for help accessing the sample. Thanks to Raquel Flores Builes and Irene Garcia Molina for data collection, Jack Wilson who was the second rater of the production data, and to the anonymous reviewers for their valuable comments on earlier versions of this work.

References

Andrés-Roqueta, C., Adrian, J. E., Clemente, R. A., & Katsos, N. (2013). Which are the best predictors of theory of mind delay in children with specific language impairment? *International Journal of Language & Communication Disorders, 48*(6), 726-737.

Ariel, M. (1990). *Accessing noun phrase antecedents*. London: Routledge.

Arnold, J.E., Bennetto, L., & Diehl, J.J. (2009). Reference production in young speakers with and without autism: Effects of discourse status and processing constraints. *Cognition, 110,* 131-146.

Baron-Cohen, S. (2000). Theory of mind and autism: a fifteen year review. In *Understanding other minds*, vol. 2 (ed. S. Baron-Cohen, H. Tager-Flusberg & D. Cohen). Oxford University Press.

Bishop, D. V. M. & Adams C. (1991). What do referential communication tasks measure? A study of children with specific language impairment. *Applied Psycholinguistics*, 12, 2, 199-215.

Bishop, D. V. M. (1989), Autism, Asperger's syndrome and semantic-pragmatic disorder: Where are the boundaries? *International Journal of Language & Communication Disorders, 24*, 107–121.

Bishop, D. V. M. (2003).*The Test for Reception of Grammar, version 2 (TROG-2)*. London: Psychological Corporation.

Chafe, W. (1976). Givenness, Contrastiveness, Definiteness, Subjects, Topics, & Point of View. In C. N. Li (Ed.), *Subject and Topic* (pp. 25-56). New York: Academic Press Inc.

Chevallier, C., Wilson, D., Happé, F., & Noveck, I. (2010). Scalar inferences in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders,* 40(9), 1104-1117.

Clark, H. H. & Bangerter, A. (2004). Changing ideas about reference. In I. A. Noveck & D. Sperber (Eds.), *Experimental Pragmatics* (pp. 25-49). Basingstoke: Palgrave.

Clark, H. H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1-39.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior*, 22, 245-258.

Davies, C. & Katsos, N. (2010). Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua, 120*(8), 1956-1972.

Dickson, W. (1982). Two decades of referential communication research: A review and meta-analysis. In C. J. Brainerd and M. Presley (eds), *Verbal Processes in Children* (1-33). New York: Springer Verlag.

Farmer, M. (2000). Language and social cognition in children with specific language impairment. *Journal of Child Psychology and Psychiatry*, 41, 627–636

Girbau, D. (2001). Children's referential communication failure: The ambiguity and abbreviation of messages. *Journal of Language and Social Psychology*, 20(1-2), 81-89.

Graf, E. & Davies, C. (2014). 'The Production and Comprehension of Referring Expressions'. In D. Matthews (ed.).*Pragmatic Development in First Language Acquisition* (Trends in Language Acquisition Research 10). Amsterdam: John Benjamins, pp. 161-182.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. Morgan (eds), *Syntax and semantics, 3: Speech Acts* (41-58). New York: Academic Press. Reprinted in Grice, H.P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press, 22–40.

Happé, F. G. E. (1997). Central coherence and theory of mind in autism: Reading homographs in context. *British Journal of Developmental Psychology, 15*, 1–12.

Huang, Y. & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in five-year-olds: Evidence from real time spoken language comprehension. *Developmental Psychology*, *45*, 1723-1739.

Johnston, J. R., Smith, L. B., & Box, P. (1997). Cognition and communication: Referential strategies used by preschoolers with specific language impairment. *Journal of Speech Language and Hearing Research, 40*(5), 964-974.

Katsos, N. & Bishop, D. V. M. (2011). Pragmatic tolerance: implications for the acquisition of informativeness and implicature. *Cognition*, *120,* 67-81.

Katsos, N., Andrés-Roqueta, C., Estevan, R. A. C., & Cummins, C. (2011). Are children with Specific Language Impairment competent with the pragmatics and logic of quantification? *Cognition, 11,* 43-57.

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.

Leonard, L. (1998). *Children with Specific Language Impairment.* MIT Press.

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,* 1184–1192.

Matthews D., Lieven E., & Tomasello M (2007). How toddlers and preschoolers learn to uniquely identify referents for others: a training study. *Child Development*, *78(6),* 1744-1759.

Meline, T. J., & Brackin, S. R. (1987). Language-impaired children's awareness of inadequate messages. *Journal of Speech and Hearing Disorders, 52*(3), 263-270.

Mendoza, E., Carballo, G., Muñoz, J., & Fresneda, M. D. (2005). *Test de Comprensión de Estructuras Gramaticales (CEG)*. Madrid: TEA – Ediciones.

Morisseau, T., Davies, C., & Matthews, D. (2013). How do 3- and 5-year-olds respond to under- and over-informative utterances? *Journal of Pragmatics, 59, Part A*(0), 26-39.

Nadig, A. & Sedivy, J. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13 (4),* pp. 329-336.

Nilsen, E. S. & Graham, S. A. (2012). The development of preschoolers' appreciation of communicative ambiguity. *Child Development*, 83, 1400 – 1415

Norbury, C. F. (2005a). The relationship between Theory of Mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology. 23*, 383-399.

Norbury, C. F. (2005b). Barking up the wrong tree? Lexical ambiguity resolution in children with language impairments and autistic spectrum disorders. *Journal of Experimental Child Psychology. 90*, 2, 142-171.

Norbury, C.F. (2013). Atypical pragmatic development. In D. Matthews (Ed.), *Pragmatic Development in First Language Acquisition.* John Benjamins, pp.343-362.

Norbury, C.F. (2014). Practitioner Review: Social (pragmatic).communication disorder conceptualization, evidence and clinical implications. *Journal of Child Psychology and Psychiatry*, *55(3),* 204–216.

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

Norbury C.F. & Paul R. (2013) Speech, Language, and Social Communication Disorders. In *Rutter's Child and Adolescent Psychiatry*. 6th edition. Blackwells.

Perner, J., Leekam, S. R., & Wimmer, H. (1987) Three year-old's difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5,* 125-129.

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,* 607-618.

Raven, J., Raven, J.C., & Court, J.H. (1998). *Coloured Progressive Matrices*. Oxford: Oxford Psychologists Press.

Reuterskiöld -Wagner, C., Nettelbladt, U., & Sahlén, B. (2001). Giving the crucial information. Performance on a referential communication task in Swedish children with language impairment. *International Journal of Language and Communication Disorders*, *36(4),* 433-445.

Robinson, E.J. & Robinson, W.P. (1982). Knowing when you don't know enough: children's judgments about ambiguous information. *Cognition*, *12*, 267 - 280.

Saborit, C. & Julián, J. P. (2005). *L'avaluació del llenguatge infantil. ELI. Universitat Jaume I de Castelló (UJI)*. Collecció: Educació.

Surian, L., Baron-Cohen, S., & van der Lely, H. (1996).Are children with autism deaf to Gricean maxims? *Cognitive Neuropsychiatry*, *1*: 55-71.

Tager-Flusberg, H. (1995). "Once upon a ribbit": Stories narrated by autistic children. *British Journal of Developmental Psychology, 13*, 45-59.

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 autism and developmental cognitive neuroscience* (2nd ed., pp. 1–45). Oxford: Oxford University Press.

van der Lely, H K. J. (1997).Narrative Discourse in Grammatical Specific Language Impaired Children: A Modular Language Deficit? *Journal of Child Language, 24*, 221 -256.

Wechsler, D. (2003). *Manual for the Wechsler Intelligence Scale for Children - Fourth edition*. San Antonio, TX: The Psychological Corporation.

Whitehurst, G. J. (1976). Development of communication - changes with age and modeling. *Child Development, 47*(2), 473-482.

Whitehurst, G. J. & Sonnenschein, S. (1981). The development of informative messages in referential communication: Knowing when vs. knowing how. In W. P. Dickson (ed.), *Children's oral communication skills* (pp.127-142). New York: Academic Press.

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13,* 103–128.

Appendix: Coding schemes used in production task

Pragmatic coding scheme:

	Category of expression	Example	Informativeness coding
1	Feasible lexical alternative	'box' for present	Coded for informativeness as anticipated noun 'present'
2	Semantically related noun	'glass' for jar	Coded for informativeness as anticipated noun 'jar'
3	Grammatical error	'boots' for boot	Coded for informativeness as anticipated noun 'boot'
4	No attempt	'don't know'	Deleted from this analysis; not coded for informativeness on the grounds that the hearer could not resolve reference.
5	Deictic response	'this' or manual pointing	Deleted from this analysis; not coded for informativeness on the grounds that the hearer could not resolve reference.
6	Lexical error	'cheese' for cushion	Deleted from this analysis; not coded for informativeness on the grounds that the hearer could not resolve reference to the target.
7	Incorrect target	'the boot with spots' for striped boot target	Deleted from this analysis; not coded for informativeness on the grounds that the hearer could not resolve reference to the target
8	Adjective only	'spots' for spotty cushion	Deleted from this analysis; not coded for informativeness on the grounds that the hearer could not easily resolve reference. This is a form of under-informativeness, but not included in the informativeness analysis since the current focus is on the form of noun phrases.

Linguistic coding scheme:

2 points were awarded for:

- Target response (e.g. the mouth)
- A feasible lexical alternative (e.g. the lips)
- Over-informative expressions (e.g. ‘the open mouth’ for a solitary mouth)

1 point was awarded for:

- Unanticipated but partially discriminating adjective (e.g. ‘the big book’ for the open book: the open book covered slightly larger area than the closed one).
- Semantically related noun (e.g. ‘the dress’ for a tshirt)

0 points were awarded for:

- No attempt
- Deictic response (e.g. pointing and/or saying ‘this one’)
- Lexical error (e.g. ‘the pencil’ for a comb)
- Incorrect target (e.g. ‘the spotty boot’ for the stripy boot)
- Adjective only (e.g. ‘open’ for the open fridge)
- No adjective provided when required, or non-discriminating adjective provided (‘the jug’ for a spotty jug)