Riches, N. G., Loucas, T., **Charman, T.**, Simonoff, E. & Baird, G. (2010). Sentence Repetition in children with specific language impairment and autism: a study of linguistic factors affecting recall. *International Journal of Language and Communication Disorders, 45*, 47-60.

Sentence Repetition in Adolescents with Specific Language Impairments and Autism: an Investigation of Complex Syntax

Riches, N. G.[1], Loucas, T.[1], Baird, G.[2], Charman, T.[3], Simonoff, E.[4]

Background: Recent studies have indicated that many children with autism spectrum disorders (ASD) present with language difficulties that are similar to those of children with Specific Language Impairments (SLI), leading some to argue for similar structural deficits in these 2 disorders. **Aims:** Sentence Repetition was used to investigate complex syntax in these groups, i.e. the production of structures with long-distance dependencies. **Methods & Procedures.** Adolescents with SLI (mean age 15;3, n = 14), and ASD plus language impairment (ALI; mean age 14;8, n = 16) were recruited alongside typically-developing adolescents (mean age 14;4, n = 17). They were required to repeat sentences containing relative clauses that varied in syntactic complexity. **Results.** The adolescents with SLI presented with greater syntactic difficulties than the adolescents with ALI, as manifested by higher error rates on the more complex object relative clauses, and a greater tendency to make syntactic changes during repetition. **Conclusions & Implications.** Adolescents with SLI have more severe syntactic difficulties than adolescents with ASD plus language impairment, which may be due to their short-term memory limitations.

1. University of Reading
2. Guy's and St Thomas' Hospital Trust and Kings College London
3. Institute of Education
4. Institute of Psychiatry

What is already known: We know that children with ASD plus language impairments perform poorly on two clinical markers of SLI; non-word repetition, and past tense tasks. Moreover, converging data suggest a substantial overlap between these two groups. However, we do not know the extent of the overlap between their language profiles.

What this study adds: This study is one of the first to compare the performance of children with SLI and ASD plus language impairment on a task involving complex syntax, i.e. the ability to comprehend / produce complex clause level structures such as relative clauses. It is also the first study to employ an automated algorithm to count Sentence Repetition errors, a procedure which may facilitate the analysis of future Sentence Repetition data.

Introduction

Theoretical motivations

Sentence repetition (SR), otherwise known as elicited imitation, is an important tool for diagnosing children with Specific Language Impairment (SLI), whose poor language cannot be explained by factors conventionally associated with language difficulties, e.g. hearing loss, low IQ. Recent studies have found this task to be a highly discriminating diagnostic marker of SLI, capable of distinguishing between children with SLI and non-affected individuals with high degrees of sensitivity and specificity (Conti-Ramsden, Botting, & Faragher, 2001). Its clinical utility has recently led to the development of two standardised clinical assessments (Gardner, Froud, Mclelland, & van der Lely, 2006; Seeff-Gabriel, Chiat, & Roy, 2008).

A less-widely investigated property of SR is its potential to identify specific profiles of language difficulty. Because SR is so open-ended children can make a wide variety of errors when repeating a sentence, e.g. they may add, substitute, transpose or omit words, morphemes and phonemes, or make wholesale changes to syntactic structure. In this way, SR data can provide qualitative information about underlying difficulties. For example, children with SLI presented with a different error profile to children with inconsistent phonological disorder, characterised by a higher proportion of errors involving the word omission (Seeff-Gabriel, Chiat, & Dodd, 2005). “Poor comprehenders,” who have semantically-based reading difficulties, tend to make more semantic substitutions than language-matched controls (Marshall & Nation, 2003). Individuals with Williams Syndrome score well below age-level when repeating sentences with complex syntax, e.g. relative clauses (Grant, Valian, & Karmiloff-Smith, 2002). While these studies demonstrate the potential of SR to investigate language profiles, there have been few "qualitative" studies of SR in language-impaired populations, e.g. studies investigating the kinds of errors children make (e.g. additions, omissions, substitutions), and which kinds of syntactic items or constructions are most affected.

SR emerged as an important paradigm in the late sixties with the work of Slobin and Welsh (1968) and Clay (1971), who argued that if an individual can repeat an utterance longer than their word span, i.e. the number of random words they can repeat, they cannot depend solely on short-term memory (STM), but must use syntactic knowledge to "chunk" the stimulus, i.e. anchor it to representations in long-term memory (LTM) to facilitate recall. More recently, Potter and Lombardi (1992; 1990, 1998) have conducted a series of experiments which manipulate the standard SR paradigm in a variety of ways, for example, by investigating the effect of intervening material, e.g. lure words (1992;1990) and sentence primes (1998). To broadly summarise, these studies suggest that when individuals are required to recall sentences they *reconstruct* the stimulus from information in LTM, including lexical, conceptual and syntactic representations. For example, participants who substituted the main verb during recall, nonetheless tended to maintain the original verb-argument structure, thereby preserving syntactic form (Potter & Lombardi, 1992). Given its sensitivity to syntax SR has been widely used in the field of typical language development to investigate young children's syntactic competence (e.g. Kidd, Brandt, Lieven, & Tomasello, 2007). The paucity of studies using SR to investigate syntax in clinical populations is therefore surprising.

While SR probes representations in LTM, it clearly involves numerous cognitive processes; Phonological STM to temporarily store phonological information (Rummer, 2004;

Willis & Gathercole, 2001), the Central Executive, to manage the exchange of information between STM and LTM (Jefferies, Lambon-Ralph, & Baddeley, 2004), and phonological output processes. Additionally, the relationship between comprehension and repetition is complex. While children can exploit STM to parrot short sentences without successful comprehension, sentences exceeding a child's STM span must arguably be successfully comprehended to ensure repetition. This is because when recall cannot be supported by STM, syntactic and semantic representations in LTM are recruited, and if a sentence is not successfully comprehended, these representations are likely to diverge from the stimulus (Vinther, 2002). In support of this argument, McDade, Simpson and Lamb (1982) found that when information in STM was allowed to decay by inserting a 3-second pause after the stimulus, comprehension closely predicted repetition performance, an association not evident when repetition was immediate and therefore STM could be exploited. Assuming repetition is underpinned by comprehension, online language processing mechanisms, e.g. syntactic parsing, are likely to play a role. By contrast, some have argued that successful repetition may require only superficial comprehension (Lust, Chien, & Flynn, 1987). Given the uncertainty over cognitive processes involved in SR, it is difficult to determine the root causes of poor performance. Nonetheless, variation in error patterns across different populations suggests that SR data can elucidate underlying language difficulties.

This study will use SR to investigate the language difficulties of adolescents with SLI and autism spectrum disorders (ASD). Both of these developmental disorders are characterised by language and communication difficulties, which in the case of ASD occur in the context of impairments in reciprocal social interaction and restricted and repetitive behaviours and interests. The association between these disorders has recently been the focus of much research. Standard diagnostic criteria, e.g. the ICD-10 and the DSM-IV (American Psychiatric Association, 2000), allow for no overlap between these disorders. Thus, an SLI diagnosis is not made when a child presents with ASD. However, studies have suggested an association between SLI and ASD at both a genetic and a behavioural level. Family studies show ASD risk increases in siblings of children with SLI and vice-versa (Fombonne, Bolton, Prior, Jordan, & Rutter, 1997; Tomblin, Hafeman, & O'Brien, 2003). Genetic linkage studies have reported possible loci on chromosomes 7q and 13q associated with language impairment and ASD (Alarcón, Yonan, Gilliam, Cantor, & Geschwind, 2005; Bradford et al., 2001). Tager-Flusberg and colleagues have suggested that possible genetic association is reflected at the behavioural level. They have identified a subgroup of children with ASD who present with language impairments in the context of nonverbal skills within the average range; that is, a psychometric profile typical of SLI. These children present with poor performance on non-word repetition (Kjelgaard & Tager-Flusberg, 2001) and past-tense tasks (Roberts, Rice, & Tager-Flusberg, 2004), both important psycholinguistic markers of SLI. Consequently, there may be a high degree of overlap between the language phenotypes of these two groups. However, we cannot assume that similar performance on two clinical markers will necessarily extend to all aspects of structural language use. For example, a recent study observed that while receptive and expressive language are equally impaired in ALI, SLI individuals perform better on receptive than expressive tasks (Loucas et al., in press). Furthermore, many studies of the SLI / ALI overlap have employed standardised tests which fail to differentiate the effects of task demands, processing deficits and linguistic abilities. They may also lack the psycholinguistic sophistication to identify subtle differences between groups. Thus, there is a need for tasks that allow a more fine-grained analysis of psycholinguistic abilities.

While the current literature on the overlap between SLI and ALI has focused on word-level performance, e.g. syntactic morphology and non-word repetition, there has been very little research on the production and comprehension of complex syntax; structures which are manifested at the clause level, such as relative clauses, passives and questions. Difficulties with such structures are widely observed in children with SLI (e.g. Novogrodsky & Friedmann, 2006; Van der Lely & Battel, 2003). Furthermore, such difficulties may play an important role in diagnosis, as many language assessments investigate abilities at the clause level. While difficulties with complex syntax have been well-documented in SLI, little is known about the syntactic skills of children with ASD above and beyond morphosyntax. Therefore, a comparison of clause-level syntactic abilities is warranted.

Experimental issues

In order to investigate complex syntax we need to operationalise our definition of “complexity”. Miller and Chomsky (1963) used the notion of self-embedding whereby one phrasal category is nested inside another, or the ratio of terminal to non-terminal nodes. MacWhinney and Pleh (1988) described syntactic complexity in terms of perspective-switching. For example, in object-extracted relatives (example (2), below) the subject is first parsed as the agent, consistent with canonical word-order, then as the patient of the relative clause, and finally as the agent of the main clause. Perhaps the most widely used definition is based on syntactic dependencies. Numerous complex constructions, such as passives, questions, and relative clauses, are characterised by a long-distance relationship between moved arguments (or *fillers*), and the place where they derive their thematic role (the *gap*). In relative clauses this distance varies according to whether the head noun is the subject or the object;

(1) **The policeman** [who $_{t}$ chased the (tall thin) thief] $_{\text{Rel. Clause}}$ wore a (big black) hat (*Subject Relative*)

(2) **The thief** [who the (tall thin) policeman chased $_{t}$] $_{\text{Rel. Clause}}$ wore a (big black) hat (*Object Relative*)

Two opposing theoretical frameworks can account for difficulties with complex syntax in SLI. According to competence-based accounts children with SLI have difficulties with particular syntactic configurations. For example the Computational Grammatical Complexity account (Van der Lely, 2005) proposes a deficit affecting all dependent syntactic relationships, including the long-distance dependencies shown above. By contrast processing accounts invoke limitations in Working Memory (WM) to explain difficulties with complex syntax. Individuals with limited WM may be less efficient at reactivating a filler upon reaching the gap. For example, Just and Carpenter (1992) found a strong correlation between adult participants' performance on a reading span task, and their ability to interpret object relatives. Marinis, Roberts, Felser and Clahsen (2005) observed a similar correlation in L2 learners, this time using a backwards digit recall task.

A particular challenge of SR studies is how to quantify children's errors. Generally, coders count the number of sentences containing an error, or the number of errors per sentence, i.e. the number of words added, substituted, omitted, or transposed (see Vinther, 2002 for a discussion).The latter metric is clearly more sensitive to the degree of repetition difficulty. However, it is often extremely difficult to count errors. For example, if the child repeats *computers and printers* as *scanners and computers*, the coder may focus on the movement of *computers* to the position after *and,* and count it as a transposition error. *Printers* must then be deleted (1 error), and *scanners* added in front of *and* (1 error), making 3 errors. However the transformation is actually possible in 2 moves; substitution of *scanners* for *computers*, and substitution of *computers* for *printers*. This feels less intuitive than the first set of operations, but is clearly valid within the dictates of the coding scheme. Therefore, to ensure consistency and reliability the coder must ignore their intuitions, compare multiple sets of operations, and choose the smallest set. While this 3-word example appears complex, difficulties are greatly compounded for long complex utterances. In light of this, the widely used CELF Recalling Sentences subtest, counts only the first three errors. Yet this approach does not distinguish adequately between responses involving minor changes and those which radically alter sentence structure. Therefore, in order to consistently choose the smallest set of operations, this study will use an algorithm called the Levenshtein Distance (Levenshtein, 1966). While the original algorithm measured the distance between two strings in characters, it has been adapted to measure the Levenshtein Distance in words (LDw). Examples of the procedure are given in Appendix II. The algorithm counts the minimum number of words added, substituted or omitted, to transform one sentence into another.

In addition to the LDw, which was used as a broad metric of repetition difficulty, qualitative error categories were also developed, based on the observation that individuals who have yet to attain mastery of a particular construction, such as language-impaired, or very young children, tend to simplify syntactic structure during repetition, for example, by transforming object relatives into subject relatives (Kidd et al., 2007; Novogrodsky & Friedmann, 2006). Such errors could therefore provide a further indication of syntactic difficulties. Another possibility is that participants who find relative clauses difficult will not produce them at all. For example, they could omit the relative entirely, or express it is a coordinated clause, i.e. introduced by the conjunction *and*. This error was likewise coded.

Design and hypotheses

Subject and object relatives were used as they allow one to manipulate syntactic complexity without altering the length of the stimulus in words. In this sense Phonological STM load was kept constant, and therefore syntactic factors could be isolated. Complexity was operationalised in terms of the length of the dependency between the filler and gap, which was further manipulated by adding adjectives either within the relative clause or the main clause (see examples (1) and (2)). The main experimental hypotheses were -

(1) Participants will make more errors repeating object relatives due to long-distance dependencies.

(2) Repetition performance will be affected by the distance between the filler and the gap, consistent with WM accounts of processing difficulty.

(3) The effect of complexity will not differ between SLI and ALI consistent with arguments proposed by Kjelgaard and Tager-Flusberg (2001) and Roberts et al. (2004) that these two groups present with a similar language phenotype.

Method

Participants

The study investigated two clinical populations – adolescents with SLI, and high-functioning (close to average non-verbal IQ) adolescents with ASD plus language impairment (ALI). Participants were selected from a cohort of individuals with Special Educational Needs who had been assessed during the Special Needs and Autism (SNAP) Project (Baird et al., 2006). A diagnosis of autism was made on the basis of ICD-10 criteria (WHO, 1993) using the ADOS (Lord et al., 2000), ADI-R (Lord, Rutter, & Couteur, 1994), clinical vignettes and teacher report. Participants were diagnosed with language impairment if there was a discrepancy between their language abilities, measured using the CELF-3 UK (CELF-3 UK, Semel, Wiig, & Secord, 2000), and their non-verbal IQ scores on the Wechsler Intelligence Scale for Children-III (WISC-III: Weschler, 1992). The language cut-off was a standard score of 77 (-1.5 SD) or below on the expressive and/or receptive subscales, while the IQ cut-off was a standard score of 80 (-1.3 SD) or above on either Performance IQ, or the Perceptual Organisational Index. None of these individuals met the diagnostic criteria for any syndrome other than ASD or SLI, as ascertained via medical examination, interviews with teachers, and inspection of the paritipcants' medical records, and all of them were tested for hearing difficulties (<30dB).

Given the time lag between the SNAP study and the current study, on average 42 months, language and non-verbal abilities were retested using a shorter version of the previous assessments; Concepts and Directions (CD) and Recalling Sentences (RS) from the CELF, and Picture Arrangement (PA) and Block Design (BD) from the WISC. The two CELF subtests were chosen to measure expressive (RS) and receptive (CD) abilities, with the former being an especially reliable indicator of SLI (Conti-Ramsden et al., 2001). The WISC subtests were assessments of non-verbal ability with a minimal motor component. Four further participants with SLI were recruited via contacts in schools with language units. ASD status was determined using the ADOS (all participants), SCQ (all participants) and the ADI-R (participants 1, 2 and 3). Language and non-verbal abilities were assessed using the full WISC and CELF (participants 1 and 2), but due to time constraints participants 3 and 4 were administered the short forms. Hearing difficulties were assessed via teacher report.

Group characteristics are shown in Table 1. In addition to the screening assessments, participants were administered a series of STM / WM tests; the Children's Test of Non-word Repetition (CNRep: Gathercole & Pickering, 2001) and the Digit Recall (DR), Backwards Digit Recall (BDR) and Listening Recall (LR) from the Working Memory Test Battery for Children (Gathercole & Pickering, 2001). The LR test involves making true-false judgements about a series of sentences, and then recalling the last word of each sentence in the correct order. While clinical groups were matched for language and non-verbal abilities, the participants with SLI performed more poorly on the STM / WM tasks, obtaining significantly lower scores on DR ($F(1,28) = 6.15$, $p = .019$, Partial $\eta^2 = .220$), and presenting with a trend towards significance on BDR ($F(1,28) = 3.98$, p = .XXXPartial $\eta^2 = .142$).

Stimuli

Relative clause type and adjective position were combined to create four conditions; SM – subject relative with adjectives in the main clause, SR – subject relative with adjectives in the relative clause, OM – object relative with adjectives in the main clause, and OR – object relative with adjectives in the relative clause. The adjectives manipulated the length of the dependency in sentences containing object relatives, but not those containing subject relatives (see sentences (1) and (2)). The stimuli are listed in Appendix I, with relative clauses in square brackets and adjectives in bold. There were 24 sentences, 6 stimuli within each cell of the 2 x 2 design. They were generated from a set of 19 high frequency action verbs, 24 high frequency nouns referring to either people or animals, and 22 high-frequency adjectives. As the study focuses on processing complexity within the relative clause, the semantic properties of the main clause were kept relatively constant; *wear* or *have* was the main verb, and the object referred to an item of clothing, or a physical characteristic. A number of factors which may affect recall were controlled – length in phonemes, lexical frequency, and plausibility (see Appendix I). Plausibility measures were obtained from 15 adult native speakers who were asked to give ratings of 1 to 5 for the two possible argument orders within the relative clause, e.g. *the donkey who kicked the boy*, versus *the boy who kicked the donkey*. Plausibility was counter-balanced by assuring an even distribution of plausibility ratings within each condition. One-way ANOVAs did not find an effect of experimental condition (1 – 4) on either sentence length, mean lexical frequency or plausibility (all p values > .6), indicating minimal variation across conditions. The stimuli were recorded in a soundproof booth by a female native speaker of Southern British English at a rate of approximately 4 syllables per second. A condenser microphone was used, and the speech wave digitised at 44,100 Hz.

Procedure

The participants were told "Now I'm going to play some sentences. Listen to the sentences and repeat them. Your sentence must be <u>exactly the same</u> as the one you hear." The participants listened to 24 sentences, 6 in each condition of the 2 x 2 designed. The stimuli were presented over headphones (Pro-Luxe OA 850) using a DELL laptop computer. Four experimental orders were created in a pseudorandom fashion, so that there were no more than three consecutive sentences from each block in the 2 x 2 design. Within each experimental group participants were assigned to these four orders as evenly as possible given that not all groups were divisible by four.

Participants' errors were coded live on a scoresheet using transcription methods from the CELF preschool, e.g. omitted words were deleted with a straight line. In addition, all responses were audio-recorded onto a laptop computer using a clip-on omni-directional microphone. Adjustments were made to the live transcriptions based on the audio-recordings. Responses were entered orthographically into a computer database allowing for minor phonological changes, e.g. substitutions of a single phoneme. Novel or unintelligible words were coded orthographically yielding an error rate of one addition. Reformulated speech and intrusions,e .g. “I’ve forgotten” were excluded from the analysis.

Certain types of errors were not transcribed in the final version of the data, i.e. the corrected forms were entered into the database. Firstly errors related to tense morphemes, e.g. omission of past tense –ed or overregularisations, were transcribed in a corrected form. This is because, for language impaired individuals, the likelihood of making such errors may be dependent on the phonological / phonotactic properties of the verb stem (e.g. Marshall & van der Lely, 2006), and verb regularity, two factors not controlled in the current study. Secondly, aspectual auxiliaries, e.g. *was* wearing, which were not present in any of the stimuli, were removed from the final data. The removal of errors related to tense and auxiliary morphemes is warranted given the theoretical focus on clause-level abilities. Finally, errors involving changes in the indefinite article, e.g. a → an, an → a, were overlooked as they were derived from a phonological agreement rule and therefore not related to syntactic difficulties.

Interrater agreement measures were calculated for 8.5% of the recordings (5 participants altogether; 2 TD, 1 SLI, and 1 ALI). For 94.8% of repetition attempts (91 out of 96) the second rater found not discrepancy between the audio recording and the transcription. For the five repetition attempts where disagreement arose, relatively few words were disputed, with the second rater’s interpretation resulting in a change in LDw of between 0 and 2. Interrater measures were also obtained for the coding scheme for the qualitative analysis, with 100% agreement being obtained.

Analysis

Experimental design

Within-group investigations of error profiles were conducted using planned comparisons with orthogonal contrasts, i.e. the dataset is progressively subdivided at each stage of the analysis. This technique is suitable for investigating a series of contrasts based on a strong a priori hypothesis. In this case, an association between dependence length and repetition difficulty was hypothesized for all groups, such that;

$$OR > OM > SR = SM$$

where ">" signifies a greater error rate. Planned comparisons were conducted in three stages; object versus subject relatives, SR versus SM, and OR versus OM. The advantage of planned comparisons over post hoc contrasts, is that, for analyses based on a priori hypotheses, they are statistically more precise because they avoid over- or under-correction for the possibility of a Type I error, i.e. accidental rejection of the null hypothesis. Differences in error profiles across the experimental groups were investigated using a repeated measures ANOVA investigating the interaction between Group and Complexity, with Group as the between-subjects factor, and Complexity as the within-subjects factor. The ANOVA method was deemed appropriate for all analyses where the homogeneity of variance assumption was met (Levene's test $> .05$).

Between-group analysis of raw error rates

The mean error rates by group and sentence type are shown in Figure 1. Kruskall-Wallis tests confirmed that the participants with SLI and the participants with ALI both made significantly more errors than the TD participants (χ^2 (1)= 20.5, p < .001, and χ^2 (1)= 18.0, p < .001 respectively). By contrast overall error rates did not vary significantly across the SLI and ALI groups (F (1, 28)= 1.69, p = .204, partial η^2 = .057). Given the low error rates in the TD group, their data was excluded from further analyses.

Within-group analysis of raw error profiles

Planned comparisons, consisting of 3 one-way ANOVAs per group, were conducted. The first comparison investigated the overall effect of complexity on mean error rates per participant (LDw). A significant effect was observed in the SLI group (F(1,54) = 7.81, p = .010, Partial η^2=.117), but not the ALI group (F(1,62) = 2.28, p = .236 Partial η^2=.035). The second comparison investigated the effect of adjective position on error rates in subject relatives. A significant effect was observed in the SLI group (F(1,26) = 5.22, p = .032, Partial η^2=.167), but not the ALI group (F(1,30) = 1.25, p = .272, Partial η^2=.040). The final comparison investigated the effect of adjective position on error rates in object relatives. Here also, a significant effect was observed in the SLI group (F(1,26) = 4.86, p = .036, Partial η^2 = .158), but not the ALI group (F(1,30) = 3.15, p = .086, Partial η^2 = .095).

Between-group analysis of raw error profiles

A 2 x 2 factorial repeated measures ANOVA was conducted investigating the interaction between ASD status and syntactic complexity (subject versus object relatives). There was a significant main effect of Complexity (F(1,28) = 65.3, p = .001, partial η^2 = .700), and a significant Group x Complexity interaction (F(1, 28) = .581, p = .023, partial η^2 = .172). The estimated marginal means, plotted in Figure 2, demonstrate that this interaction is driven by a larger effect of complexity in the SLI group.

Between-group qualitative analyses

Error rates for qualitative error categories are shown in Table 3. A complete relative clause required a verb, an argument (i.e. subject or object), and a relativiser, except in the following cases; relativisers could be omitted from object relatives, consistent with English syntax, and agents could be omitted from relative clauses containing a reduced passive. Interestingly, passives within the relative clause were relatively common, with 6 children with SLI and 3 children with ALI producing a total of 13 responses containing either full or reduced passives. Intriguingly, all of these responses were attempted repetitions of object relatives, and passivisation appeared to be used as a strategy to avoid producing object relatives but maintain thematic relationships, e.g. *the granny that the thief robbed* → *the granny that was robbed by the thief.*

A series of one-way ANOVAs was conducted with group as the independent variable, and the percentage error rates, as shown in Table 3, as the dependent variable. The rate of object to subject relative transformations was significantly higher in the SLI group ($F(1,28) = 4.50$, $p = .043$, partial $\eta^2 = .138$). No significant between-group differences were observed for the rate of incomplete / null responses ($F(1,28) = 4.50$, $p = .596$, partial $\eta^2 = .010$) or errors involving incomplete or omitted relative clauses ($F(1,28) = 2.60$, $p = .118$, partial $\eta^2 = .085$).

Analysis of memory factors associated with performance

Pairwise correlations, shown in Table 4, were conducted to investigate the association between raw scores on the memory tests and performance on SR. Analyses were run for the clinical groups separately and combined, and the CELF RS task was included to compare the psychometric properties of the current stimuli with those of a widely used assessment. It is important to note that poor SR performance led to high LDw scores on the experimental task, but low scores on the CELF, hence coefficients are negative in the former, but not the latter. Three of the STM / WM tasks, CNRep, DR, and BDR correlated significantly with performance on the experimental task in either or both of the clinical groups. In general, correlation coefficients did not differ greatly across the groups. Overall, correlation coefficients were weaker for the CELF RS test, suggesting than the experimental task.

Discussion

The clinical groups, as expected, presented with higher error rates than the typically developing controls. A profile emerged such that greater errors were made for object relatives versus subject relatives, and sentences with adjectives in the main clause versus adjectives in the relative clause. This profile was more pronounced in the SLI group than the ALI group with significant effects of complexity and adjective position observed in the former, but not the latter, and a significant group by complexity interaction. Qualitative analyses were consistent with the data from error rates with the SLI group producing significantly more object to subject relative transformations than the ALI group; 17% of responses versus 6%. High correlations were observed between SR error rates and error rates on three tests of STM / WM, the CNRep, DR and BDR.

Implications for our understanding of the SR task

Consistent with previous research, SR demonstrated its potential as a sensitive phenotypic maker of language impairments, with high error rates in the clinical groups, and low error rates in the TD participants. Such difficulties are striking given the age of the participants, suggesting that SR may continue to be a sensitive marker of language impairments at least into early adolescence and possibly beyond. Likewise the strong effect of syntactic complexity in the clinical groups indicates the severe and persistent nature of their syntactic difficulties. The finding that SR difficulties extend to the ALI group supports the claim that SR may be useful clinical marker of language-impairment in a variety of different populations who experience language difficulties.

The effect of syntactic complexity on SR error rates is consistent with previous research suggesting that this assessment paradigm is sensitive to syntactic factors. Likewise, the tendency to transform object into subject relatives replicates previous studies, with the rate of object to subject transformations slightly higher than that observed by Novogrodsky and Friedmann (2006); 17% versus 10%. Errors with particular syntactic structures may provide an insight into an individual's syntactic competence, and their linguistic representations in LTM (Slobin & Welsh, 1968). A tendency to simplify syntactic structure during repetition may indicate that the more complex object relative clause is poorly represented in the syntactic knowledge of the participants. Relative clauses are particularly well-suited to investigating syntactic complexity, as they may vary in complexity while controlling for phonological length. Therefore the error profiles in the current study cannot be explained in terms of a simple model of PhSTM, for example a mechanism which involves storing sentences in phonological form and recalling them as a whole. Nonetheless, a strong association was observed between SR and non-word repetition performance suggesting that PhSTM is involved in SR on some level, or that SR and non-word repetition performance are influenced by the same underlying cognitive factor.

Adjective position likewise affected performance, with a significant effect of adjective position within object relatives for the SLI participants. Adjectives placed within the relative clause increase the length of the dependency, thus placing a greater burden on WM. This may result in poor comprehension, which in turn affects repetition. However, this account is contradicted by the finding that adjectives placed in subject relative clauses, which do not affect dependency length, nonetheless impacted on error rates, with a significant effect in the SLI group. Furthermore, differences across the OM and SR conditions were clearly small, despite the longer dependency in the former. The effect of adjective position, independent of complexity suggests that an additional determinant of processing difficulty is the distance between the subject and the verb in the main clause, which are separated by the relative clause. This can be regarded as a syntactic dependency, in that the thematic role of subject is unspecified until we encounter the main verb. WM may be taxed by this kind of structure because a representation of the subject must be actively maintained before it can be integrated with the main verb.

While it is interesting to speculate on the origins of SR difficulties we still know little about the cognitive mechanisms underpinning SR. For example, the extent to which repetition depends on successful comprehension is not known, and therefore we cannot necessarily assume that the effect of dependency length is attributable to parsing difficulties during comprehension. In addition, adjectives placed within the relative clause may affect repetition for reasons unrelated to their location within syntactic structure. For example, being semantically abstract, adjectives are more difficult to recall than concrete words such as nouns. Participants may therefore have struggled to recall the adjectives, causing them to falter and forget the rest of the sentence which may be rapidly fading from memory. In this way, the adjectives placed within the relative clause may have adversely affected recall by virtue of their serial position, and not their syntactic position. In addition, adjectives within relative clauses are pragmatically unusual given that relative clauses provide backgrounded information, and are therefore referentially "light." Such pragmatic factors may also have affected repetition difficulty. Given the current focus on syntax, pragmatic considerations were overlooked. It is clear therefore, that while adjective placement consistently affected error rates, interpretation of this profile is problematic.

Similarities and differences between SLI and ALI

The study has important implications for the literature on the association between SLI and ALI. The profiles of the two groups were qualitatively similar, with error rates increasing in response to both syntactic complexity and adjective position. Therefore, the participants with ALI present with a similar kind of syntactic impairment, affecting structures with long-distance dependencies / non-canonical word order such as object relatives. This finding is consistent with the claim that there is a strong overlap in the language phenotype of these two groups (Kjelgaard & Tager-Flusberg, 2001; Roberts et al., 2004). However, the participants with SLI were significantly more affected by syntactic complexity, and were significantly more likely to make wholesale changes to syntactic structure. In this sense the participants with SLI presented with greater syntactic difficulties than those with ALI, a finding which contradicts the hypothesis that there is a strong phenotypic overlap. Qualitative studies which go beyond total error counts by investigating error rates across different kinds of stimuli, or coding for different kinds of errors, have the potential to identify differences in language phenotypes across different groups where studies investigating raw error rates have failed to identify differences.

The exact origin of the differences in the profiles of the groups is difficult to determine. A recent study identifying qualitative differences in the non-word repetition performance of children with ALI and SLI, Whitehouse, Barry and Bishop (2008), proposed that such differences may be due to underlying cognitive variation. While the children with SLI presented with a convex error profile consistent with PhSTM limitations, characterised by rapidly decreasing performance on the longer words, the children with ALI did not present with a significant length by error rate interaction. The authors suggest that this qualitatively different profile reflects different cognitive limitations, for example an attentional deficit, which affects repetition irrespective of length. With regard to the current study, while the reduced effect of syntactic complexity in the ALI group may indicate a milder syntactic deficit, it is also possible that a flatter profile may be due to the role of an additional factor such as attention. One possible outcome of poor attention is that the child loses concentration altogether, and produces a null response. However, the two groups presented with similar rates of such responses. By contrast, DR scores did differ significantly across the two groups, and to a lesser extent BDR, with poorer performance in the SLI group, suggesting that STM may be an important cognitive factor affecting differences in error profiles. This observation is supported by the high correlation between performance on the experimental SR task, and the assessments of STM / WM, although one cannot necessarily infer a causal relationship from this analysis. Another possibility is that while general SR difficulties may reflect STM / processing limitations, the greater syntactic difficulties of the children with SLI reflect the added contribution of a specific syntactic deficit (Van der Lely, 2005). Further research is clearly needed to identify differences in the cognitive profiles of ALI and SLI groups, and how these relate to language performance.

The study has important implications for the SR paradigm. In addition to its diagnostic properties, it also demonstrated sensitivity to syntactic structure, indicating its potential as means of investigating syntactic abilities. However, it is as yet difficult to determine whether syntax-related SR difficulties are due to poor syntactic competence or processing limitations, e.g. STM / WM difficulties. An important implication of the study is that SR can provide qualitative information. For example, error rates across different structures may be investigated, and particular types of syntactic errors may be counted. The sensitivity of SR to qualitative factors, e.g. syntactic structure, may be enhanced using an algorithm which quantifies the distance between sentences, such as the LDw. By producing total error rates without arbitrary cut-offs, such algorithms may provide a reliable metric of repetition difficulty. Furthermore, the LDw demonstrated good concurrent validity, correlating highly with tests of STM; Digit Recall, and the CNRep. These correlations were higher than those of the CELF, though this may be due to the stimuli themselves, which were designed to stress STM by incorporating long-distance dependencies, rather than the scoring algorithm. Construct validity was also demonstrated, in that the LDw data matched the qualitative analyses, in that both revealed greater syntactic difficulties in the SLI group. Given its reliability and validity, the LDw, or related algorithms may prove a useful means of analysing data in future SR studies, and could also be employed in future clinical assessments.

Limitations

It is important to bear in mind the limitations of the study. Firstly, no filler stimuli were used, which differ structurally from the target stimuli and therefore control for priming effects. The SR test was administered as part of a larger battery and therefore stimuli were kept to a minimum. However, filler items are recommended for future SR studies. Secondly, the SLI group performed significantly worse on DR, and also performed comparatively poorly on BDR, indicating differences in STM /WM abilities. It is not certain whether this reflects genuine differences between the populations, or the effect of random variation on a relatively small sample. It would be interesting to note whether such differences are observed in future studies comparing ALI and SLI.

Directions for future research

Given the sensitivity of SR to syntactic structure, and its ability to distinguish between different language-impaired populations, it may prove to be an important paradigm in the investigation of language difficulties. Performance on different kinds of complex structures, e.g. passives, and questions may be investigated. Such studies should be complemented by further work investigating the cognitive basis of the paradigm itself. In particular it is important to disentangle the roles of syntactic knowledge, and processing factors.

With regard to the debate on the language phenotype overlap between SLI and ALI, more research is clearly needed to investigate possible differences. The findings of the current study suggest that although there is substantial overlap between children with language impairment who meet diagnostic criteria for ASD and those who meet diagnostic criteria for SLI, there are also differences between the two groups that when fully determined may inform epidemiology, genetics and clinical practice.

Acknowledgements

The authors wish to thank Autism Speaks / the National Alliance for Autism Research for their generous funding, the parents / guardians and individuals who participated, and Susie Chandler, Abigail Davison-Jenkins, Ann Ozsivadjian, and Vicky Slonims for their help in screening the participants with ASD. The first author wishes to thank Colin Bannard at the Max Planck Institute, Leipzig, for explaining the concept of the Levenshtein Distance, and Michael Gilleland of Merriam Park software for publishing a procedure to calculate the Levenshtein Distance on the internet, which was used during the data analysis. We would also like to thank the two anonymous reviewers for their valuable comments on the manuscript.

References

ALARCÓN, M., YONAN, A. L., GILLIAM, T. C., CANTOR, R. M., and GESCHWIND, D. H. (2005). Quantitative genome scan and Ordered-Subsets Analysis of autism endophenotypes support language QTLs. *Molecular Psychiatry*, 10, 747-757.

AMERICAN PSYCHIATRIC ASSOCIATION. (2000). *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR)*. . Washington, DC.: American Psychiatric Association.

BAIRD, G., SIMONOFF, E., PICKLES, A., CHANDLER, S., LOUCAS, T., MELDRUM, D., and CHARMAN, T. (2006). Prevalence of Pervasive Developmental Disorders in a Population Cohort of Children in South East Thames. The Lancet, 368, 210-215

BOOTH, J. R., MACWHINNEY, B., and HARASAKI, Y. (2000). Developmental differences in visual and auditory processing of complex sentences. *Child Development*, 71(4), 981 - 1003.

BRADFORD, Y., HAINES, J., HUTCHESON, H., GARDINER, M., BRAUN, T., SHEFFIELD, V., CASSAVANT, T., HUANG, W., WANG, K., and VIELAND, V. (2001). Incorporating language phenotypes strengthens evidence of linkage to autism. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 105, 539-547.

CLAY, M. M. (1971). Sentence Repetition: Elicited Imitation of a Controlled Set of Syntactic Structures by Four Language Groups. *Monographs of the Society for Research in Child Development*, 36(3), 1-85.

CONTI-RAMSDEN, G., BOTTING, N., and FARAGHER, B. (2001). Psycholinguistic markers for Specific Language Impairment (SLI). *Journal of Child Psychology and Psychiatry*, 42, 741-748.

FOMBONNE, E., BOLTON, P., PRIOR, J., JORDAN, H., and RUTTER, M. (1997). A family study of autism: cognitive patterns and levels in parents and siblings. *Journal of Child Psychology and Psychiatry*, 667-683.

FRIEDMANN, N., and NOVOGRODSKY, R. (2004). The acquisition of relative clause comprehension in Hebrew: A study of SLI and normal development. *Journal of Child Language*, 31, 661 - 681.

GARDNER, H., FROUD, K., MCLELLAND, A., and VAN DER LELY, H. K. (2006). Development of the Grammar and Phonology Screening (GAPS) test to assess key markers of specific language and literacy difficulties in young children. *International Journal of Language and Communication Disorders*, 41(5), 513 - 540.

GATHERCOLE, S. E., AND PICKERING, S. (2001). *Working Memory Test Battery for Children (WMTB-C)*. Harcourt Assessment.

GRANT, J., VALIAN, V., and KARMILOFF-SMITH, A. (2002). A study of relative clauses in Williams syndrome. *Journal of Child Language*, 29(02), 403-416.

JEFFERIES, E., LAMBON-RALPH, M., and BADDELEY, A. D. (2004). Automatic and controlled processing in sentence recall: The role of long-term and working memory. *Journal of Memory and Language*, 51(4), 623 - 642.

JUST, A., AND CARPENTER, P. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, 99(1), 122 - 149.

KIDD, E., BRANDT, S., LIEVEN, E., and TOMASELLO, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes*, 22(6), 860 - 897.

KJELGAARD, M. M., and TAGER-FLUSBERG, H. (2001). An investigation of language impairment in autism: implications for genetic subgroups. *Language and Cognitive Processes*, 16(2 / 3), 287 - 308.

LEVENSHTEIN, V. I. (1966). Binary codes capable of correcting delections, insertions, and reversals. *Soviet Physics - Doklady*, 10, 707 - 710.

LOMBARDI, L., AND POTTER, M. C. (1992). The regeneration of syntax in short term memory. *Journal of memory and language*, 31(6), 713-733.

LORD, C., RISI, S., LAMBRECHT, L., COOK, E. H., LEVENTHAL, B. L., DILAVORE, P. C., et al. (2000). The Autism Diagnostic Observation Schedule, Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205-223.

LORD, C., RUTTER, M., & COUTEUR, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*(5), 659-685.

LOUCAS, T., CHARMAN, T., PICKLES, A., SIMONOFF, E., CHANDLER, S., MELDRUM, D., and BAIRD, G. (in press) .Autistic symptomatology and language ability in Autism Spectrum Disorder and Specific Language Impairment. *Journal of Child Psychology and Psychiatry*.

LUST, B., CHIEN, Y. C., and FLYNN, S. (1987). What children know: Methods for the study of first language acquisition. *Studies in the Acquisition of Anaphora*, 2, 271-–356.

MACWHINNEY, B., and PLEH, C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition*, 29, 95-141.

MARINIS, T., ROBERTS, L., FELSER, C., and CLAHSEN, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, 27, 53 - 78.

MARSHALL, C. R., and VAN DER LELY, H. K. J. (2006). A challenge to current models of past tense inflection: The impact of phonotactics. *Cognition*, 100(2), 302-320.

MARSHALL, C., AND NATION, K. (2003). Individual differences in semantic and structural errors in in children's memory for sentences. *Educational and Child Psychology*, 7 - 18.

MCDADE, H. L., SIMPSON, M. A., and LAMB, D. E. (1982). The use of elicited imitation as a measure of expressive grammar: a question of validity. *Journal of Speech and Hearing Disorders*, 47, 19 - 24.

MILLER, G. A., and CHOMSKY, N. (1963). Finitary Models of Language Users. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of Mathematical Psychology*, 2 (pp. 419 - 491). Wiley, NY.

NOVOGRODSKY, R., and FRIEDMANN, N. (2006). The production of relative clauses in syntactic SLI: a window to the nature of the impairment. *Advances in Speech Language Pathology*, 8(4), 364 - 375.

POTTER, M. C., and LOMBARDI, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, 29(6), 633 - 654.

POTTER, M. C., and LOMBARDI, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, 38, 265 - 282.

ROBERTS, J. A., RICE, M. L., and TAGER-FLUSBERG, H. (2004). Tense marking in children with autism. *Applied Psycholinguistics*, 25, 429 - 448.

RUMMER, R. (2004). Immediate and delayed recall of visually presented sentences: evidence for the involvement of phonological information. *Experimental Psychology*, 51(1), 14-22.

SEEFF-GABRIEL, B., CHIAT, S., and DODD, B. (2005). The relationship between speech disorders and language. In*Differential Diagnosis and Treatment of Children with Speech Disorders (2nd Edition).* (pp. 100 - 116). London: Whurr.

SEEFF-GABRIEL, B., CHIAT, S., and ROY, P. (2008). *The Early Repetition Battery*. Harcourt Assessment.

SEMEL, E., WIIG, E. H., and SECORD, W. (2000). *Clinical Evaluation of Language Fundamentals - Third Edition UK (CELF-3UK)*. Harcourt Assessment.

SLOBIN, D. I., and WELSH, C. A. (1968). Elicited imitation as a research tool in developmental psycholinguistics. *Working papers of the Language Behavior Research Laboratory, University of California, Berkeley* , 10.

TOMBLIN, B., HAFEMAN, L. L., and O'BRIEN, M. (2003). Autism and autism risk in siblings of children with specific language impairment. *International Journal of Language and Communication Disorders*, 38(3), 235 - 250.

VAN DER LELY, H. K. J. (2005). Domain-specific cognitive systems: Insight from grammatical specific language impairments. *Trends in Cognitive Sciences*, 9, 153 - 181.

VAN DER LELY, H. K. J., and BATTEL, J. (2003). Wh-movement in children with grammatical SLI: A test of the RDDR hypothesis. *Language*, 79, 153 - 181.

VINTHER, T. (2002). Elicited imitation: a brief overview. *International Journal of Applied Linguistics*, 12(1), 54 - 73.

WECHSLER, D. (1992). *Wechsler Intelligence Scale for Children - 3rd UK Edition (WISC-III UK)*. Pearson.

WHITEHOUSE, A. J. O., BARRY, J. G., and BISHOP, D. V. M. (2007) Further defining the language impairment of autism: Is there a specific language impairment subtype? *Journal of Communication Disorders*, 4, 319 - 336.

WILLIS, C., and GATHERCOLE, S. E. (2001). Phonological short-term memory contributions to sentence processing in young children. *Memory*, 9(4/5/6), 349-363.

Appendix I – List of Stimuli

Cond	Sentence	Mean length (syll.)	Mean freq. of content words[1]	Mean plaus-ibility rating
SM	The monster [that killed the prince] wore a **bright green** cloak	13.5	1.92	2.88
	The donkey [that rode the boy] wore an **old yellow** hat	(mean)	.239	1.37
	The woman [that fed the child] wore some **bright blue** trousers	1.38		
	The thief [that watched the detective] wore a **bright red** scarf	(s.d.)		
	The criminal [that robbed the doctor] wore some **dark brown** boots			
	The teacher [that shouted at the girl] wore a **dark purple** jacket			
OM	The child [that the mother hugged] wore some **bright green** trousers	13.7	1.86	2.27
	The prince [that the princess rescued] wore a **bright gold** crown	.816	.175	1.14
	The cat [that the dog chased] wore a **bright yellow** collar			
	The soldier [that the criminal shot] wore a **bright green** hat			
	The mother [that the child carried] wore a **short yellow** skirt			
	The housewife [that the fireman rescued] wore a **bright green** T-shirt			
SR	The boy [that kicked the **big old** donkey] wore a hat	13.5	1.92	2.88
	The child [that hugged the **fluffy old** teddy] wore a cap	1.38	.239	1.37
	The policeman [that shot the **tall thin** thief] wore a hat			
	The thief that [caught the **small chubby** policeman] wore a tie			
	The patient [that examined the **tall thin** nurse] had a hat			
	The criminal [that attacked the **tall thin** granny] wore some glasses			
OR	The granny [that the **tall thin** thief robbed] wore some shoes	14.2	1.75	2.8
	The artist [that the **short fat** housewife painted] wore a cap	1.33	.113	1.16
	The granny [that the **small chubby** boy pushed] wore some glasses			
	The baby [that the **nice friendly** mother kissed] wore a hat			
	The policeman [that the **short chubby** thief chased] had a moustache			
	The criminal [that the **short heavy** policeman hit] had a beard			

1. Frequency per million words based on CELEX database. Log values are shown.

Appendix II

Examples demonstrating the LDw

(words involved in operations are italicized)

Stimulus	**Response**	**LDw**
The prince *that the princess rescued* wore a bright gold crown	The prince wore a bright gold crown	4 = deletion of “that the princess rescued” (4 deletions)
The *housewife* that *the fireman rescued* wore a bright *green* T-shirt	The *fireman* that *rescued a nurse* wore a bright *yellow* T-shirt	5* = substitution of “housewife” by “fireman” (1 substitution), substitution of “the fireman rescued” by “rescued a nurse” (3 substitutions), substitution of “green” by “yellow” (1 substitution).
The granny *that the small chubby boy pushed wore some* glasses	The granny *had dark blue* glasses	8 = substitution of “that” by “had” (1 substitution) deletion of “the small chubby boy pushed” (5 deletions), substitution of “dark blue” by “wore some” (2 substitutions)
The child that hugged the *fluffy old* teddy wore a *cap*	The child that hugged the *old fluffy* teddy *bear* wore a *green hat*	5 = substitution of “fluffy old” by “old fluffy” (2 substitutions), addition of “bear” and “green” (2 additions), substitution of “cap” by “hat” (1 substitution)
The child that hugged the fluggy old teddy wore a cap	(no response)	11 = omission of all 11 words.

*NB although it looks as if “rescued” has moved, the minimum distance is obtained via substitution

Table 1 – Descriptive statistics – mean, standard deviations, and minimum / maximum values

	TD (n = 17, 7 female)	SLI (n = 14, 1 female)	ALI (n = 16, no females)	P-values for ANOVAs comparing clinical groups
Age	14;4 4.2 14;0 – 14;11	15;3 7.49 14;5 – 16;7	14;8 5.77 14;0 – 15;4	
WISC – mean subtest standard score	11.2 2.89 7 – 16.5	11.9 2.25 8.5 – 17.5	11.6 2.44 8 - 16	
CELF – mean of subtest standard scores	9.59 1.75 7.5 – 14	4.07 .958 3 - 6	4.60 1.08 3 – 6.5	
CELF –RS raw scores	28.2 1.47 25 - 30	22.1 3.48 12 - 28	21.2 4.71 11 - 27	p = .538
CELF CD – Recalling Sentences raw scores	63.5 6.84 49 - 75	34.1 7.86 24 - 54	38.4 11.3 18 - 58	p = .244
CNRep raw scores (out of 40)	33.4 3.35 23 - 37	25.9 6.99 9 - 35	28.9 4.33 21 - 36	p = .170
DR raw scores (out of 54)	36.7 7.49 24 - 52	26.4 3.8 20 - 32	30.8 5.46 22 - 44	p = .019*
BDR raw scores (out of 42)	18.7 5.93 12 - 30	11.2 3.19 6 - 17	14.2 4.71 7 - 20	p = .056
LR raw scores (out of 36)	16.5 3.22 11 - 26	12 5.05 5 - 25	12.94 4.31 5 - 20	p = .616

Table 2

Summary statistics for error rates (LDw) by Group and Condition

Data show means, standard deviations, and range (minimum and maximum values)

TD				SLI				ALI			
SM	SR	OM	OR	SM	SR	OM	OR	SM	SR	OM	OR
.412	.598	.578	.961	2.49	4.08	3.93	5.72	2.49	3.23	2.99	4.5
.531	.804	.662	1.14	1.69	1.99	2.32	1.98	1.79	2.05	2.10	2.50
0 - 1.83	0 - 3.33	0 - 2	0 - 3.83	0.5 - 5.17	1 - 8	0.5 -8.8 3	2.66 - 8.67	0.5 - 6.67	0.5 - 8	.667 -6.5	1.67 -10. 3

Table 3 – Qualitative Errors by Group

Figures show percentage of responses in each error category

	SLI	ALI
Incomplete / null utterances (3 or fewer words)	5.7 (mean) .231 (s.d.) 0 – 25 (min-max)	4.7 .212 0 - 33
Responses without a full relative clause*	12.5 33.1 0 – 46	7.03 25.6 0 – 33
Responses transforming an object relative into a subject relative	16.4 13.8 0 – 42	6.78 10.0 0 - 33
Responses transforming a subject relative into an object relative	1.03 2.61 0 – 8	1.37 3.94 0 - 14

* Calculated as a percentage of complete responses

Table 4 – Correlations between memory raw scores and SR scores

	SLI and ALI groups (n = 30)		SLI group (n = 14)		ALI group (n = 16)	
	Exp. Task	CELF RS	Exp. Task	CELF RS	Exp. Task	CELF RS
CNRep	*-.528*	*.204*	*-.504*	*.021*	*-.529*	*.315*
	p =.003**	.280	.066	.943	**.035***	.234
DR	*-.544*	*.268*	*-.642*	*.201*	*-.434*	*.315*
	.002**	.153	**.013***	.490	.093	.466
BDR	*-.562*	*.387*	*-.519*	*.379*	*-.540*	*.323*
	.001**	**.035***	.057	.182	**.031***	.223
LR	*-.256*	*.017*	*-.170*	*-.237*	*-.313*	*.159*
	.171	.931	.562	.415	.238	.556

Figure 1 – Mean error rate by group and condition, with standard error bars

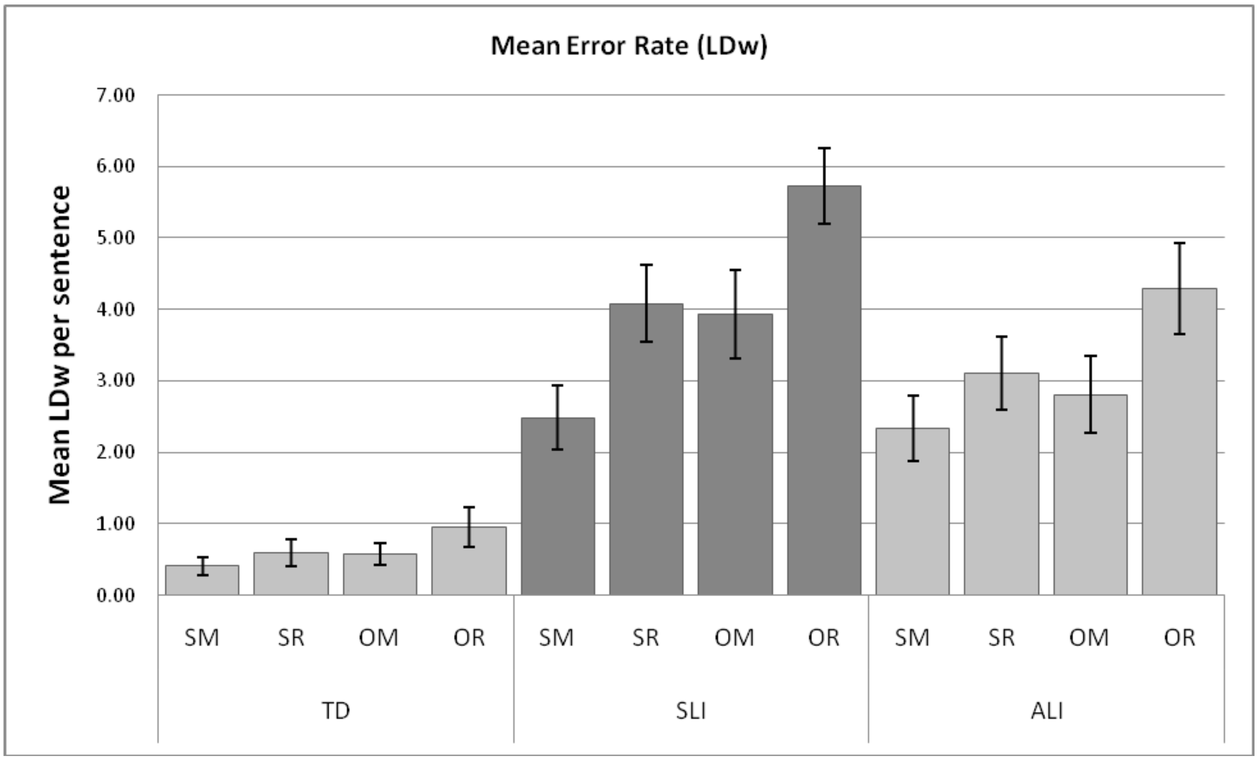

Figure 2 – Plot of interaction between Group and Complexity

with standard error bars