Short Title: Sentence interpretation in children

"Whatdunit?"

Developmental Changes in Children's Syntactically-Based Sentence Interpretation Abilities and Sensitivity to Word Order

James W. Montgomery
Communication Sciences and Disorders
Ohio University
Athens, OH 45701
740-593-1412
montgoj1@ohio.edu

Julia L. Evans
School of Behavioral and Brain Sciences
University of Texas-Dallas
Richardson, TX 75080
858-952-6104
jle130030@utdallas.edu

Ronald B. Gillam
Communication Disorders and Deaf Education
Utah State University
Logan, UT 84322
435-797-1704
ron.gillam@usu.edu

Alexander V. Sergeev
Social and Public Health
Ohio University
Athens, OH 45701
740-593-0635
sergeev@ohio.edu

Mianisha C. Finney
Communication Sciences and Disorders
Ohio University
Athens, OH 45701
740-590-5503
mf874294@ohio.edu

Correspondence:
Jim Montgomery
Communication Sciences and Disorders
Ohio University
montgoj1@ohio.edu

Key Words: children, development, sentence interpretation, word order, semantic implausibility

Abstract

Aim 1 of this study was to examine the developmental changes in typically developing English-speaking children's syntactically-based sentence interpretation abilities and sensitivity to word order. Aim 2 was to determine the psychometric standing of the novel sentence interpretation task developed for this study, as we wish to use it later with children with Specific Language Impairment (SLI). Children listened to semantically implausible sentences in which noun animacy and the natural affordance between the nouns were removed, thus controlling for event probability. Using this novel "whatdunit?" agent selection task, 256 children 7-11 years listened to two structures with canonical word order and two with non-canonical word order. After each sentence, children selected as quickly as possible the picture of the noun they believed was "doing the action." Children interpreted sentences with canonical word order with greater accuracy and speed than those with non-canonical word order. Older children (Age_M = 10:8) were more accurate and faster than younger children (Age_M = 8:1) across all sentence forms. Both older and younger children demonstrated similar error patterns across sentence type. The "whatdunit?" task also proved to have strong validity and reliability, making it suitable for studies with children with SLI.

Language is one of the most complex skills acquired by humans. Developmentally, sentence interpretation is an especially challenging feat as children must learn to make immediate sense of a fleeting acoustic signal. Successful interpretation requires children to incrementally build structure and meaning in the moment. Their mastery of sentence interpretation emerges slowly as they learn more about the regularities, cues, and constraints of their native language, and it is not until adolescence that children show adult-like performance across a variety of sentence forms. The main aim of this study was to examine developmental changes in typically developing English-speaking children's syntactically-based sentence interpretation abilities and use of word order cues. We measured children's accuracy and speed of interpretation of two sentence structures with canonical word order (subject-verb-object, subject relative) and two structures with non-canonical word order (passive, object relative). We created a novel "whatdunit?" agent selection task in which children were asked to interpret semantically implausible sentences. By removing event probably cues inherent in animate nouns, children's syntactic interpretation abilities based on word order cues could be highlighted. A second aim was to determine the psychometric soundness of the task. Ensuring its validity and reliability will allow us to use it with children with SLI to examine their sentence interpretation abilities.

Many researchers propose that adults and children begin sentence interpretation from sentence onset (Marslen-Wilson & Tyler, 1989; Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Zwitserlood, 1989; Zwitserlood, 1989), with listeners developing structure and meaning on a word-by-word basis (Borovsky, Elman, & Fernald, 2012; Elman, 1990; Traxler & Tooley, 2007). According to these authors, creating structure and meaning is influenced by various linguistic and non-linguistic cues and constraints.

The Competition Model of Sentence Interpretation: A General Framework for the Current Study

Bates, MacWhinney, and colleagues (Bates & MacWhinney, 1987, 1989; Bates, MacWhinney, Caselli, Devescovi, Natale, et al., 1984) proposed the Competition Model, an interactive activation model, as an account of cross-linguistic differences in sentence processing. The model assumes that the listener interprets the meaning of a sentence by calculating the probabilistic value of multiple linguistic cues in a sentence such as word order, morphology and semantic characteristics (e.g., animacy) and the listener's final interpretation of the sentence is based on the coalition of linguistic cues having the highest likelihoods. Qualitative and quantitative differences in sentence comprehension in children and adults across various languages have been described using the model.

Three key constructs- cue validity, cue strength, and cue cost- are included in the model. Cue validity relates to the information value of a given cue or type of information (e.g., preverbal position, postverbal position) for a specific function or meaning (e.g., agent, patient). Cue validity is the product of its availability and reliability, with availability relating to how often a cue is present and reliability relating to how often the cue leads to the proper interpretation. Cue strength is a property of an individual listener (in our case, a child), with cue strength increasing as cue validity increases. The mapping between a cue and its function/meaning can be many-to-many, and can present initial problems for young language learners. For example, the cue of noun (N) animacy tends to mark the agent role in many cases (e.g., *The boy hit the ball*), but animacy is not always present (e.g., *The bat hit the ball*) or correct in marking agency (e.g., *The ball hit the boy*). The strength of a cue for a listener is proportional to the information value of that cue. The third construct is cue cost, which relates to the processing cost associated with the immediate (real-time) use of a given cue, i.e., how difficult the cue is to use. The processing cost of different cues can differ across listeners and can reflect differences in their information processing abilities (e.g., memory, attention). Cue validity and cue cost affect the degree to which children ‘trust’ certain cues and the developmental order in which children learn to rely on different cues. The first cues children learn tend to be those that have strong validity and highly reliable. Cues can merge into coalitions that are stronger than any single cue by itself. Developmentally, the competition model posits that children learn these coalitions of cue-function mappings implicitly from their input language and adjust the weights of the different mappings over time with increased exposure to their native language.

Development of Sentence Interpretation Abilities

Even though children take considerable time to reach adult-like status in sentence interpretation, their performance is not random prior to mastery. Even very young toddlers seem to show sensitivity to a variety of cues as they try to make linguistic sense of what they hear. For example, young toddlers appear to use animate Ns to infer the agent (Bates, MacWhinney, Caselli, Devescovi, Francesco, & Venza, 1984; Corrigan & Odya-Weis, 1985; Koff, Kramer, & Fowles, 1980; Thal & Flores, 2001) and inanimate Ns to infer the patient (Corrigan & Odya-Weis) in subject-verb-object (SVO) sentences.

Evidence also exists suggesting that young toddlers (19-23 months) have some awareness of the predicate-argument structure of the language. They appear to be able to use the number of Ns in a sentence as a cue to the predicate-argument structure of an SVO sentence and whether the sentence involves one or two semantic roles, i.e., agent, patient (Yuan & Fisher, 2009; Yuan, Fisher, & Snedeker, 2012). For example, toddlers have the ability to use the presence of 1 N argument (*The girl is dacking*) or 2 N arguments (*The girl is dacking the boy*) to infer whether a novel verb (V) is intransitive (requiring 1 argument) or transitive (requiring 2 arguments).

Young toddlers also seem to have sensitivity to the V selection restriction rules of the language, recognizing when these rules have been violated (e.g., Friedrich & Friederici, 2005; Pereyra, Klarman, Lin, & Kuhl, 2005). Verb selection restrictions are constraints on Vs that determine what semantically appropriate N arguments a V can take (e.g., Altmann & Kamide, 1999; Ferretti, McRae, & Hatherell, 2001). For example, in the sentence "*The girl is eating the cookie*" the V *eating* specifies that the subject/agent N must be animate and the object/patient N be an edible. In the case of a violation of a V selection rule, one or more of the Ns is semantically inappropriate for the sentence (e.g., *The girl is riding the cookie*). Neural (ERP) evidence shows that even 19 and 24 month olds recognize when a semantically inappropriate N has been used with a V (Friedrich & Friederici, 2005; Pereyra et al., 2005).

Such results simultaneously imply that toddlers are also sensitive to the violation of the natural affordance existing between two Ns in a simple sentence. In most sentences a natural affordance between the Ns and their associated semantic roles is expressed. Affordance refers to the ways in which people can interact with objects in the world, with the interaction reflecting intrinsic constraints that occur between the entities (e.g., Gibson, 1979; Glenberg, Becker, Klotzer, Kolanko, Muller, et al., 2009). In "*The girl is eating the cookie*" the affordance between the two Ns (*the girl*, *the cookie*) is a natural one as encoded/ expressed through the V *eating*. Sentences involving natural affordance also typically express highly probable events, with such event knowledge being used by both adults (Gertner, Fisher, & Eisengart, 2006; Kaschak & Glenberg, 2000; Matsuki, Chow, Hare, Elman, Scheepers, et al., 2012; Metusalem, Kutas, Urbach, Hare, McRae, et al., 2012) and children (Chapman & Kohn, 1978; Friedrich & Friederici, 2005; Pereyra et al., 2005; Strohner & Nelson, 1974) to facilitate sentence comprehension.

Word order relates to the regularity (statistical properties) of the language. Because English is a word order language, it is an especially important cue. Different structures such as SVO and subject relative (SR) clause sentences and passives and object relative (OR) clause sentences vary with respect to their word order as well as overall frequency of occurrence in the language (Roland, Dick, & Elman, 2007). SVO and SR forms are canonical NVN structures. They have overall higher frequency of occurrence than passives and ORs. Though SRs have overall lower frequency than SVOs they are no more difficult to understand than SVOs because both structures are NVN word order and express an SVO grammatical relation. Passive and OR structures are more difficult to understand than SVOs and SRs, with ORs posing the greatest challenge for listeners (MacDonald & Christiansen, 2002; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). ORs not only have a lower frequency of occurrence but also contain a lower frequency non-canonical word order (NNV) that expresses an object-subject-verb grammatical relation. Passives also are of lower frequency. They are easier to understand than ORs, but more difficult than SVOs and SRs because they express a non-canonical object-verb-subject grammatical relation.

For English-speaking children, mastery of word order cues has a protracted developmental trajectory, but children appear to show some awareness of word order as early as 16 months based on data from the preferential looking paradigm (Hirsh-Pasek & Golinkoff, 1996). By 2.5 years, children use word order but in combination with animacy to interpret SVOs, i.e., interpret animate N1 as agent (e.g., Bates et al., 1984; Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Thal & Flores, 2001). By age 4, children appear to begin to rely more on word order and less on noun animacy (e.g., Bates et al.). Using word order to interpret other word order sentences (object-subject-verb) begins around age 5, with children interpreting N2 as the subject/agent. However, the most significant developmental changes in sentence interpretation occur for sentences that are infrequent such as reversible passives and reversible ORs. This shift takes place around 7-12 years of age (e. g., Dick, Wulfeck, Krupa-Kwiatkowski, & Bates, 2004; Slobin, 1966; Von Berger, Wulfeck, Bates, & Fink, 1996).

Dick et al. (2004) examined the developmental trajectory of the sensitivity to word order (as well as other cues) in children 5-17 years of age. All of the sentences were reversible and included two animate nouns. These researchers used a “whodunit?” agent selection task to examine children’s understanding of SVO and SR canonical structures as well as passive and OR non-canonical structures. In their task, children listened to a sentence and then saw two images, one corresponding to the agent and the other to the patient. Children were asked to select the agent as quickly as possible. Results showed that children were sensitive to word order and sensitivity improved with age. The developmental trajectory to interpret different sentences also mirrored the frequency of occurrence of their word order: SVO = SR > Passive > OR (Roland et al., 2007).

The 5 and 6 year olds interpreted SVO and SR structures with 90% accuracy, and interpretation was at ceiling by age 9-10 years. Lagging just behind were passives, with performance approaching asymptote by about 11-12 years. OR sentences were the most difficult, with little improvement occurring between 5-8 years. The greatest improvements occurred between 9-10 years and 11-12 years, but adult-like levels were not attained until 15-17 years of age. Even though passives and ORs are both lower frequency structures, passives proved to be easier than ORs. This interpretation advantage may owe to the fact that their passives included the substructure cues of a verb participal form (e.g., *is chased*) and a preposition by-phrase (e.g., *by the dog*), which did not appear in the ORs.

The pattern of children's sentence interpretation speed mirrored that of their interpretation accuracy (SVO = SR < Passive < OR), but the developmental trajectory of interpretation speed was much shallower than that of accuracy. SVO and SR speeds were fairly stable between the ages of 5-10 years. Interpretation speed for passives remained about 250 ms slower than that of canonicals between 5-8 years. However, speed of passive interpretation became comparable to that of canonical sentences by age 9. By 11-12 years, children's speed for passives was comparable to that of 15-17 year olds. For 5-8 year olds, speed of OR interpretation remained about a second slower than that of canonical sentences. Not until 9-12 years did speed of ORs become comparable to that of canonicals. Also between the ages of 9-12 years, interpretation speed of ORs began to approximate that of 15-17 year olds.

As part of their study, Dick et al. (2004) also examined the sentence interpretation abilities of 5-18 year olds with SLI (and children with unilateral focal lesions). Children with SLI are those who are developing in a typical manner except for language (Leonard, 2014). Sentence comprehension difficulties represent a hallmark deficit of SLI (e.g., Friedman & Novogrodsky, 2007; Leonard, Deevy, Fey, & Bredin-Oja, 2013; Montgomery & Evans, 2009; Montgomery, Gillam, & Evans, 2009; Robertson & Joanisse, 2010; van der Lely, 2005). Dick et al. showed that the SLI group, relative to age controls, yielded a similar pattern of accuracy (SVO = SR > Passive > OR) and speed (SVO = SR < Passive < OR). However, the SLI group showed poorer accuracy across all sentence types, especially for passives and ORs. The groups did not differ in interpretation speed. Finally, it is interesting to note that children with SLI up through 8.5 years have been shown to rely exclusively on N animacy, even when word order cues are available, when selecting the agent in NVN (*the cat pets the cow*), NNV (*the camel the horse chases*), and VNN (*pets the pig the lamb*) sentences in which both animate and inanimate Ns appear (Evans, 2002; Evans & MacWhinney, 1999).

Two theoretical accounts have been proposed to explain the sentence interpretation deficits of SLI, a syntax-specific deficit (Frideman & Novgrodsky, 2007; van der Lely, 2005) and limitations in general cognitive processing abilities (Bishop, 2006; Leonard et al., 2013; Montgomery & Evans, 2009; Montgomery et al., 2009; Robertson & Joanisse, 2010). As to which account is the better explanation of these children's deficits has yet to be resolved. In **Method**, we provide a few comments about the overarching aims of our research project, and how the present study and future SLI studies fit in.

In summary, the accuracy and speed of children's sentence interpretation develop over time. It is not until adolescence that children begin to demonstrate adult-like interpretation of more complex structures such as ORs. For children whose first language is English, around age 4, they begin to rely more on word order than N animacy to guide sentence interpretation. However, the mastery of word order does not appear until adolescence.

Aims of the Present Study

The present study had two main aims. The first was to investigate the developmental changes in English-speaking children's syntactically-driven sentence interpretation abilities and sensitivity to word order. Typically developing children listened to two reversible structures with canonical word order (SVO, SR) and two with non-canonical word order (Passive, OR). All of the sentences were semantically implausible, as N animacy, N affordance, and thus probably event cues were removed. Asking children to interpret implausible sentences allowed us to illuminate their ability to use syntactic knowledge and word order. We created a novel "whatdunit?" agent selection task that was adapted from the conventional "whodunit?" task (Dick et al., 2004; Von Berger et al., 1996). Children were asked to identify the agent as quickly as possible in each sentence. We compared younger and older children's accuracy and speed of interpretation as well as error patterns. A second aim of the study was to determine the psychometric soundness (validity and reliability) of the task. If the task reveals good psychometric standing, it will prove suitable to use with children with SLI.

Method

Participants

Participants were typically developing (TD) children who were part of a larger on-going, multi-site project investigating the relation between cognitive processing and syntactically-driven sentence interpretation of children with and without SLI. The present study focusing on TD children is the first of a series of reports from this project. Subsequent studies will compare propensity matched SLI and TD groups 1) on our "whatdunit?" task to examine the similarities/differences in sentence interpretation abilities and 2) with respect to which cognitive abilities contribute to syntactically-based sentence interpretation.

For this study, participants were 256 children 7 through 11 years of age (males = 126, females = 130) with normal developmental history and language development. This age band was chosen because the main aim of our research program is to better understand the nature of sentence interpretation in children with SLI. These children, relative to age mates, show marked deficits interpreting each of the structures used in the present study.

Children were recruited from four regions of the U.S.-- Athens area of Ohio, Logan, Utah, San Diego, California, and Dallas, Texas. Children were recruited through various school systems, community centers, and university-sponsored summer camps for children. English was the primary language spoken by all the children. All the children had normal medical, developmental and language history, and no neurological impairment or psychological/emotional disturbance, based on parent report. Standardized language and cognitive assessments as well as parent reports were completed at time of participation. All children had: (a) normal-range IQ as measured by the *Leiter International Performance Scale-Revised* (Roid & Miller, 1997); (b) normal-range hearing sensitivity (*American National Standards Institute*, 1997); (c) normal or corrected vision; and (d) normal language as measured by the *Clinical Evaluation of Language Fundamentals-4* (CELF-4; Semel, Wiig, & Secord, 2003) linguistic concepts and following directions subtest and recalling sentences subtest, the *Comprehensive Receptive Expressive Vocabulary Test* (CREVT-2, Wallace & Hammill, 2000), the *Comprehensive Assessment of Spoken Language* test

(CASL, Carrow-Woolfolk, 1999) and the *Test of Narrative Language* (TNL, Gillam & Pearson, 2004). Each of the standardized language tests has good internal reliability (.84 - .95) and internal validity (.60 - .82). Table 1 displays summary cognitive and language data for the participants.

Based on parent report, 81.7% of children were Caucasian, 5.8% African American, 9.4% Hispanic, 8% Asian, and 2.8% American Indian/native Hawaiian. Almost 76% of children came from households whose mothers attained a 2-year degree or lower and 24% from a household whose mothers attained a college degree or higher. Almost 19% of the children came from a low-income family (< $30k/year) whereas 81% came from a high-income family (≥ $30k/year) Demographic data appear in Table 2.

To examine developmental changes in children's sentence interpretation abilities, participants were divided into two age groups. The younger group (N = 132) had a mean age of 8:1 years (7:0-9:3). The older group (N = 124) had a mean age of 10:8 years (9:4-11:11). These age groups were motivated on two grounds: 1) findings from Dick et al. (2004) indicate developmental improvement in the interpretation of sentences with non-canonical word order sometime during the 9th year of life and 2) comparing two relatively large-N groups will yield robust and stable results.

Place Tables 1 & 2 about here

Sentence Interpretation Task

Procedure. Children's sentence interpretation was assessed using a novel "whatdunit?" agent selection task. Children were told that they would hear a man saying some funny sounding sentences about one thing/object doing an action on another thing/object. They were told that after each sentence three pictures would appear at the bottom of the touch screen and to touch the "thing that did the action" as quickly as they could.

Stimuli. Sentences consisted of 132 items (33 SV0, 33 SR, 33 *be* Passive, 33 OR). All of the sentences were reversible and of the same length (12 words). The SRs and ORs were center-embedded relative clause structures. Each sentence included a prepositional phrase (PP) following the second noun phrase (NP) in which a noun appeared. Including a PP allowed us to control sentence length across all sentence types without altering their fundamental syntactic forms. The 132 items were arranged into three blocks of 44 items each. One block was presented during each of the three testing sessions (see below). Appendix A provides sample experimental sentences.

The sentences were created using a pool of 33 nouns (Ns), 22 verbs (Vs), and 3 prepositions (see Appendix B for a list of these items). Sentences were constructed to be semantically implausible to ensure the children would rely only or primarily on syntactic information and word order to interpret them. Semantic implausibility was created in two related ways. First, N animacy was removed by selecting inanimate/ object Ns as the agent (and patient) of the sentences. Removing animacy was critical because children with SLI as old as 8.5 years primarily rely on N animacy (Evans, 2002). Second, we violated "typical" predicate-argument structure of the sentences (i.e., V selection restriction rules) and simultaneously the natural affordance between the Ns in the sentences. Together, these manipulations rendered all sentences semantically implausible and expressing highly improbable events (e.g., *The chair that the bread had splashed under the square was new*). Thus, children were offered no reliable semantic/ real-world cues to which N was the agent. Finally, our interest was on semantic implausibility broadly, not determining individual and/or interactive effects of N inanimacy and weak N affordance on children's sentence interpretation.

The sentences were also controlled in other ways. In both the SRs and ORs the relative pronoun *that* always appeared. Also, verb tense was the same in all sentence types. Because verb tense in the passives was past (*was V-ed*), (*The belt was pulled by the book near the very new bowl*), the other sentence types were constructed in past tense (SVO: *The hat had hugged the belt behind the very bright new sock*; SR: *The cake that had cleaned the bed near the train was bright*; OR: *The truck that the clock had pulled near the door was bright*).

The properties of the Ns, Vs, and prepositions were also carefully controlled to minimize any influence that lexical knowledge may exert on children's sentence interpretation (see Borovsky et al., 2012). This control was especially important for our work with children with SLI because of the documented lexical deficits of these children (e.g., Mainela-Arnold, Evans, & Coady, 2008; 2010;

McGregor, Newman, Reilly, & Capone, 2002). All of the main lexical items had spoken word frequency ratings of 6:0 or younger (Moe, Hopkins, & Rush, 1982) and age of acquisition (AoA) ratings of 3.6 years or younger. The nouns also had high imageability (> 500), concreteness (> 500), and familiarity (> 500) ratings (Coltheart, 1981; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012; Storkel & Hoover, 2010; Vitevitch & Luce, 2004). Each of the Ns appeared with equal frequency as NP1, NP2, and NP3 across the sentences, and the Vs and prepositions occurred with equal frequency. Finally, the images corresponding to the Ns were color drawings of simple objects (e.g., bed, coat, spoon, truck) standardized for name and image agreement, familiarity, and visual complexity (Rossion & Pourtois, 2004).

The sentences were recorded at a normal speaking rate (~ 4.4 syllables/s; Ellis Weismer & Hesketh, 1993) and with normal prosodic variation by an adult male speaker of Midwestern American English. All audio files were digitized (44 kHz), low pass filtered (20 kHz), and normalized for intensity.

Baseline Motor Speed Task

Stimuli and Procedure. Children completed a simple motor speed task yielding an index of each child's motor planning and execution speed (e.g., Dick et al., 2004; Montgomery & Leonard, 2006). The structure of this task was identical to the "whatdunit?" task. Displayed along the bottom of the touch screen were 3 empty boxes arranged horizontally. Children were told that they would first hear a tone (2kHz, 500ms) and then see a cross appear in one of the boxes. They were instructed to touch the cross as quickly as possible as soon as it appeared. The task comprised 30 trials. The tone and cross were separated by an inter-stimulus interval varying between 500ms and 1.5s. The cross appeared in each box randomly across the trials and an equal number of times.

General Procedures

Each child was seen individually in a quiet laboratory. The children participated in three testing sessions, each lasting about two hours including rest breaks. To record the accuracy and speed of the children's responses as well as ensure a random order of presentation of the trials and position of the correct answer, presentation of the stimuli was controlled using E-Prime software (Schneider, Eschmann, & Zuccolotto, 2002) running on a laboratory laptop connected to a 17" Elo Touch Screen monitor.

Children sat at a table in front of a touch screen. To maintain consistency in interpreting reaction time data, children placed their arm in a comfortable position on the table so that the fingers of their dominant hand rested on a red dot located in the center of the bottom edge of the monitor, just below the touch screen. Children were instructed to leave their fingers on the dot until they were ready to touch the screen. Both response accuracy and speed were emphasized in both speeded tasks. Prior to experimental trials in each task, children saw demonstration items and completed practice trials (semantically plausible and implausible sentences in the "whodunit?" task) to ensure they understood the tasks. The children were able to complete the practice trials and no child was excluded due to difficulty understanding the tasks. Stimuli were presented binaurally under noise reduction headphones at a comfortable listening level determined by the child. The simple motor speed task always immediately preceded the agent selection task.

Data Preparation

The children's speeded responses on the "whatdunit?" and simple motor speed tasks were not smoothed, i.e., outliers were not removed. This decision was motivated on two grounds. First, we wished to maximize the robustness of the analyses by including the full distribution of the data (Whelan, 2008). Second, because our ultimate aim is to describe and compare the sentence interpretation abilities of children with and without SLI, the analyses of the sentence interpretation (and simple motor) task were conducted in the manner future analyses will be performed. Finally, the accuracy score data on the "whatdunit?" task were not transformed because transformation of percent accuracy scores often makes data interpretation very difficult (Warner, 2012).

Results

Sentence Interpretation

Accuracy. To assess children's sentence interpretation accuracy by age, a repeated measures ANOVA (Age x Sentence Type) was conducted, with the between-group factor being age (younger, older) and the within-group factor being sentence type (SVO, SR, Passive, OR). The ANOVA revealed significant effects for age $F(1, 254) = 33.52$, $p = .001$, $\eta^2 = .05$ and sentence type $F(3, 762) = 264.87$, $p =$

.001, $\eta^2 = .38$, and a significant interaction $F(3, 762) = 6.45$, $p = .0003$, $\eta^2 = .006$. Figure 1 presents comprehension accuracy by sentence type and age.

Place Figure 1 about here

Planned between-age group comparisons (with Bonferroni correction) indicated that the older children outperformed the younger children on each sentence type: SVO $t(245) = 4.43$, $p = .0004$; SR $t(248) = 3.74$, $p = .0008$; Passive $t(254) = 4.26$, $p = .0004$; OR $t(254) = 5.62$, $p = .0004$. Planned within-age group comparisons yielded slightly different performance patterns for each group. For the older children, SVOs and SRs were interpreted with comparable accuracy. Passives and ORs were likewise interpreted with comparable accuracy, and both were interpreted more poorly than the SVOs and SRs. The younger children also showed comparable interpretation of SVOs and SRs. Passive interpretation was likewise poorer relative to SVOs and SRs. But, unlike the older children, interpretation of ORs was poorer than passives.

Inspection of individual participant's data indicated that 6 children achieved a score of zero on the passives (2 younger, 4 older) and 8 children achieved a score of zero on the ORs (4 younger, 4 older). Finally, although the overall accuracy for passives and ORs was considerably poorer than that of SVOs and SRs, these lower scores were not attributed to random performance (see **Error Patterns** in **Discussion** for details). Table 3 displays summary data for interpretation accuracy by age and sentence type.

Place Table 3 about here

Speed. Because the older children's mean simple motor RT ($RT_M = 551$ms) was significantly faster than the younger children's ($RT_M = 653$ms), $t(229) = 9.36$, $p = .001$, a subtraction method was used to eliminate the motor planning and execution component of the children's speeded interpretation responses. The procedure entailed subtracting each child's mean motor RT from each sentence processing time for correctly interpreted items, thus yielding an adjusted interpretation speed on a trial-by-trial basis. Recall that 6 children achieved a score of zero on the passives and 8 on the ORs. As a result, fewer children were included in the passive and OR analyses than the SVO and SR analyses. All those children performing above zero were included in the analyses.

A repeated-measures ANOVA (Age x Sentence Type) was conducted, with the between-group factor being age and the within-group factor being sentence type. Results revealed significant effects of age $F(1, 244) = 25.68$, $p = .0001$, $\eta^2 = .07$ and sentence type $F(3, 732) = 34.52$, $p = .0001$, $\eta^2 = .03$ as well as a significant age x sentence type interaction $F(3, 732) = 2.96$, $p = .03$, $\eta^2 = .003$. Figure 2 presents interpretation speed by age and sentence type.

Place Figure 2 about here

Planned between-age group comparisons (with Bonferroni correction) revealed that the older children were faster than the younger children to interpret each sentence type: SVO $t(240) = 4.10$, $p = .0004$; SR $t(216) = 3.80$, $p = .0008$; Passive $t(222) = 6.01$, $p = .0004$; OR $t(225) = 4.27$, $p = .0004$. Similar to the accuracy results, planned within-age group comparisons yielded slightly different patterns between the older and younger children. The older children's interpretation speed was comparable across the SR, passive, and OR sentences, and all were slower than SVO sentence interpretation. For the younger children, speed was comparable for the SVOs and SRs and slower than their interpretation of both the passives and ORs, which did not differ from each other. Table 4 presents interpretation speed by age and sentence type summary data.

Place Table 4 about here

Individual Variability. Though not a focus of the present study, we were still interested in examining the variability of sentence interpretation accuracy across participant ages given the semantic implausibility of our sentences. We might expect the children would show less variability for the higher-frequency, earlier-acquired SVOs and SRs than the lower-frequency, later-acquired non-canonical passives and ORs. Consistent with these expectations, it can be seen that the variability in the interpretation of SVOs (Figure 3) and SRs (Figure 4) was relatively compact whereas the variability in passive and OR interpretation (Figures 5 and 6, respectively) was marked by considerable variability.

Place Figures 3-6 about here

Error Patterns. To examine the nature of the errors children made across sentence types, an Age (younger, older) x Error Type (object NP error, PPN error) ANOVA was computed, followed by planned between-age group and within-age group comparisons. Object noun errors reflected children's selection of the object noun. Prepositional Phrase Noun (PPN) errors reflected their selection of the noun appearing in the prepositional phrase near the end of the sentence. Numbers of errors were used in the analyses.

Results indicated significant effects of age $F(1, 254) = 33.40$, $p = .001$, $\eta^2 = .02$ and error type $F(3, 762) = 265.67$, $p = .001$, $\eta^2 = .13$, and a significant age x error type interaction $F(3, 762) = 6.37$, $p = .0003$, $\eta^2 = .003$. Between-age group comparisons (with Bonferroni correction) revealed that the younger children made a greater number of object errors than older children in each sentence type: SVOs $t(254) = 3.82$, $p = .0008$; SRs $t(254) = 3.16$, $p = .007$; Passives $t(254) = 3.07$, $p = .009$; and ORs $t(254) = 4.27$, $p = .0004$. Though the number of PPN errors was very low for each group, the younger children also made more of these errors than older children in each sentence type: SVOs $t(254) = 3.85$, $p = .0008$; SRs $t(235) = 3.42$, $p = .002$; Passives $t(198) = 4.27$, $p = .0004$; and ORs $t(227) = 4.01$, $p = .0004$.

Within-age group comparisons (with Bonferroni correction) indicated that the groups produced slightly different error patterns. The following pattern was produced by the younger children: a greater number of object errors than PPN errors in SVOs $t(245) = 2.83$, $p = .02$; Passives $t(172) = 12.50$, $p = .0004$; and ORs $t(184) = 16.09$, $p = .0004$; but comparable object and PPN errors in SRs $t(262) = 2.13$, $p = .14$. The older children produced more object errors than PPN errors across all sentence types: SVOs $t(220) = 3.95$, $p = .004$; SRs $t(236) = 2.71$, $p = .02$; Passives $t(133) = 9.98$, $p = .0004$; and ORs $t(141) = 11.31$, $p = .0004$. Table 5 presents error type by sentence type for each age group.

Place Table 5 about here

Psychometric Properties of the "Whatdunit?" Task

Because the current task differed in important ways from the conventional "whodunit?" task, it was important to establish its basic psychometric soundness. To this end, we calculated both validity and reliability.

Internal construct validity. Internal construct validity was assessed by computing correlations on accuracy scores for all children across the four different sentence types. This procedure allowed us to determine whether we had constructed 1) a strong overall task and 2) two highly correlated canonical structures (SVO, SR) and two highly correlated non-canonical structures (Passive, OR). Table 6 reveals that the correlation between the two canonical structures was very high (.84, p = .0001) as was the correlation between the two non-canonical structures (.89, p = .0001). It should be noted that the correlations between the canonical and non-canonical structures were markedly lower than the correlations within each sentence type, supporting the construct of canonical/non-canonical differences.

Place Table 6 about here

Concurrent validity. Concurrent validity was examined by computing the correlation between accuracy scores on the agent selection task as a whole and composite scores for three language domains: 1) lexical (CREVT-R, CREVT-E; CASL antonyms); 2) sentential (CELF-4 concepts and following directions, CELF-4 recalling directions); and 3) overall language (combining the above scores with the total score for the TNL). Composite scores were created by converting each child's raw score on each language measure to a z-score and deriving an average z-score for each respective domain. The "whatdunit?" task was significantly correlated with each of the composite language scores. Concurrent validity data appear in Table 7.

Place Table 7 about here

Task Reliability. Internal consistency reliability was assessed on a trial-by-trial basis across all 256 children by computing Cronbach's alpha for each sentence type separately and on the task as a whole. Very strong reliability was attained for each sentence type. Overall task reliability was likewise very strong. Reliability data are displayed in Table 8.

Place Table 8 about here

Discussion

This study had two main aims. One was to investigate the sentence interpretation abilities of TD children 7 through 11 years of age, with a special focus on their syntactically-based interpretation abilities and sensitivity to word order. Children listened to two broad types of semantically implausible sentences, those with canonical word order and those with non-canonical word order. Children were asked to identify as quickly as possible the inanimate agent of each sentence. The second aim was to determine the psychometric soundness of the task, with an eye toward its suitability to use with children with SLI.

Aim 1: Examining Children's Sentence Interpretation Performance

Interpretation Accuracy. The overall sentence interpretation pattern demonstrated by the children in the present study was similar to that reported in the developmental literature (e.g., Dick et al., 2004; Von Berger et al., 1996). Our children showed significantly better interpretation of the sentences with canonical word order than those with non-canonical word order. Though the sentence type effect was small ($\eta^2 = .38$; Cohen, 1988), we might argue that the practical relevance of the effect is significant. That we obtained a robust sentence type effect suggests that our semantic manipulation was successful, with the different word orders reliably influencing the children's performance. At a more fine-grained level, the children also showed an overall similar pattern to previous research: SVO = SR > Passive > OR. These findings, importantly, suggest that in the absence of semantic plausibility children 7-11 years of age use word order cues to derive an appropriate interpretation of different sentence structures, i.e., determine who did what whom.

Also in keeping with previous research, the present study revealed an age effect. Though the effect was small ($\eta^2 = .05$), the older children were shown to robustly outperform the younger children. The older children (9:4-11:11 years) showed significantly better interpretation across all four sentence structures than the younger children (7:0-9:3 years). Though the younger children performed with good accuracy on the SVOs (80%) and SRs (77%), the older children outperformed them on both (89%, 85%). Likewise, the older children outperformed the younger children on the sentences with non-canonical word order. An especially striking finding is that relative to the canonical forms, the younger children's performance on the non-canonical forms dropped precipitously (Passives = 50%, ORs = 39%). Even the older children showed a substantial drop (66%, 58%).

The age groups produced slightly different accuracy patterns. Within each age group, the SVOs and SRs were interpreted with comparable accuracy, a finding that is consistent with the developmental literature. However, the groups showed a different pattern of interpretation of the sentences with non-canonical word order. The younger children showed an interpretation advantage of the passives over the ORs, a pattern that mirrors the literature. However, the older children showed comparable interpretation of the passives and ORs. These different response patterns suggest that older children have comparable sensitivity to the word orders of passive and OR sentences whereas younger children are differentially sensitive to the word orders of these forms.

The present findings depart from the developmental literature in ways other than the divergences noted above. The two most striking differences relate to performance levels and variability. With respect to absolute performance levels, informal comparison of the scores produced by our children with those of the children in Dick et al. (2004) reveal that our children uniformly performed worse across all sentence forms, with performance on the sentences with non-canonical word order being substantially poorer. For example, our older children (Age_M = 10:8) performed considerably worse on the passives (66%) than the 5-6 year olds (90%) in Dick et al. The contrast in scores for the ORs is even more dramatic. Our older children achieved just 58% accuracy whereas the 9-12 year olds in Dick et al. achieved about 90% accuracy. Interestingly, even for the canonical forms, the performance of our older children was lower than that of the children in Dick et al. The older children in the present study performed with 89% accuracy and 85% accuracy on the SVOs and SRs, respectively. In Dick et al., 9-12 year olds performed at ceiling on both of these structures.

Performance variability is a second notable difference. The children in the present study performed with far greater variability across all sentence structures, especially those with non-canonical word order, relative to the children in Dick et al. (2004), as revealed by a comparison of the error bars in each study. But the most powerful demonstration of our children's variability comes from an inspection of the scatterplots (Figures 3-6). The variability of the children's interpretation of the SVO and SR sentences was relatively compact, with the majority of the children performing above 80% accuracy. By contrast, the children's interpretation of the passives and ORs was characterized by considerable variability. Together, the lower performance levels and greater variability of our children appear to be attributable to the fact that our sentences expressed very weak semantic plausibility, leaving word order as the primary cue available to guide the children's sentence interpretation.

The poorer performance on our sentences, however, may also have been influenced by factors other than the absence of semantic plausibility, making them more difficult than the sentences used by some researchers. For example, our sentences were a bit longer than those used in some studies (Dick et al., 2004; Kidd & Bavin, 2002), but not all (Booth, MacWhinney, & Harasaki, 2000; Love, 2007; Roberts, Marinis, Felser, & Clahsen, 2007). Also, our sentences may have been more complex than those used in some previous studies. First, the SR and OR sentences included center-embedded relative clauses that were embedded within the sentential subject (*The kite that the dress had pressed near the book was hot*) rather than in cleft constructions (*It's the cat that the dog is biting*), which are easier to process (e.g., Dick et al.). Second, our sentences included a PP near the end of the sentence. Its inclusion may have resulted in an increase in processing difficulty of the sentences because of uncertainty as to whether the PP should attach to the VP or the object NP.

Interpretation Speed. The general pattern of sentence interpretation speed was similar to that reported in the developmental literature (Dick et al., 2004). Children's speed of interpretation of sentences with canonical word order was significantly faster than their speed of sentences with non-canonical word order, i.e., SVO/SR faster than passive/OR. Like the accuracy findings, the sentence type effect was small ($\eta^2 = .03$). But, again, we might argue that our results have practical relevance. Our findings are clearly consistent with the developmental literature despite the fact that our sentences were absent semantic plausibility.

In addition, the present study produced an age effect, though small ($\eta^2 = .07$), with the older children yielding robustly faster sentence interpretation than the younger children across all sentence types. However, the older and younger children showed different patterns of interpretation speed. The older children showed no difference in speed of interpretation across SR, passive, and OR sentences, all of which were slower than SVO interpretation. The younger children, by contrast, yielded SVO and SR interpretation speeds that were comparable and faster than those for passives and ORs, which did not differ. The younger children's findings align well with those of Dick et al. (2004), suggesting that younger children's interpretation speeds are differentially sensitive to sentences containing canonical word order vs. sentences involving non-canonical word orders. Sentence interpretation speed of the older children, however, appears essentially insensitive to sentences varying in word order. For these children, accuracy appears to be the more sensitive index of their appreciation of word order. For younger children, both accuracy and speed are sensitive indices.

Error Patterns. Children's sentence interpretation error patterns were examined to determine whether they would make more object NP errors than PPN errors (noun appearing in PP near the end of the sentence) for those sentences they misinterpreted. We might anticipate this pattern given that object errors are linguistically motivated whereas PPN errors are not. Consistent with this expectation, the children overwhelmingly yielded more object NP errors than PPN errors. Relative to the older children, the younger children produced significantly more object NP errors across all four sentence types. The younger children also made more PPN errors than the older children in each sentence type; however, the number of these errors was very small for both groups. More important, the younger and older children showed essentially the same error patterns. Both groups produced significantly more object errors than PPN errors across all sentences, regardless of their word order. The only exception to this pattern was the younger children who produced comparable numbers of object and PPN errors in SR sentences. These results, importantly, indicate that the children's errors were not random, even for those achieving very low accuracy. Thus, overall, when children misinterpret semantically implausible sentences containing either canonical or non-canonical word order, their errors are overwhelmingly linguistically motivated. They misinterpret the object NP as the agent of the sentence.

Aim 2: Psychometric Soundness of the "Whodunit?" Task

Our novel task proved to be a valid and reliable measure of children's syntactically-based sentence interpretation abilities. With respect to validity, the task demonstrated very good internal construct validity, with much higher correlations within each sentence type (canonical word order: SVO-SR; non-canonical word order: Passive-OR) than across the sentence types containing canonical vs. non-canonical word orders. Task accuracy was significantly correlated with other language measures at moderately-high levels demonstrating strong concurrent construct validity. The task also proved to have very good internal reliability.

Suitability of the "Whatdunit?" Task to Study Sentence Interpretation in Children with SLI

As noted earlier, a hallmark characteristic of children with SLI is their sentence interpretation deficits. Based on the syntax-specific deficit account (Friedman & Novogrodsky, 2007; van der Lely, 2005), these children's difficulties understanding sentences containing non-canonical word order (passive, OR) are due to difficulties constructing hierarchical grammatical relations. By contrast, the more general cognitive processing limitations account assumes that weaknesses in memory and attention relate to these children's broader difficulties understanding sentences involving non-canonical word order and as well as those with canonical word order (Leonard et al., 2013; Montgomery & Evans, 2009; Montgomery et al., 2009; Robertson & Joanisse, 2010). As to which of these accounts better explains the nature of these children's sentence interpretation deficits is still under debate.

The present “whatdunit?” task proved to be a sensitive and robust measure of TD children’s developmental changes in sentence interpretation abilities. It also proved to be psychometrically sound. Thus, it appears to be suitable to use with children with SLI. Because the task illuminates children’s syntactically-driven sentence interpretation abilities, it will offer us the unique and important opportunity to empirically evaluate and compare the merits of the two different theoretical accounts of SLI sentence interpretation deficits. Our future studies will focus on comparing the similarities/differences in sentence interpretation abilities and which cognitive abilities best account for sentence interpretation in well-matched groups of children with SLI and TD children.

Conclusions

The main goals of this study were to 1) examine the developmental changes in TD children’s syntactically-based sentence interpretation abilities and sensitivity to word order and 2) determine the psychometric soundness of the task developed for this study, with an eye toward using it with children with SLI. Our “whatdunit?” sentence interpretation task differed from the conventional “whodunit?” task mainly in that in our task children listened to semantically implausible sentences and were asked to identify an inanimate agent in each sentence. Listening to implausible sentences invited children to use their syntactic and word order knowledge to interpret the sentences. This manipulation makes our task unique relative to those used by other developmental researchers.

The first key finding was that older children, relative to younger children, were both more accurate and faster to interpret all sentence types. These findings indicate that the older children had greater sensitivity to the word order cues of the language than younger children. A second key finding was that even the older children showed performance on the sentences involving canonical word order that was not yet at ceiling and performance that was far from ceiling on the non-canonical forms. The younger children’s performance across all sentence forms fell well below ceiling. Third, the “whatdunit?” task showed strong basic psychometric characteristics. Finally, the task appears to be well suited to study the sentence interpretation abilities of children with SLI.

The present findings extend the developmental language literature in at least two important ways. First, interpretation of semantically implausible sentences involving canonical or non-canonical word order improves with age along with reliance on syntactic knowledge and word order improving with age. Second, children's syntactically-based sentence interpretation, regardless of the word order of the sentence, is not yet mastered through age 11 years. The present findings together with previous developmental findings (Bates & MacWhinney, 1987, 1989; Bates et al., 1984; Dick et al., 2004; Von Berger et al., 1996) indicate that reliable and successful sentence interpretation derives from a combination of structural and semantic cues.

Appendix A

Sample Experimental Sentences

Subject-Verb-Object

The hat had hugged the belt behind the very bright new sock.

The ring had moved the square behind the very bright cold bed.

The square had changed the bed under the very new dry key.

The shoe had bumped the fork near the very bright new wheel.

The knife had watched the ball near the very bright hot square.

Subject Relative

The watch that had hugged the truck behind the kite was bright.

The train that had helped the knife under the square was cold.

The boot that had fixed the shoe behind the drum was new.

The cake that had cleaned the bed near the train was bright.

The spoon that had licked the book near the watch was bright.

Passive

The train was watched by the bed behind the very cold cake.

The watch was bumped by the wheel near the very bright clock.

The key was changed by the chair behind the very bright square.

The belt was pulled by the book near the very new bowl.

The clock was rubbed by the shirt behind the very new door.

Object Relative

The box that the kite had splashed behind the shoe was dry.

The truck that the clock had pressed near the door was bright.

The chair that the bread had splashed under the square was new.

The kite that the dress had pressed near the book was hot.

The watch that the sock had wiped near the shirt was dry.

Appendix B

Nouns, Verbs, and Prepositions used in the Experimental Sentences

Nouns		Verbs	Prepositions
Ball	Knife	Asked	Behind
Bed	Ring	Bathed	Near
Belt	Shirt	Bumped	Under
Boat	Shoe	Called	
Book	Sock	Cleaned	
Boot	Spoon	Changed	
Bowl	Square	Dressed	
Box	Train	Fixed	
Bread	Truck	Helped	
Broom	Watch	Hooked	
Cake	Wheel	Hugged	
Car		Kissed	
Chair		Licked	
Clock		Marked	
Door		Pressed	
Dress		Pulled	
Drum		Rubbed	
Fork		Splashed	
Glove		Touched	
Hat		Washed	
Key		Watched	
Kite		Wiped	

Acknowledgments

This research was supported by a grant (R01 DC010883) from the National Institute on Deafness and Other Communication Disorders. We express our gratitude to all the children and their parents who participated in this project. We also thank Beula Magimairaj and Naveen Nagaraj for their invaluable assistance during various phases of this study.

References

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 73*, 247-264.

American National Standards Institute (1997). Specifications of audiometers (ANSI/ANS-8.3-1997; R2003).

Bates, E. & MacWhinney, B. (1987). Competition variation and language learning. In B. MacWhinney (Ed.) *Mechanisms of language acquisition* (Hillsdale NJ: Lawrence Erlbaum).

Bates, E. & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney and E. Bates (Eds.) *The Cross-Linguistic Study of Sentence Processing.* Cambridge: Cambridge University Press.

Bates, E., MacWhinney, B., Caselli, C., Devescovi, A., Natale, F., & Venza, V. (1984). A crosslinguistic study of the development of sentence interpretation strategies. *Child Development*, *55*, 341–354.

Bishop, D. (2006). What causes specific language impairment in children? *Current Directions in Psychological Science, 15*, 217–221.

Booth, J., MacWhinney, B., & Harasaki, Y. (2000). Developmental differences in visual and auditory processing of complex sentences. *Child Development, 71*, 981-1003.

Borovsky, A., Elman, J., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology, 112*, 417-436.

Carrow-Woolfolk, E. (1999). *Comprehensive Assessment of Spoken Language*. Pro-Ed.

Chapman, R. & Kohn, L. (1978). Comprehension strategies in two- and three-year olds: Animate agents or probable events? *Journal of Speech and Hearing Research, 21*, 746-761.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33A*, 497-505.

Corrigan, R., & Odya-Weis, C. (1985). The comprehension of semantic relations by two-year-olds: An exploratory study. *Journal of Child Language*, *12*, 47-59.

Dick, F., Wulfeck, B., Krupa-Kwiatkowski, M., & Bates, L. (2004). The development of complex sentence interpretation in typically developing children compared with children with specific language impairment or early unilateral focal lesions. *Developmental Science, 7*, 360-377.

Elman, J. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Ellis Weismer, S., & Hesketh, L. (1993). The influence of prosodic and gestural cues on novel word acquisition by children with specific language impairment. *Journal of Speech and Hearing Research*, *54*, 177-190.

Evans, J. (2002) Variability in comprehension strategy use in children with specific language impairments: A dynamical systems account. *International Journal of Language and Communication Disorders, 37,* 95-116.

Evans, J. & MacWhinney, B. (1999). Sentence processing strategies in children with expressive and expressive-receptive specific language impairments. *International Journal of Language and Communication Disorders, 34,* 117-134.

Ferretti, R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, *44*, 516-547.

Friedman, N., & Novogrodsky, R. (2007). Is the movement in syntactic SLI related to traces or to thematic role transfer? *Brain and Language, 101*, 50-63.

Friedrich, M., & Friederici, A. (2005). Semantic sentence processing reflected in the event-related potentials of one- and two-year old children. *NeuroReport, 16*, 1801-1804.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science, 17*, 684-691.

Gibson, J. (1979). *The ecological approach to visual perception*. New York: Houghton Mifflin.

Gillam, R., & N. Pearson, N. (2004). *Test of Narrative Language*. Pro-Ed.

Glenberg, A., Becker, R., Klotzer, S., Kolanko, L., Muller, S., & Rinck, M. (2009). Episodic affordances contribute to language comprehension. *Language and Cognition, 1*, 113-135.

Golinkoff, R., Hirsh-Pasek, K., Cauley, K. & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14*, 23-45.

Hirsh-Pasek, K., & Golinkoff, R. (1996). *The origins of grammar: evidence from early language comprehension*. Cambridge, MA: MIT Press.

Kaschak, M., & Glenberg, A. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language 43*, 508–529

Kidd, E., & Bavin, E. (2002). English-speaking children's comprehension of relative clauses: Evidence for general-cognitive and language-specific constraints on development. *Journal of Psycholinguistic Research, 31*, 599-617.

Koff, E., Kramer, P., & Fowles, B. (1980). Effects of event probability and animateness on children's comprehension of active and passive sentences. *The Journal of Psychology, 104,* 157-163.

Kuperman V, Stadthagen-Gonzalez H, & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, Instruments, & Computers, 44,* 978-90.

Leonard, L. (2014). *Children with Specific Language Impairment*. MIT Press. Cambridge, MA.

Leonard, L., Deevy, P., Fey, M., & Bredin-Oja, S. (2013). Sentence comprehension in specific language impairment: A task designed to distinguish between cognitive capacity and syntactic complexity. *Journal of Speech, Language, and Hearing Research, 56*, 577-589.

MacDonald, M., & Christiansen, M. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109,* 35–54.

Mainela-Arnold, E., Evans, J. L., & Coady, J. A. (2008). Lexical representations in children with SLI: Evidence from a frequency-manipulated gating task. *Journal of Speech, Language, and Hearing Research, 51*, 381–393.

Mainela-Arnold, E., Evans, J., & Coady, J. (2010). Explaining lexical-semantic deficits in Specific Language Impairment: The role of phonological similarity, phonological working memory, and lexical competition. *Journal of Speech, Language, and Hearing Research, 53,* 1742-1756.

Marslen-Wilson, W., & Tyler, L. (1980). The temporal structure of spoken language understanding. *Cognition*, *25*, 71–102.

Marslen-Wilson, W., & Welsh, A. (1978), Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29–63.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 576–585.

Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition, 37*, 913–934.

McGregor, K., Newman, R., Reilly, R., & Capone, N. (2002). Semantic representation and naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 45*, 998–1014.

Metusalem, R., Kutas, M., Urbach, T., Hare, M., McRae, K., & Elman, J. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, *66*, 545-567.

Moe, A., Hopkins, C., & Rush, R. (1982). *The vocabulary of first-grade children*. Springfield, IL: Thomas.

Montgomery, J., & Evans, J. (2009). Complex sentence comprehension and working memory in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52*, 269-288.

Montgomery, J., Evans, J., & Gillam, R. (2009). Relation of auditory attention and complex sentence comprehension in children with specific language impairment: A preliminary study. *Applied Psycholinguistics, 30*, 123-151.

Montgomery, J., & Leonard, L. (2006). Effects of acoustic manipulation on the real-time inflectional processing of children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 49*, 1238-1256.

Pereyra, J., Klarman, L., Lin, L., & Kuhl, P. (2005). Sentence processing in 30-month-old children: An event-related potential study. *NeuroReport, 16*, 645-648.

Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent priming at trace positions in children's sentence processing. *Journal of Psycholinguistics*, *36*, 175-188.

Robertson, E., & Joanisse, M. (2010). Spoken sentence comprehension in children with dyslexia and language impairment: The roles of syntax and working memory. *Applied Psycholinguistics*, *31*, 141-165.

Roid, G., & Miller, L. (1997). *Leiter International Perfor-mance Scale-Revised*. Wood Dale, IL:Stoelting.

Roland, D., Dick, F., & Elman, J. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language, 57*, 348–379.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object set: The role of surface detail in basic-level object recognition. *Perception, 33*, 217-236.

Schneider, W., Eschman, A., & Zuccolott, A. (2002). E-Prime user's guide. Pittsburgh: Psychology Software Tools Inc.

Semel, E., Wiig, E., & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals-4*. San Antonio, TX: The Psychological Corporation.

Slobin, D.I. (1966). Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior, 5*, 219–227.

Strohner, H., & Nelson, K. (1974). The young child's development of sentence comprehension: Influence of event probability, nonverbal context, syntactic form, and strategies. *Child Development, 45*, 567-576.

Storkel, H. & Hoover, J. (2010). On-line calculator of phonotactic probability and neighborhood density based on child corpora of spoken American English. *Behavior Research Methods, Instruments, & Computers, 42*, 497-506.

Thal, D, & Flores, M. (2001). Development of interpretation strategies in typically developing and late-talking toddlers. *Journal of Child Language, 28*, 173-193.

Traxler, M. & Tooley, K. (2007). Lexical mediation and context effects in sentence processing. *Brain Research*, *1146*, 59-72.

van der Lely, H. (2005). Domain-specific cognitive systems: Insight from Grammatical-SLI. *Trends in Cognitive Sciences, 9,* 53-59.

Vitevitch, M., & Luce, P. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers, 36,* 481-487.

Von Berger, E., Wulfeck, B., Bates, E., & Fink, N. (1996). Developmental changes in real-time sentence processing. *First Language, 16*, 193-222.

Warner, R. (2012). *Applied Statistics: From Bivariate Through Multivariate Techniques*. SAGE Publications (2nd edition). New York.

Wells, J., Christiansen, M., Race, D., Acheson, D., & MacDonald, M. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology, 58*, 250-271.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58,* 475-482.

Yuan, S., & Fisher, C. (2009). ‘‘Really? She blicked the baby?’’: Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*, *20*, 619–626.

Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child Development, 83,* 1382-1399.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition, 32,* 25–64.

Table 1. *Summary of participant standard scores on the standardized test measures for younger and older groups*

	Younger (N= 132)	**Older** (N= 124)
Nonverbal IQ		
Leiter Scale [a]		
M	112.2	112.7
SD	14.9	13.0
Lexical		
CREVT-R [b]		
Receptive		
M	104.6	107.7
SD	11.9	11.2
CREVT-R [b]		
Expressive		
M	103.3	100.5
SD	12.7	11.9
CASL[c]		
Antonyms subtest		
M	116.4	114.7
SD	13.3	13.5
Sentence		
CELF-4 [d]		
Concepts & Following Directions		
M	11.4	11.2
SD	2.0	2.1
CELF-4 [d]		
Recalling Sentences		
M	11.1	11.1
SD	2.5	2.5

Narrative		
TNL [e]		
Receptive		
M	11.1	11.1
SD	2.6	2.5
TNL [e]		
Expressive		
M	9.9	9.7
SD	2.8	3.0

[a] Leiter Scale = Leiter International Performance Scale (reported as full scale IQ).

[b] CREVT-R = Comprehensive Expressive-Receptive Vocabulary Test-Revised.

[c] CASL = Comprehensive Assessment of Spoken Language.

[d] CELF-4 = Clinical Evaluation of Language Fundamentals- 4th Edition.

[e] TNL = Test of Narrative Language.

Table 2. *Summary of participant demographics for the younger and older groups*

	Younger (N= 132)	**Older** (N= 124)
Age$_M$	8:1	10:8
Gender		
Male	24%	25%
Female	27%	23%
Race and Ethnicity		
Caucasian	43%	39%
African American	3%	3%
Hispanic	6%	4%
Asian	4%	4%
American Indian, Native Hawaiian	1%	2%
Mother's Education		
No college, some college 2-year degree	37%	39%
College or Above	15%	9%
Family Income		
Low (< $30k/yr)	10%	9%
High (≥$30k/yr)	42%	39%

Table 3. *Sentence interpretation accuracy (percent correct) by sentence type and age*

	Sentence Type			
	SVO	**SR**	**Passive**	**OR**
Younger				
M	80.2%	76.9%	50.2%	39.3%
SD	17.8	18.0	29.5	24.8
Range	6.1-100.0	21.2-100.0	0.0-100.0	0.0-100.0
Older				
M	88.9%	84.5%	66.1%	58.1%
SD	13.8	14.5	30.0	28.9
Range	15.6-100.0	36.4-100.0	0.0-100.0	0.0-100.0
Grand				
M	84.4%	80.6%	57.9%	48.3%
SD	16.5	16.8	30.7	24.4
Range	6.1-100.0	27.3-100	0.0-100	0.0-100

Table 4. *Sentence interpretation speed in ms (adjusted for simple motor speed) by sentence type and age*

	Sentence Type			
	SVO	**SR**	**Passive**	**OR**
Younger				
M	1199ms	1398ms	1564ms	1811ms
SD	872	1143	1003	1379
Range	-41 – 5736	-141 – 9818	49 – 5485	36 - 7692
Older				
M	782ms	936ms	912ms	1169ms
SD	721	730	676	955
Range	107 – 6094	103 – 5813	116 – 4350	-162 - 4719
Grand				
M	998ms	1175ms	1249ms	1500ms
SD	797	937	840	1167
Range	-41 – 6094	-141 – 9818	49 – 5485	-162 - 7692

Table 5. *Sentence interpretation error pattern (mean number of errors) by sentence type and age*

	Sentence Type			
	SVO	**SR**	**Passive**	**OR**
Younger				
Object Error				
M	3.9	4.2	13.6	16.3
SD	3.0	3.1	9.3	8.2
Range	0.0-14.0	0.0-13.0	0.0-33.0	0.0-33.0
PP-N Error				
M	2.7	3.4	2.8	3.7
SD	3.9	3.4	3.7	3.7
Range	0.0-27.0	0.0-18.0	0.0-26.0	0.0-22.0
Older				
Object Error				
M	2.4	3.0	10.0	11.7
SD	2.9	3.0	9.6	9.1
Range	0.0-15.0	0.0-14.0	0.0-33.0	0.0-33.0
PP-N Error				
M	1.2	2.1	1.2	2.1
SD	2.1	2.4	1.9	2.5
Range	0.0-12.0	0.0-10.0	0.0-14.0	0.0-12.0
Grand				
Object Error				
M	3.2	3.6	11.8	14.0
SD	2.9	3.0	9.4	8.7
Range	0.0-15.0	0.0-14.0	0.0-33.0	0.0-33.0
PP-N Error				
M	1.9	2.7	2.1	2.9
SD	3.0	2.9	2.8	3.1
Range	0.0-27.0	0.0-18.0	0.0-26.0	0.0-22.0

Note. Object error represents the incorrect selection of the image corresponding to the object of the sentence and PP-N error represents the incorrect selection of the image corresponding to the noun appearing in the prepositional phrase occurring near the end of the sentence.

Table 6. *Internal construct validity of the sentence interpretation measure as indicated by correlation across the four sentence types (accuracy)*

	Sentence Type			
	SVO	**SR**	**Passive**	**OR**
SVO	---	.84*	.33*	.35*
SR		---	.31*	. 32*
Passive			---	.89*

Note. *significant at .0001

Table 7. *Concurrent validity of the "whatdunit?" sentence interpretation measure as indicated by the correlation between overall interpretation accuracy score and performance across three standardized language domains*

	Language Domain		
	Lexical	**Sentence**	**Overall**
Overall Accuracy	.62*	.60*	.64*

Note. *significant at .0001

Table 8. *Internal reliability (Chronbach's alpha) of the sentence interpretation measure*

	Sentence Type				
	SVO	**SR**	**Passive**	**OR**	**Task Overall**
Reliability	.88	.86	.95	.94	.97

Figure 1. *Sentence interpretation accuracy for each sentence type by age group*

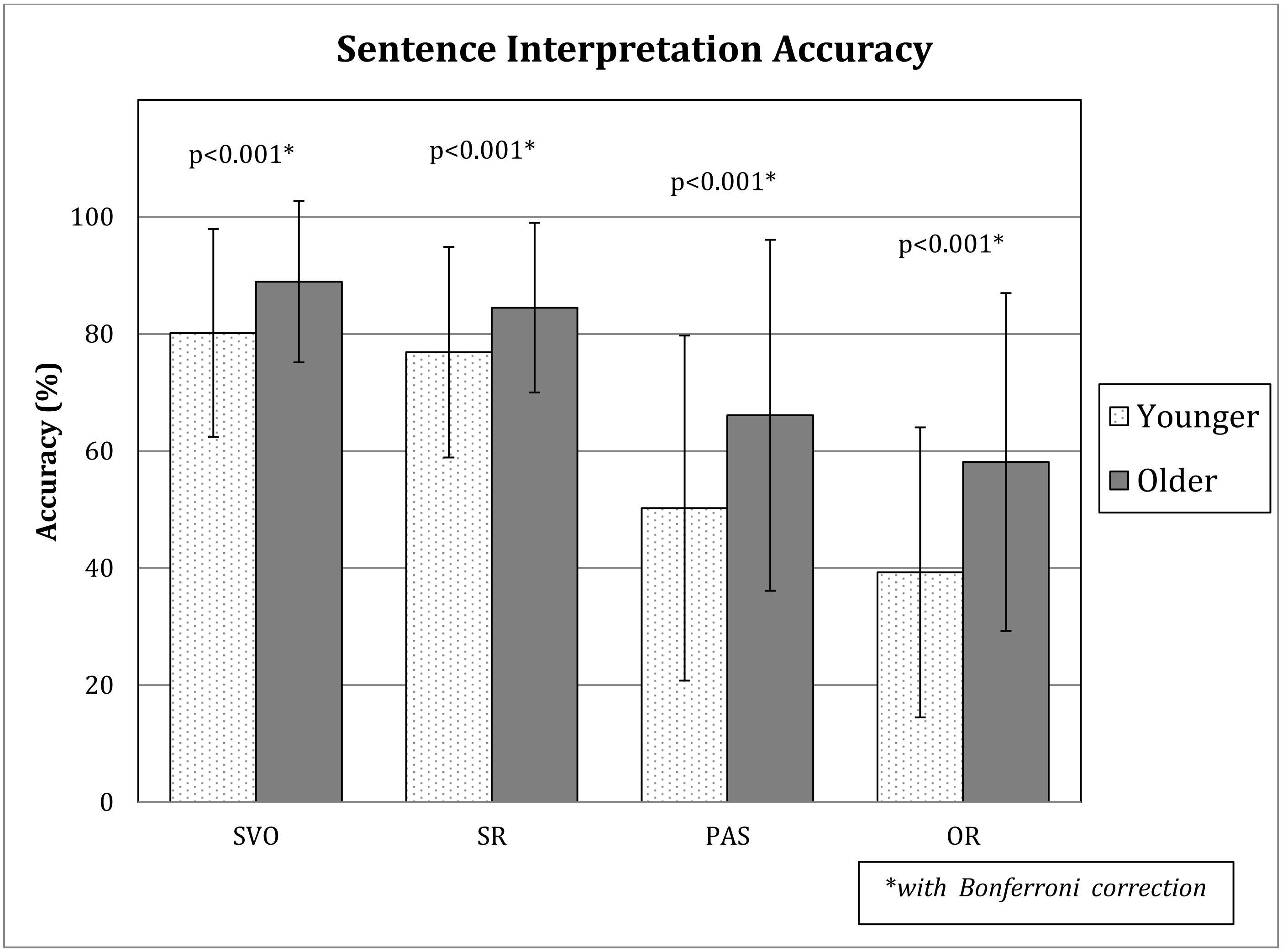

Figure 2. *Speed of sentence interpretation in ms (adjusted for simple motor speed) for each sentence type by age group*

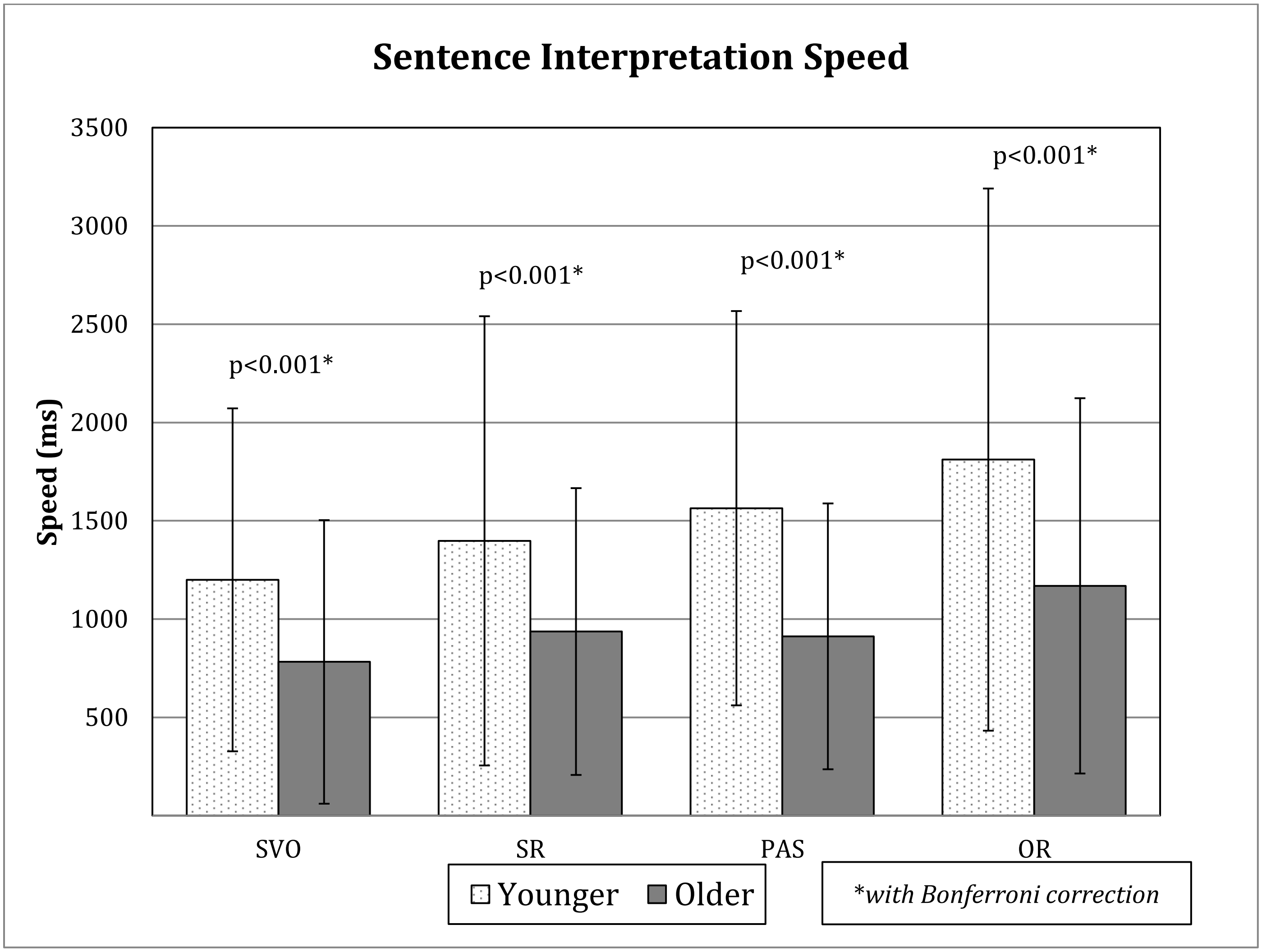

Figure 3. *Scatterplot of individual children's SVO sentence interpretation accuracy*

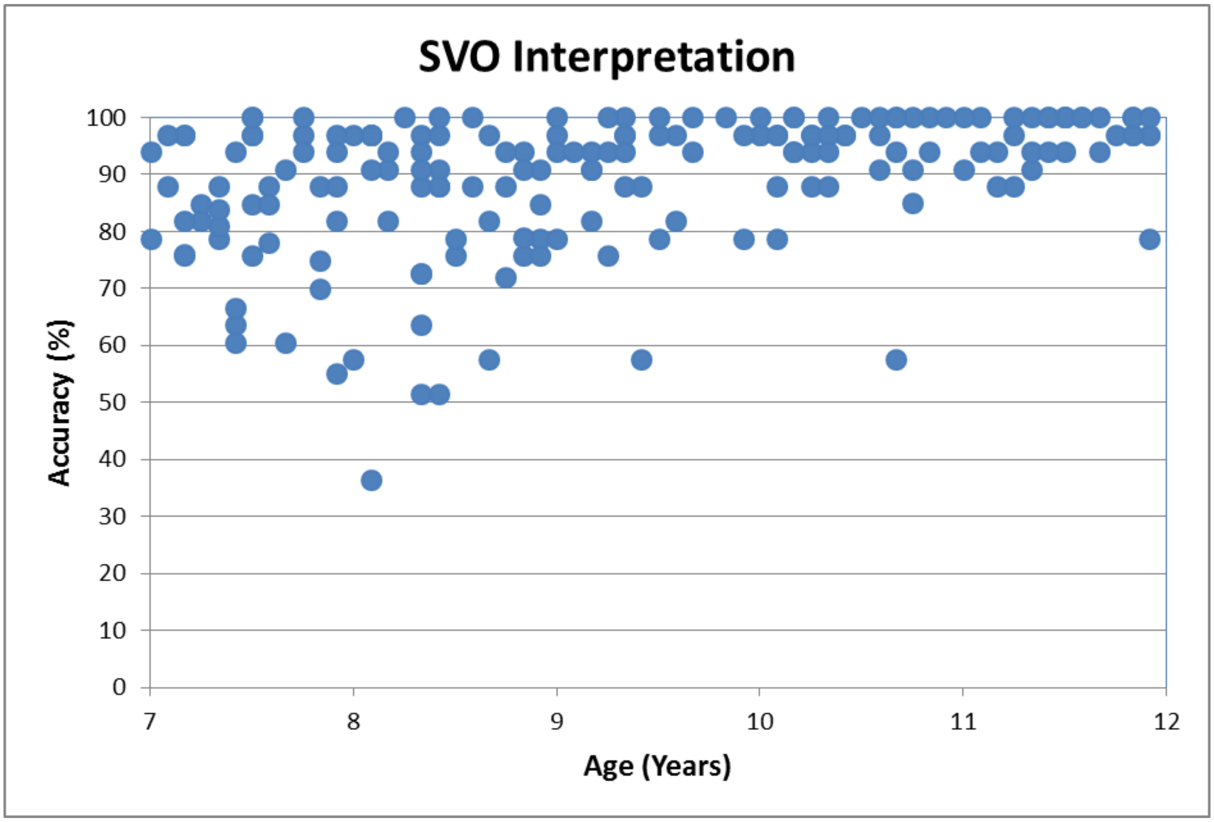

Figure 4. *Scatterplot of individual children's SR sentence interpretation accuracy*

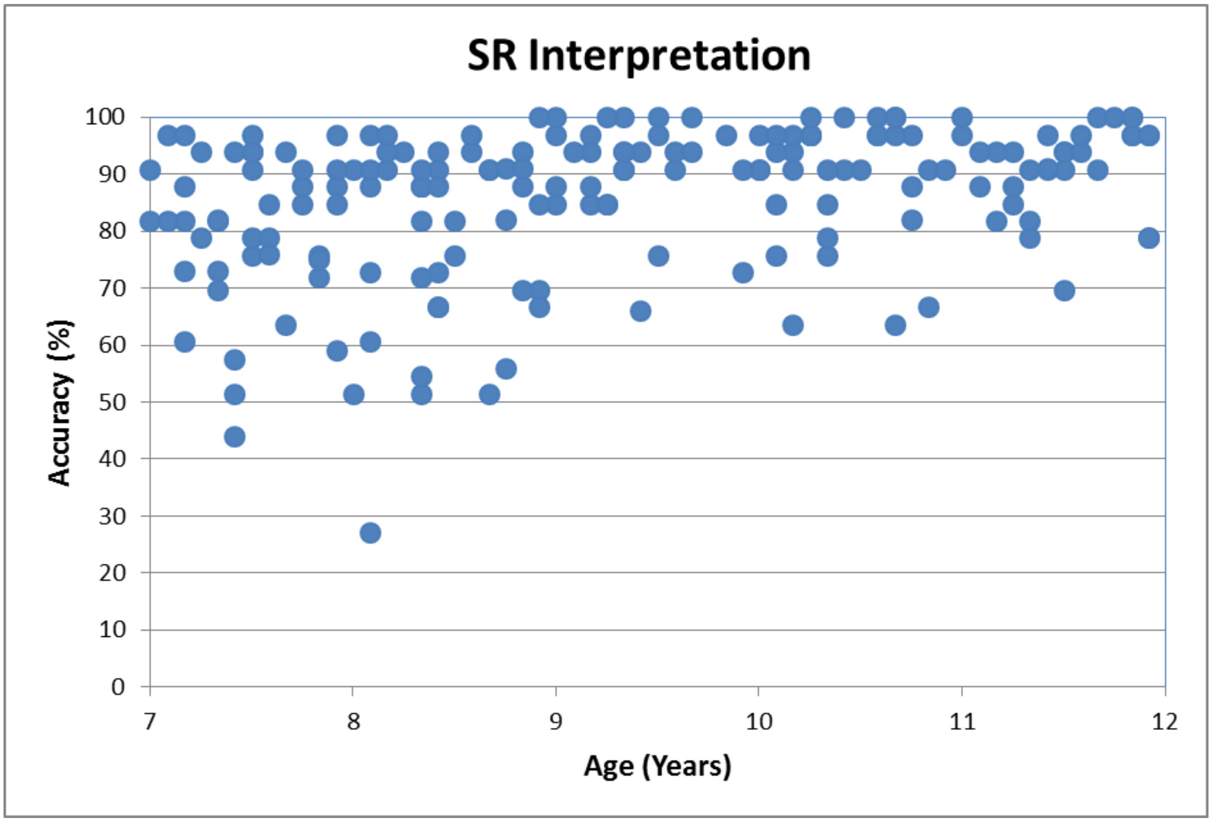

Figure 5. *Scatterplot of individual children's passive sentence interpretation accuracy*

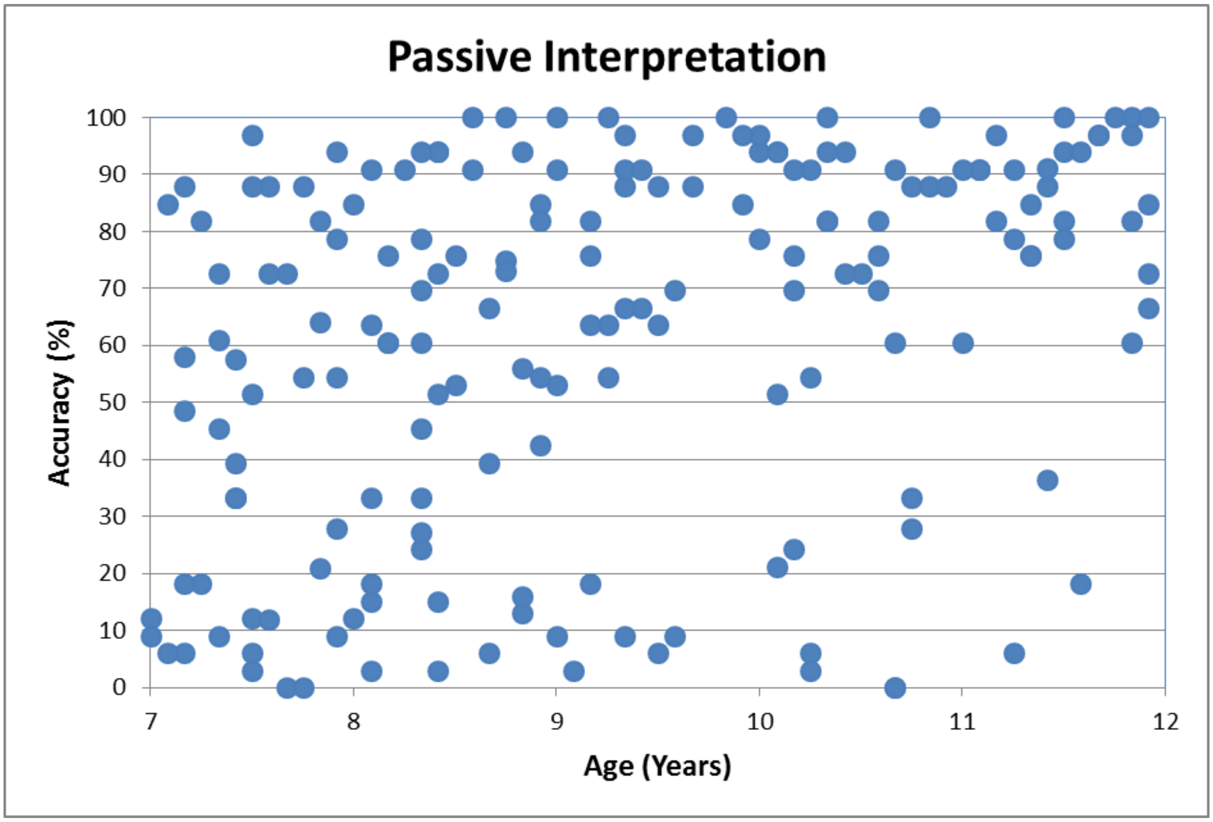

Figure 6. *Scatterplot of individual children's OR sentence interpretation accuracy*