



# City Research Online

## City St George's, University of London

**Citation:** Polišenská, K., Chiat, S., Comer, A. & McKenzie, K. (2014). Semantic effects in sentence recall: The contribution of immediate vs delayed recall in language assessment. *Journal of Communication Disorders*, 52, pp. 65-77. doi: 10.1016/j.jcomdis.2014.08.002

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

**Permanent repository link:** <https://city-ref-test.eprints-hosting.org/id/eprint/4806/>

**Link to published version:** <https://doi.org/10.1016/j.jcomdis.2014.08.002>

**Copyright and Reuse:** Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Accepted for publication in *Journal of Communication Disorders* (August 2014)

Semantic effects in sentence recall:

The contribution of immediate vs. delayed recall in language assessment

Kamila Poliřenská,<sup>a,\*</sup>, ShulaChiat<sup>b</sup>, AmandaComer<sup>b</sup>, KirstyMcKenzie<sup>b</sup>

<sup>a</sup> The University of Manchester, School of Psychological Sciences, Oxford Road,  
Manchester, M13 9PL, United Kingdom

<sup>b</sup> City University London, Language and Communication Science, Northampton Square,  
London EC1V0HB, United Kingdom

### **Abstract**

Sentence recall is increasingly used to assess language. It is widely debated what the task is actually testing, but one rarely explored aspect is the contribution of semantics to sentence recall. The few studies that have examined the role of semantics in sentence recall have employed an ‘intrusion paradigm’, following Potter and Lombardi (1990), and their paradigm relies on interference errors with conclusions based on an analysis of error patterns. We have instead manipulated the semantic plausibility of whole sentences to investigate the effects of semantics on immediate and delayed sentence recall. In Study 1, adults recalled semantically plausible and implausible sentences either immediately or after distracter tasks varying in lexical retrieval demands (backward counting and picture naming). Results revealed significant effects of plausibility, delay, and a significant interaction indicating increasing reliance on semantics as the demands of the distracter tasks increased. Study 2, conducted with 6-year-old children, employed delay conditions that were modified to avoid floor effects (delay with silence and forward counting) and a similar pattern of results emerged. This novel methodology provided robust evidence showing the effectiveness of delayed recall in the assessment of semantics and the effectiveness of immediate recall in the assessment of morphosyntax. The findings from our study clarify the linguistic mechanisms involved in immediate and delayed sentence recall, with implications for the use of recall tasks in language assessment.

**Keywords:** sentence; immediate recall; delayed recall; semantics; assessment

## 1. Introduction

Sentence repetition has been shown to be a reliable tool for identifying specific language impairment (SLI) in both children and adults (Conti-Ramsden, Botting & Faragher, 2001; Hesketh & Conti-Ramsden, 2013; Poll, Betz, & Miller, 2010; Redmond, Thompson, & Goldstein, 2011). It has also been shown to discriminate groups of children with SLI from typically developing children in every language tested to date (for example, Dutch: Van Daal, Verhoeven, van Leeuwe, & van Balkom, 2008; Cantonese: Stokes, Wong, Fletcher, & Leonard, 2006; French: Thordardottir, Kehayia, Mazer, Lessard, Majnemer, Sutton, Trudeau, & Chilingaryan, 2011). However, there is ongoing concern about what sentence repetition is actually testing, as highlighted by a recent article by Riches (2012, p.499) which emphasized again that ‘sentence repetition is poorly understood’.

The bulk of theoretical research on verbal recall has been concerned with ‘phonological’ or ‘auditory’ memory, and experimental tasks have focused on the single-word level in order to eliminate (or at least minimize) the contribution of semantics. Even so, effects of lexical knowledge, including semantics, are evident. Both children and adults have shown superior recall of words compared to nonwords, and this finding has been replicated across numerous languages and age groups (Hulme, Maughan & Brown, 1991; Multhaup, Balota & Cowan, 1996; Saint-Aubin & Poirier, 2000; Graf Estes, Evans & Else-Quest, 2007; Coady & Evans, 2008; Hoff, Core & Bridges, 2008). Turning to the specific contribution of semantics, Romani, McAlpine and Martin (2008) pointed out that concreteness effects are linked to semantic processing, and replicated previous findings of Walker and Hulme (1999) who showed that concrete words are recalled significantly better than abstract words. The effects of concreteness are often attributed to less rich semantic representations in words of low imageability. Miller and Roodenrys (2009) manipulated word frequency and word concreteness within the same experiment, and in addition to replicating the previously

observed effects of frequency and concreteness on recall, found a significant interaction between these factors. When the stimuli were more concrete, the effect of word frequency was smaller. This suggests that the semantic factor of concreteness makes phonology more readily accessible, counteracting the effects of lower lexical exposure.

Extending beyond the lexical level, the potential role of semantics is greater, raising questions about the relative role of phonology and semantics (Acheson, Postle, & MacDonald, 2010; Alloway, 2007; Schweppe, Rummer, Bormann & Martin, 2011). In a paper that has been highly influential, Potter and Lombardi (1990) argued for a central role of semantics in sentence recall. They suggested that sentence recall is based on a propositional-conceptual structure which is generated during comprehension of the sentence and that the sentence is regenerated via mechanisms involved in normal sentence production rather than retaining the sentence as a surface string of words. Two priming mechanisms are posited which add to the regeneration process. First, the syntactic structure is primed while processing the sentence. Second, the entries in the mental lexicon that have been part of the presented sentence are activated and therefore have a higher probability of being selected during the repetition. Thus, a propositional representation and syntactic/lexical priming provide a basis for verbatim sentence recall without assuming surface short-term storage of phonological representations (Potter & Lombardi, 1998). The regeneration hypothesis was supported by evidence from a set of experiments which used a novel intrusion paradigm. Participants were presented with a sentence, e.g. *'The knight rode around the palace searching for a place to enter'*, followed by a list of nouns. A lure word, e.g. 'castle', was included in the list for half of the participants. All of the participants were then given a probe word and they had to decide if that word appeared in the sentence before recalling the sentence. Potter and Lombardi tested stimuli presented both visually and auditorily and concluded that the

tendency for semantic intrusions in sentence recall appeared in both presentations, which was in line with their regeneration hypothesis.

It should be noted that Potter and Lombardi's recall task incorporated delay with interruption. This point was taken up by Rummer and Engelkamp (2003) who challenged Potter and Lombardi's conclusion (1990, p.648) that 'only conceptual and lexical representations are involved'. In a study based on the intrusion paradigm, Rummer and Engelkamp compared adult performance on immediate vs. delayed recall, in both visual and auditory presentation conditions. In the immediate condition, the word list preceded the model sentence and recall immediately followed. In the delayed condition, the word list and/or probe intervened between the model sentence and sentence recall. The immediate condition yielded more accurate sentence recall than the delayed condition, and showed no effects of lure words with auditory presentation, suggesting a limited role of semantics when a sentence has an acoustic form and is recalled without a delay. In contrast, when sentences were presented visually, lure words affected sentence recall even in the immediate condition. The authors stress that the amount of surface phonological information involved in sentence recall depends on the delay between sentence presentation and recall and conclude that if phonological information is available (as in immediate recall), it will contribute to sentence recall. These outcomes echo findings from an early study comparing immediate and delayed sentence recognition (Sachs, 1967). Adults were presented with a set of sentences and had to detect if an 'original sentence' (e.g. *He sent a letter about it to Galileo, the great Italian scientist*) was identical to the 'test sentence' or had been changed. Changes were either in form: (e.g. *He sent Galileo, the great Italian scientist, a letter about it*) or in meaning (e.g. *Galileo, the great Italian scientist, sent him a letter about it*). Sachs found that adults were equally able to detect changes in form and in meaning when the test sentence was presented immediately after the original sentence. However, when there was a delay filled with 80 or

160 syllables of text there was a sharp decline in their ability to recall the form compared to the immediate condition. In contrast, their ability to detect changes in meaning was well preserved even in the delayed conditions leading Sachs (1967) to suggest that the form of a sentence is stored for less time than its meaning.

These findings on the greater role of semantics relative to phonology in delayed recall do not rule out a contribution of semantics in immediate recall. Indeed, in a recent study of recall in 4-6-year-old children [names removed for anonymity] (under review), we found that span for semantically plausible sentences was significantly higher than span for semantically implausible sentences, which was in turn considerably higher than span for lists of unrelated words. This finding has been replicated in Arabic-speaking children (Wallan, Chiat & Roy, 2011). Thus, semantic content affects the preservation of form in immediate recall, and this has been observed in both adults and children. However, there is some indication that the relative contribution of semantics changes with the demands of the recall task.

This theoretical issue has important implications for what is assessed in sentence repetition tests. There is a striking gap between the clinical usage of immediate recall (e.g. Clinical Evaluation of Language Fundamentals Revised, CELF-R; Semel, Wiig, & Secord, 1994), and the theoretical research focus on delayed recall (e.g. Potter & Lombardi 1990; Alloway, 2007). A notable exception is a study by McDade, Simpson and Lamb (1982) who looked at immediate and delayed repetition tasks in children. They concluded that immediate repetition might overestimate children's language knowledge. Children's spontaneous production scores correlated less with immediate repetition scores and children were able to repeat sentences even if they did not understand what they were repeating; in contrast, delayed repetition seemed to reflect children's language skills better. Although McDade et al.'s conclusions have to be reviewed with caution (due to a very small sample and no detailed assessment of children's production and comprehension skills), their study raised a

key question of how presentation condition (immediate vs. delayed) might affect performance on repetition tasks, with implications for what they are testing. The conclusion McDade et al. draw from their results equates true linguistic knowledge with comprehension, and overlooks the different types of knowledge that may be required for immediate vs delayed recall, and how these relate to comprehension. An alternative view of their results is that accessing of semantics, required in sentence comprehension, is more crucial for delayed than immediate sentence recall, but that immediate recall is more informative about knowledge of lexical phonology and morphosyntax.

There is a considerable body of research on the effects of time delay and different types of distractors on recall, with implications for competing hypotheses about the source of ‘forgetting’. Berman, Jonides and Lewis (2009) and Lewandowsky, Oberauer and Brown (2009) argued that the source of ‘forgetting’ in short-term recall tasks is interference rather than time (decay), and showed that interference tasks affected recall more than simple delay. However, few studies have investigated the effect of different types of verbal distractors on recall. Studies using the ‘irrelevant sound paradigm’, where participants are instructed to ignore the irrelevant sound presented during the task, have shown that background noise interferes with recall performance (Banbury, Macken, Tremblay, & Jones, 2001). Two studies have reported that recall is affected by the semantic content of distractor material: Oswald, Tremblay, and Jones (2000) found that semantic interference was larger when the distractor participants listened to was meaningful speech compared to meaningless background speech. Similarly, Beaman, Hanczakowski, Hodgetts, Marsh, and Jones (2013) showed that semantically related distractors impaired recall more than semantically unrelated distractors. Surprisingly few studies have used a distracter task that engages participants’ production skills. Lewandowsky, Geiger and Oberauer (2008) found that repeated articulation of the same word (e.g. *super, super, super*) led to more forgetting than articulation of a changing

sequence (e.g. *super, table, house*). A recent study by Rose, Buchsbaum and Craik (2014) investigating retention of a single word following a 10-s silent delay with rehearsal, an easy math task, or a hard math task. Results showed that adults' recall was better following a rehearsal-filled than task-filled delay, and that the harder math task had a more detrimental effect on recall than the easier task. Collectively, these studies lead to broad predictions that delay with distractors will affect accuracy of recall, and that more demanding tasks will have greater effects. The two studies reporting effects of meaningfulness or meaning relations of distractors provide evidence that semantics is engaged in delayed recall. However, these studies have limited implications for our research because of differences in stimuli (usually single letters or words rather than sentences, and usually presented visually rather than auditorily).

The purpose of the current study was to investigate the effects of semantic content on immediate vs delayed recall of sentences which no previous study has systematically investigated. The study was carried out with adults and children in two separate experiments, with stimuli and recall conditions adapted to the different abilities of the two groups. Adults were included for two reasons. First, their results provide a better platform for a comparison with previous studies as the main focus of previous recall research is non-developmental. Second, adult speakers' longer attention span affords an opportunity to evaluate effects of more complex distractor tasks on preservation of phonology. Taking this into account, conditions in the adult and child studies differed in several respects. Adult delay conditions both involved distractor material since, in line with the 'interference hypothesis', piloting indicated that adult recall was not affected by silent delay. The two adult delay conditions used two phonological production tasks differing in lexical retrieval demands: counting backwards, and picture naming. In the child study, on the other hand, one delay condition was silent, while the other was filled with the simpler distractor task of forward counting. In

addition, the child stimuli were shorter than the adult stimuli, to accommodate the different thresholds for full recall in children and adults. These differences in stimuli and fillers will need to be taken into account in making inferences about the extent to which mechanisms used in immediate vs delayed recall are affected by language development.

For both children and adults, we hypothesized that:

- (1) Semantics will play a role in both immediate and delayed conditions, with number of words correctly recalled better for plausible than implausible sentences, as meaningful and familiar information is more resistant to interference than less meaningful information. This plausibility effect is also predicted by the conceptual regeneration hypothesis by Potter and Lombardi (1990) and would replicate our previous findings [anonymous].
- (2) Verbatim recall will be better for immediate than delayed conditions: children will recall more words immediately than after silent delay or forward counting; adults will recall more words immediately than after backward counting or picture naming. Immediate recall is expected to be better as phonology is available alongside other linguistic representations, while in delayed recall the phonological trace might be degraded due to interference in the distractor conditions (Lewandowsky et al., 2009), and in the case of children, the greater likelihood of ‘self-induced’ interference through loss of attention or distraction during silent delay (not observed in adult studies or in our pilot).
- (3) Delayed recall will be further affected by the intervening activity (Berman et al., 2009; Rose et al., 2014; Lewandowsky et al., 2008). In children, more words will be recalled correctly in the delayed condition filled with silence than in the delayed recall filled with counting. This is expected because producing digits provides phonological

material which will interfere with recall performance, while silence does not interfere with the phonological trace and at the same time provides an opportunity for rehearsal (although disruption may occur through self-induced distraction/interference). In adults, both interfering tasks involve production of phonological material. However, digits form a small class of items whose sequence, forward and even backward, is expected to be highly familiar, involving minimal semantic processing; in contrast, lexical items selected for the picture naming task were concrete, semantically rich and unpredictable, requiring both phonological and semantic processing. Due to the different production demands of the distractor tasks, it is expected that more words will be correctly recalled in the backward counting condition than in the picture naming condition.

- (4) Finally and importantly, we predict an interaction between conditions, with the effects of semantic plausibility increasing in line with increases in disruption. The lack of message can be compensated for in immediate recall because surface information is still available, but the lack of message cannot be fully compensated for in delayed recall as access to the phonological representations has decreased. Hence, the gap between plausible and implausible sentences is expected to be greater in delayed recall.

Both studies received ethical approval from the Language and Communication Sciences Research Ethics Committee, [name removed for anonymity].

## **2.1. Study 1: The effect of semantic plausibility on delayed and immediate recall in adults**

### **2.1.1. Participants**

Twenty-four adult native speakers of English participated in this study, equally divided between genders. All were employees from one department within a local authority council office in south-east England who had responded to an invitation to participate sent via email, together with an information letter about the study and a consent form. Participants were aged between 18 and 50 years, and occupied either clerical administration or managerial positions in the council office. Rather than testing undergraduate students as is the case in many studies in psycholinguistics, we included adults who were less likely to have been exposed to academic/more complex language and who have not participated in psycholinguistic experiments before so they were truly naive to the nature of the experiment. Participants were of different backgrounds and ages, so the sample was more representative. The inclusionary criteria were English as a first language and the exclusionary criteria were hearing or neurological problems.

### **2.1.2. Stimuli**

Targets comprised 36 sentences, equally divided into 18 semantically plausible (SP) and 18 semantically implausible (SI) sentences. Each set was further divided according to length (9, 12 and 15 words), with six sentences at each length. These six sentences were then divided between the three recall conditions described below, such that each participant received two SP and two SI sentences at each of the three lengths in each of the three conditions. Allocation of the six same-length SP/SI sentences to the three recall conditions was counterbalanced to ensure any differences observed between conditions could not be due to specific items.

Semantically implausible sentences were created by replacing the content words in semantically plausible sentences with different content words that were matched for number

of syllables, and chosen to violate selectional restrictions (Chomsky, 1965) resulting in sentence anomalies meanings, as illustrated by the following pair:

SP: The white cat was chasing him in the park.

SI: The long pig was washing him in the fork.

This process ensured that SP and SI targets were matched for length in terms of number of words, and number of syllables, and were also matched for syntactic and prosodic structure.

The final pools of content words used for SP and SI sentences were compared on several lexical variables known to affect lexical access of individual words (Davis, 2005). Independent samples *t*-tests revealed that lexical properties of words in plausible sentences did not significantly differ from the lexical properties of words from implausible sentences: frequency  $t(206) = .33, p = .74$ , based on CELEX (Baayen, Piepenbrock & van Rijn, 1995); frequency  $t(206) = .15, p = .88$ , based on British National Corpus; familiarity  $t(206) = 1.52, p = .13$ , based on MRC database (Coltheart, 1981); imageability  $t(206) = .47, p = .64$ , based on MRC database (Coltheart, 1981); and age of acquisition  $t(206) = .04, p = .97$ , based on Bird, Franklin and Howard (2001). Hence, SP and SI sentences were also matched for lexical properties of constituent words. Examples of plausible and implausible sentences are presented in Appendix A.

### 2.1.3. Procedure

The test was carried out in a small room in the local authority council offices where the participants were employed. The test was presented on a laptop screen through headphones to ensure consistency and high-quality sound. Stimuli were presented in three conditions.

- 1) Immediate recall: Participants were told they would hear a sentence and must repeat it straight back. No visual instructions were presented.
- 2) Delayed recall with backwards counting: Participants heard the target sentence and were required to count backwards from ten to one before repeating it. Written instructions appeared on the laptop screen 'Count back from 10 to 1 aloud' and stayed displayed till the counting stopped.
- 3) Delayed recall with interruption by picture naming: Participants heard the target sentence and were required to name a set of four pictures that appeared before they repeated the sentence. All pictures were selected from the Receptive One Word Picture Vocabulary Test (Brownell, 2000), and depicted highly familiar objects that were easy to name, to ensure that the naming itself was not challenging. Pictures were displayed simultaneously (i.e. 4 pictures in a row, e.g. milk, box, letter, tin). The pictures appeared immediately after the auditory presentation of the sentence finished, and were displayed for a time period necessary for a participant to name them and then repeat the sentence. All participants named items promptly so time differences between them would have been negligible, and although it is likely that there was some variation, this level of variation was not likely to be critical in terms of our results and interpretation.

The presentation order of SP and SI sentence stimuli was counterbalanced across participants. The recall conditions were presented in a fixed order with immediate recall first, followed by delayed recall with counting and finally delayed recall with naming. The stimuli were presented via a laptop showing a PowerPoint presentation with sound clips attached to slides. The slides showed 36 boxes with numbers (1 to 36) in 6 rows and the sound was attached to these boxes. When a sentence was played, the box was highlighted. After repetition of a sentence associated with a particular box, the box faded and the next box was

highlighted and so on till all 36 sentences were played. Prior to each new presentation condition, the experimenter explained the procedure for the next 12 sentences. After each row of 6 sentences the participants were given a visual reward of a smiley face over the last item, and an auditory reward at the end of each block of 12 sentences of an ‘applause’ sound recording.

Participants were asked to repeat sentences in serial order. Responses were recorded using an audio recorder and scored by counting the number of words the participant repeated back correctly in each sentence. These scores were then totalled according to the repetition condition (immediate, delay counting, delay naming) and sentence type (semantically plausible or implausible) for each participant.

#### **2.1.4. Results**

Figure 1 presents boxplots with the percentages of correctly recalled words according to recall condition and semantic plausibility, with a maximum of 72 for each combination of recall and plausibility conditions (two 9-word sentences, two 12-word sentences and two 15-word sentences).

FIGURE 1 INSERT ABOUT HERE

This figure reveals a decrease in scores across the three recall conditions which is steeper for the SI than the SP sentences. The effects of the two factors were explored with a 2 (sentence type: SP vs SI) x 3 (recall condition: immediate, delayed with backward counting, and delayed with picture naming) repeated measures ANOVA. Although our hypotheses were targeted, we took a conservative approach and applied Bonferroni corrections for multiple comparisons throughout. All *p*-values are presented with these corrections applied. This

analysis yielded significant main effects of plausibility,  $F(1, 23) = 283.86, p < .001, \eta^2 = .93$ , recall condition,  $F(2, 46) = 133.12, p < .001, \eta^2 = .85$  and a significant interaction,  $F(2, 46) = 27.42, p < .001, \eta^2 = .54$ . Post-hoc analyses revealed that plausible sentences were recalled significantly better than implausible sentences (mean difference 22.05%,  $p < .001$ , CI 95% 19.34 - 24.76). Sentences in the immediate condition were recalled significantly better than sentences presented for delayed recall, both with counting (mean difference 18.18%,  $p < .001$ , CI 95% 14.45 - 21.91) and naming (mean difference 29.97%,  $p < .001$ , CI 95% 24.15 - 35.78). Within the delayed recall conditions, counting hindered recall significantly less than naming (mean difference 11.79%,  $p < .001$ , CI 95% 7.23-16.34). Figure 1 and follow-up comparisons using *t*-tests showed that significant differences between plausible and implausible sentences were present in all recall conditions, but the size of the effect is larger in the delayed conditions, both with counting and naming. The results also followed a step-wise pattern for recall condition: immediate > delayed counting > delayed naming, and although the same pattern was found in both plausible and implausible conditions, the size of the difference between immediate and delayed with counting conditions was much larger in implausible sentences. Due to multiple comparisons, the *p*-level was adjusted to  $\alpha = .007$ ; all significant differences were  $p < .001$ .

### 2.1.5. Discussion

The results were strikingly straightforward and in line with hypotheses. Not only did both factors (plausibility and timing of recall) impact on the number of words recalled, but the two interacted such that the impact of the delay was greater when the plausibility of targets was reduced. In immediate repetition, verbatim recall was close to ceiling: on average, just 3.65% words were not repeated correctly in the plausible condition, and even in the implausible condition, just 13.24% words were affected. This performance greatly exceeded

the repetition of unrelated sequences of words found by Pickering & Gathercole (2001), confirming the major contribution of lexical and syntactic knowledge to immediate recall ([names removed for anonymity], under review). Nonetheless, the significant difference between plausible and implausible targets demonstrates a contribution of sentential semantics, which is also in line with findings for children reported in [names removed for anonymity], (under review). Findings in the delayed conditions demonstrated that this semantic contribution increased if different phonological representations are activated, suggesting that these overwrite the phonology of the recall target, with repercussions for the preservation of lexical and syntactic representations. It seems that the relative demands of the phonological intrusion influence the relative contribution of semantics in recall. The demands of counting backwards are low, suggesting that this output task can be performed relatively automatically. Here, the drop in scores for plausible sentences was on a par with the average drop for implausible sentences in the immediate condition (13.36% and 13.24% respectively), and a third of the average drop for implausible sentences (39.90%), indicating that reliance on semantics is already considerably higher than in immediate recall. When the task demands in terms of lexical retrieval increased and participants had to access both lexical semantics and phonology, the disruption of phonological representations was greater and this led to a greater reliance on semantics. This was demonstrated by performance in the implausible condition where participants preserved fewer than half the target words on average (a drop of 53.42% for implausible compared with 23.40% for plausible sentences).

These findings are in line with the argument that forgetting is affected by the cognitive load of interfering activity (Lewandowsky et al., 2009). The contrast between counting backward from 10 to 1 vs naming a random set of picture is strikingly similar to the contrast between repeated production of the same word vs production of constantly changing material reported to have differential effects in Lewandowsky et al. (2008).

## **2.2. Study 2: The effect of semantic plausibility on delayed and immediate recall in children**

### **2.2.1. Participants**

Twenty-four typically developing six-year-olds (mean age 6;6, SD = 3 months) took part in this experiment. Children were recruited through an independent primary school in London and only children whose parents gave written consent participated. The exclusionary criteria were: any developmental disorders as reported by teacher and parents via questionnaires, English spoken as an additional language, and referral to speech and language therapy services.

### **2.2.2. Stimuli**

As in Study 1, the targets were 36 sentences, half semantically plausible and the other half semantically implausible, matched on syntactic structure and number of words and syllables. For instance, the semantically plausible sentence ‘The new teacher waited as we got on the bus’ was matched by the semantically implausible sentence ‘The short water counted as we wrote on the rice’. Again, each set of 18 sentences was equally divided between three lengths, with six sentences per length, but in this case, sentences were 9, 10 or 11 words long (as opposed to 9, 12 or 15 words in the adult study). The shorter sentences accommodated the lower recall capacity of the younger participants evident in piloting. Again, the six sentences at each length were divided between the three recall conditions described below, such that each participant received two SP and two SI sentences at each of the three lengths in each of the three recall conditions. Allocation of the six same-length SP/SI sentences to the three recall conditions was counterbalanced to ensure any differences observed between conditions could not be due to specific items.

There were 18 sentences with a range of syntactic structures. Half of the sentences used ( $n = 9$ ) were simple sentences, most often with the structure Subject, Verb, Object and Adjunct (e.g. I have been blowing pretty bubbles the whole day). The other 9 included one example of coordination (e.g. The funny clown told us a joke and we all laughed) and 8 complex sentences with an embedded clause acting as an adjunct (e.g. When I fell in the park, I hurt my knee). Both simple (SVO/SVOA) as well as complex sentences were considered to be acquired by 6-year-old monolingual English speaking children (e.g. Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002; Diessel & Tomasello, 2005). Crucially for our study, syntactic structures were matched across conditions so could not account for any differences observed between conditions.

As in Study 1, the final stimulus pools of content words used for semantically plausible and semantically implausible sentences were compared on several lexical variables. Independent samples *t*-tests revealed that lexical properties of words in plausible sentences did not significantly differ from the lexical properties of words from implausible sentences: frequency  $t(172) = 1.34$ ,  $p = .18$ , based on CELEX (Baayen et al., 1995); frequency  $t(172) = 1.47$ ,  $p = .14$  based on British National Corpus; familiarity  $t(172) = .32$ ,  $p = .75$ , based on MRC database (Coltheart, 1981); imageability  $t(172) = .59$ ,  $p = .56$ , based on MRC database (Coltheart, 1981), and age of acquisition  $t(172) = 1.68$ ,  $p = .09$ , based on Bird, Franklin and Howard (2001). In sum, the sentences differed on semantic plausibility, but were matched on syntactic and prosodic structure and lexical and phonological properties of individual words. Examples of plausible and implausible sentences are presented in Appendix B.

### 2.2.3. Procedure

The task was administered to children individually in a quiet room at their school. The sentences were recorded by a native speaker of British English and presented over a laptop in

three different recall conditions:

- 1) Immediate condition: Participants were asked to repeat a sentence as soon as they had heard it.
- 2) Delayed condition: Participants were asked to wait for 10 seconds before repeating the sentence. The experimenter used a stop watch to time the ten seconds. When the period of 10 seconds elapsed, the experimenter pointed to children as a signal to repeat the sentence back. The participant could not see the stop watch face, so as to reduce their distraction from remembering the sentence. Practice trials were provided before testing started to ensure that all children understood the task.
- 3) Interrupted condition: The experimenter explained that after hearing the sentence the participant had to count from 1 to 10 aloud and then recall the sentence. If participants failed to automatically count aloud the experimenter would prompt them by saying “1, 2, 3...” etc. until the participant started counting aloud independently.

After explaining each condition the participant was asked to practise until they were confident in what they had to do for that condition. The presentation order of semantically plausible and implausible sentence stimuli was counterbalanced across participants. The recall conditions were presented in a fixed order with immediate recall first, followed by delayed recall with silence and finally delayed recall with interruption. Responses were audio recorded and scored for number of words correctly recalled.

#### **2.2.4. Results**

Figure 2 shows the boxplots for correctly recalled words according to recall condition and semantic plausibility. The maximum for each combination of recall and plausibility conditions was 60 words (two 9-word sentences, two 10-word sentences and two 11-word sentences).

## FIGURE 2 INSERT ABOUT HERE

This figure reveals a decrease in scores across the three recall conditions which is steeper for the SI than the SP sentences. Indeed, in the case of semantically plausible sentences, the decrease between immediate and silent delay conditions is negligible. These observations are substantiated by the results of a repeated-measures ANOVA, which revealed significant main effects of sentence plausibility ( $F(1, 23) = 68.93, p < .001, \eta^2 = .75$ ), recall condition ( $F(2, 46) = 38.72, p < .001, \eta^2 = .63$ ), and a significant interaction between these ( $F(2, 46) = 11.92, p < .001, \eta^2 = .34$ ). Post-hoc analyses revealed that plausible sentences were recalled significantly better than implausible sentences (mean difference 12.73%,  $SE = 1.54, p < .001$ , CI 95% 9.56 - 15.90). Sentences in the immediate condition were recalled significantly better than sentences in the delayed conditions, both with silence (mean difference 5.07%,  $SE = 1.64, p = .015$ , CI 95% 0.85 - 9.29) and delayed with counting (mean difference 15.76%,  $SE = 2.11, p < .001$ , CI 95% 10.34 - 21.2). The delayed conditions in turn differed significantly: counting hindered recall significantly more than silence (mean difference 10.69,  $SE = 1.71, p < .001$ , CI 95% 6.28 - 15.11). Figure 2 and follow-up comparisons using *t*-tests showed that significant differences between plausible and implausible sentences occurred in all recall conditions, but the size of the effect was larger in the delayed recalled conditions, both with silence and counting. The results followed a step-wise pattern for the implausible condition only: immediate > delayed silence > delayed counting. In the plausible condition the following pattern was found: immediate = delayed silence > delayed counting. Due to multiple comparisons, the *p*-level was adjusted to  $\alpha = .007$  and all significant differences were  $p < .007$ .

### 2.2.5. Discussion

The results for Study 2 are in line with our hypotheses, and are largely in parallel with the results for Study 1. Again, both plausibility and presentation condition impacted on the number of words recalled, with delay having a greater impact on implausible targets. Plausibility and recall again interacted as in the first experiment. In the immediate condition, verbatim recall of the six-year-olds in this study was close to ceiling: on average, just 6.39% words were not repeated correctly in the plausible condition, and even in the implausible condition an average 11.67% words were affected. Hence, performance greatly exceeded children's repetition capacity for unrelated sequences of words found by Pickering and Gathercole (2001), confirming that they too draw on lexical and syntactic knowledge in immediate recall [names removed for anonymity]. The finding that this information was almost as well preserved in the silent delay condition when targets were plausible (drop of 7.36% vs 6.39%), but was more depleted when targets were implausible (drop of 20.83% vs 11.67%), indicates the vulnerability of lexical/syntactic information and increasing reliance on sentential semantics over time. This vulnerability increased when the delay was filled with phonological material, even when that material consisted of the highly familiar sequence of numbers in forward counting. The drop for plausible targets, though significant, was relatively small: on average, the number of words children recalled was immediately reduced by just over 15% when recall followed forward counting. The drop was over twice as great for implausible targets (an overall drop of 34.51%).

Our finding of no difference between immediate and silent delay in the plausible condition is in line with the interference hypothesis. However, if forgetting is due to interference, we have to consider why children's performance deteriorated in the implausible condition when there was no interfering material. This finding could indicate limitations in the children's sustained attention, leading to internally-induced interference. On this

interpretation, the finding of plausibility effects in the silent delay condition indicates that semantics is more robust and supports phonological recall after interference due to internally-induced distraction as well as externally-provided distractors.

### **3. General discussion**

Table 1 provides an overview of the successive reductions in the proportion of words correctly recalled by children and adults across the plausibility and presentation conditions.

TABLE 1 INSERT ABOUT HERE

#### **3.1. The effect of semantic plausibility**

The results confirmed our first hypothesis. Sentence recall was affected by semantic plausibility, with semantically plausible sentences recalled better than semantically implausible sentences in both children and adults. Our studies demonstrated that semantics contributes to sentence recall, which was in line with studies using intrusion paradigms (Potter & Lombardi, 1990; Rummer & Engelkamp, 2001; 2003; Alloway, 2007; Schweppe et al., 2011). The key difference in the present study is the method of investigation. While previous studies manipulated lure words in order to investigate the effect of semantics, we manipulated the plausibility of whole sentences. To our knowledge this method has only previously been used in Miller and Isard (1963), when looking at intelligibility of semantically plausible and implausible sentences at adult English native speakers. [names removed for anonymity], (under review) and Wallan et al. (2011) have shown similar results in the immediate recall of sentences using typologically different languages (Arabic, Czech and English). The present study has now extended the method of using sentence plausibility to delayed recall in both children and adults.

### **3.2. The effect of the presentation condition and intervening tasks**

The effect of the presentation condition was also in line with our hypotheses. The effect was significant in both the child and adult studies and followed the same pattern: the more complex the production task, the greater the effect was on recall. Adults' immediate recall was better than delayed recall with backward counting and delayed recall with backward counting was better than delayed recall with picture naming. Backward counting involves retrieval of lexical phonology. As digits are a small closed class of sequenced items, it could be argued that the adults' knowledge of lexical semantics was minimally involved while performing this task. In contrast, a picture naming task requires engagement of both lexical phonology and lexical semantics and the drop in recall was therefore in line with our predictions. The pattern of results found in children was similar, but there was only one distractor task in Study 2 and therefore any differentiation between effects of lexical retrieval demands cannot be made. Children performed significantly better on immediate recall compared to recall with silent delay, and this was in turn better than recall with forward counting. Thus, with additional time, both adults and children showed a significant shift from lexical and syntactic representations to a greater reliance on semantics. This shift became even more marked when the time was filled with phonological production and production demands increased.

However, there were some differences between adults and children. Our piloting showed that silence and forward counting had a minimal effect on adults, but the findings from the child study found performance affected by silent delay and even more by forward counting. It is possible that children apply fewer strategies or rehearse less (Gathercole & Hitch, 1993) and/or that they are more liable to loss of attention and other distractions even when there is no distractor task; in addition, children's more limited experience may mean

that counting is less automatic. Backward counting affected adults, but less so than picture naming, suggesting a more automatic response in this counting condition. Indeed the percentage of words recalled correctly by adults in the backward counting condition was on a par with children's recall performance on the forward counting condition.

As pointed out above, these findings are consistent with Lewandowsky et al.'s argument that forgetting is due to interference and varies with the cognitive load of interfering activity. However, our evidence cannot adjudicate between 'decay' and 'interference' accounts of forgetting since our study was not designed to evaluate these competing hypotheses, for example, the timing of delay in the distractor conditions was not measured or controlled.

Furthermore, inferences about the source of differences between presentation conditions come with an important caveat. Since the order of presentation conditions was not counterbalanced, it is possible that declines in performance across the presentation conditions were a consequence of fatigue or boredom rather than task demands. However, there are a number of reasons why order effects are unlikely to account for observed differences. First, previous studies with children and adults (e.g. McDade et al., 1982; Riches, 2012; Rummer & Englekamp, 2003) have produced a similar pattern of results, with participants scoring better in immediate recall compared to delayed recall. Second, in our experience, repetition is a straightforward task that requires little attention or effort. This view is supported by evidence from a number of studies of sentence recall. One of our own studies (currently under review) involved repetition of sentences in 9 conditions that varied in difficulty, with order of conditions counterbalanced. Despite the fact that there were twice as many trials as in the present study, and children were younger (4-5 years), we found that they had no difficulty completing the task (100 out of 106 children completed all conditions), and crucially, no order effects were found. Similarly, recent studies by Frizelle and Fletcher (2014a; 2014b) found

that children aged 4-7 years were able to repeat 69 sentences in one session and again no order effects were found. Finally and importantly, our key research question was whether the contribution of semantic information depends on the timing of recall. The order of plausible and implausible sentences within each condition was counterbalanced, so effects of plausibility within each condition cannot be due to order of presentation, and it is unlikely that the ordering of the presentation conditions would give rise to the differential effects of plausibility on these conditions that we observed (see Figures 1 and 2). Taken together, these findings suggest that fatigue and boredom are unlikely to explain the pattern of results we have found. Note that this is not at odds with our suggestion that inattention *during silent delay* may have been a factor in children's pattern of response, with effects of silent delay in the implausible but not plausible condition.

### **3.3. The interaction between semantic plausibility and the presentation condition**

A significant interaction was found between semantic plausibility and recall conditions in both children and adults: the effects of semantic plausibility increased as disruption increased, as evidenced by a much steeper drop in correctly recalled words for implausible sentences in the delayed conditions. It is important to remember that the sentences were matched on syntactic and prosodic structure and also the lexical properties of individual items (frequency, familiarity, age of acquisition and imageability), so the differences between plausible and implausible sentences cannot be attributed to these linguistic variables. Instead, the differences between the two sentence types must be due to the semantic relations between the constituent words. This suggests that plausibility supports the retention of lexical items and syntax, particularly when recall is interrupted by an extra phonological task. Our findings and interpretations are in line with both Potter and Lombardi (1990) and Rummer and Engelkamp (2003).

When discussing linguistic mechanisms involved in immediate and delayed recall, it may be worthwhile examining elements of speech production models, e.g. ‘message formulation’, ‘grammatical encoding’ and ‘phonological encoding’ (Levelt, 1989). Our expectation is that all three of these stages would be reflected in delayed recall while immediate recall relies more on lexical phonology and morphosyntax, but less on semantics (Chiat, Armon-Lotem, Marinis, Polišenská, Roy, & Seeff-Gabriel, in press). In immediate recall, a participant is provided with a message and the lexical and grammatical forms are immediately available for re-activation, meaning that the ‘message formulation’ stage is not required; in contrast, if lexical and grammatical forms are compromised by interfering activity, it may be necessary to re-formulate the message, relying on retention of semantics. This interpretation is in line with the significant interaction found in our studies. The influential study by Potter and Lombardi suggested that conceptual representations and lexical/syntactic priming mechanisms are crucial for delayed sentence recall, but they could not comment on the distinction between immediate and delayed recall as their study only included the delayed condition. Our study adds to current understanding by highlighting the differences between the linguistic mechanisms involved in immediate and delayed recall and allows for an examination of the immediate/delayed recall distinction.

### **3.4. Implications of the findings for language assessment**

As pointed out in the introduction, our findings have implications for language assessment. Performance on repetition tasks has been found to correlate with scores on other languages tests, and sentence repetition is put forward as a quick and highly informative assessment of children’s language (Chiat & Roy, 2008; Gathercole & Baddeley, 1989; Gupta, MacWhinney, Feldman, & Sacco, 2003), and a clinical marker for language impairment (Bishop, North & Donlan, 1996; Conti-Ramsden et al., 2001). However, the underlying

mechanisms of these tasks are little understood. The findings of the present study highlight the importance of the presentation condition for the levels of language tested. Based on [names removed for anonymity] and the present study, we suggest that while immediate and delayed recall both draw on all levels of linguistic representation, the weight of their contribution differs.

Immediate sentence recall relies particularly on lexical phonology and morphosyntax, while delayed recall relies more on semantics. Our findings are in line with the small-scale study by McDade et al (1982) which showed that scores for delayed recall were a better reflection of spontaneous speech skills (which involves semantics as well as lexical phonology and morphosyntax) when compared to scores for immediate recall. However, rather than concluding that delayed recall is a *more accurate* measure of language knowledge, we propose that it *provides more information about semantic interpretation of linguistic input*. Immediate recall, on the other hand, provides more information about children's *lexical and morphosyntactic knowledge*. Along with the increased use of immediate repetition in clinical assessment, we therefore propose combining this with delayed recall. These tasks are targeted, quick and easy to administer and score, and together they provide complementary information about children's strengths and difficulties, and indicators for intervention. We propose the following interpretations for different profiles of performance, emphasising that these are tentative and call for investigation and evaluation in clinical populations:

- 1) Good performance on immediate recall has the clearest implications: if children's immediate recall is in line with expectations for their age, we can infer that their lexical phonology and morphosyntax are intact. Impairment in delayed recall would then imply difficulty in linking lexical phonology and morphosyntax to the meanings these encode, or difficulty with the meanings themselves. Semantics would therefore be the focus of intervention, and this might draw on the child's

strengths in processing and recognising linguistic forms that encode target meanings.

- 2) Conversely, impairment in immediate recall would point to deficits in lexical phonology and/or morphosyntax. These may arise from deficits in recognising words and structure, which could in turn be due to problems in processing phonological input, and/or increased susceptibility to distraction (as proposed above for typically developing children in the silent delay condition), and/or rapid decay of words and structures following recognition. If children's immediate recall is similar to their delayed recall, this would suggest relative strength in semantics, with children using the semantic information they were able to access from the input to re-construct the target (Wallan, Chiat & Roy, 2011). Intervention for such children would aim to strengthen phonological and morphosyntactic representations, for example by intensive exposure to target words and structures in relatively small chunks and with exaggerated prosody, to make the targets more accessible.
- 3) If children's immediate recall is impaired, but delayed recall is even more impaired, this may point to further decay of linguistic forms or further effects of distraction during the delay. However, it could also be due to limited accessing of semantic information from the input, and hence limited availability of semantic information to support recall. Intervention for children who show deficits in lexical phonology, morphosyntax and semantics would combine the aims and strategies of intervention in (1) and (2).
- 4) If children's performance in both immediate and delayed recall is as expected for their age, it is unlikely that they have a language impairment. However, this does

not rule out the possibility of problems with higher level language and language use.

While our findings are interesting and have the potential to inform theories of sentence recall and clinical use of sentence repetition, a number of cautions are in order. First, before drawing firm conclusions about differences *between* presentation conditions, our study needs to be replicated with the order of conditions counterbalanced. Second, there is a need to explore patterns of performance in children with language impairment in order to determine which of the profiles we have hypothesised actually occur, and consider further the implications for intervention . We also recommend replication of our study in different language to determine the generalizability of the effects of plausibility in different presentation conditions that we have observed. In addition, replication with typically developing children of different ages is needed to explore whether there are developmental changes in the type of information children retain immediately, after silent delay, and after distracting material.

Recall tasks are increasingly used in assessment and in a growing number of languages e.g. Cantonese (Stokes et al., 2006), French (Thordardottir et al., 2011), Italian (Devescovi & Caselli, 2007) and Slovak (Kapalková, Polišenská & Vicenová, 2013). An increased understanding of these tasks' underlying mechanisms will clarify what they are actually revealing about the strength and weaknesses of children with language impairment and potentially lead to improved targeting of interventions.

## References

- Acheson, D. J., Postle, B. R., & MacDonald, M. C. (2010). The interaction of concreteness and phonological similarity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1), 17–36.
- Alloway, T. P. (2007). Investigating the roles of phonological and semantic memory in sentence recall. *Memory*, 15(6), 605–615.
- Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1995). The CELEX database. *Nijmegen: Center for Lexical Information, Max Planck Institute for Psycholinguistics, CD-ROM*.
- Banbury, S. P., Macken, W. J., Tremblay, S., & Jones, D. M. (2001). Auditory distraction and short-term memory: Phenomena and practical implications. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 43(1), 12–29.
- Beaman, C. P., Hanczakowski, M., Hodgetts, H. M., Marsh, J. E., & Jones, D. M. (2013). Memory as discrimination: What distraction reveals. *Memory & Cognition*, 41(8), 1238–1251.
- Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(2), 317–333.
- Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, 33(1), 73–79.
- Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: evidence from a twin study. *Journal of Child Psychology and Psychiatry*, 37(4), 391–403.
- Brownell, R. (2000). *Expressive One-Word Picture Vocabulary Test Manual*. Novato, CA: Academic Therapy Publications.

- Chiat, S., & Roy, P. (2008). Early phonological and sociocognitive skills as predictors of later language and social communication outcomes. *Journal of Child Psychology and Psychiatry*, 49(6), 635-645.
- Chiat, S., Armon-Lotem, S., Marinis, T., Polišenská, K., Roy, P., & Seeff-Gabriel, B. (in press). The potential of sentence imitation tasks for assessment of language abilities in sequential bilingual children. *Bilinguals and assessment: State of the art guide to issues and solutions from around the world*. Clevedon: Multilingual Matters.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Oxford, England: M.I.T. Press.
- Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders*, 43(1), 1–40.
- Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, 33(4), 497–505.
- Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 42(6), 741-748.
- Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, 37(1), 65–70.
- Devescovi, A., & Caselli, C. M. (2007). Sentence repetition as a measure of early grammatical development in Italian. *International Journal of Language & Communication Disorders*, 42(2), 187-208.
- Diessel, H., & Tomasello, M. (2005). A New Look at the Acquisition of Relative Clauses. *Language*, 81(4), 882–906.
- Frizelle, P., & Fletcher, P. (2014a). Profiling relative clause constructions in children with specific language impairment. *Clinical Linguistics & Phonetics*, 28(6), 437–449.

- Frizelle, P., & Fletcher, P. (2014b). Relative clause constructions in children with specific language impairment: Relative clause constructions in children with SLI. *International Journal of Language & Communication Disorders*, 49(2), 255–264.
- Gathercole, S. E., & Baddeley, A. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, 28(2), 200–213.
- Gathercole, S. E., & Hitch, G. (1993). The development of rehearsal: A revised working memory perspective. In A. Collins, S. Gathercole, M. Conway, & P. Morris (Eds.), *Theories of memory*. Hove: Erlbaum.
- Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without Specific Language Impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 50(1), 177-195.
- Gupta, P., MacWhinney, B., Feldman, H. M., & Sacco, K. (2003). Phonological memory and vocabulary learning in children with focal lesions. *Brain and Language*, 87(2), 241-252.
- Hesketh, A., & Conti-Ramsden, G. (2013). Memory and language in middle childhood in individuals with a history of Specific Language Impairment. *PLoS ONE*, 8, 1-7.
- Hoff, E., Core, C., & Bridges, K. (2008). Non-word repetition assesses phonological memory and is related to vocabulary development in 20- to 24-month-olds. *Journal of Child Language*, 35(04), 903–916.
- Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, 30(6), 685–701.
- Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, 45(3), 337–374.

- Kapalková, S., Polišenská, K., & Vicenová, Z. (2013). Non-word repetition performance in Slovak-speaking children with and without SLI: novel scoring methods. *International Journal of Language & Communication Disorders*, 48(1), 78-89.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge: MIT Press.
- Lewandowsky, S., Geiger, S. M., & Oberauer, K. (2008). Interference-based forgetting in verbal short-term memory. *Journal of Memory and Language*, 59(2), 200–222.
- Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences*, 13(3), 120–126.
- McDade, H. L., Simpson, M. A., & Lamb, D. E. (1982). The use of elicited imitation as a measure of expressive grammar: a question of validity. *The Journal of Speech and Hearing Disorders*, 47(1), 19-24.
- Miller, G. A., & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, 2, 217–228.
- Miller, L. M., & Roodenrys, S. (2009). The interaction of word frequency and concreteness in immediate serial recall. *Memory & Cognition*, 37(6), 850–865.
- Multhaup, K. S., Balota, D. A., & Cowan, N. (1996). Implications of aging, lexicality, and item length for the mechanisms underlying memory span. *Psychonomic Bulletin & Review*, 3(1), 112–120.
- Oswald, C. J. P., Tremblay, S., & Jones, D. M. (2000). Disruption of comprehension by the meaning of irrelevant sound. *Memory*, 8(5), 345–350.
- Pickering, S., & Gathercole, S. E. (2001). *Working memory test battery for children (WMTB-C)*. London: Psychological Corporation.
- Poll, G. H., Betz, S. K., & Miller, C. A. (2010). Identification of clinical markers of Specific Language Impairment in adults. *Journal of Speech, Language, and Hearing Research*, 53, 414–429.

- Potter, M., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, 29(6), 633–654.
- Potter, M., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, 38(3), 265–282.
- Redmond, S. M., Thompson, H. L., & Goldstein, S. (2011). Psycholinguistic profiling differentiates Specific Language Impairment from typical development and from Attention-Deficit/Hyperactivity Disorder. *Journal of Speech, Language, and Hearing Research*, 54, 99–117.
- Riches, N. G. (2012). Sentence repetition in children with specific language impairment: an investigation of underlying mechanisms. *International Journal of Language & Communication Disorders*, 47, 499–510.
- Romani, C., McAlpine, S., & Martin, R. (2008). Concreteness effects in different tasks: Implications for models of short-term memory. *The Quarterly Journal of Experimental Psychology*, 61(2), 292–323.
- Rose, N. S., Buchsbaum, B. R., & Craik, F. I. M. (2014). Short-term retention of a single word relies on retrieval from long-term memory when both rehearsal and refreshing are disrupted. *Memory & Cognition*.
- Rummer, R., & Engelkamp, J. (2001). Phonological information contributes to short-term recall of auditorily presented sentences. *Journal of Memory and Language*, 45(3), 451–467.
- Rummer, R., & Engelkamp, J. (2003). Phonological information in immediate and delayed sentence recall. *The Quarterly journal of experimental psychology. A, Human experimental psychology*, 56(1), 83–95.
- Sachs, J. S. (1967). Recopition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 2(9), 437–442.

- Saint-Aubin, J., & LeBlanc, J. (2005). Word frequency effects in immediate serial recall of pure and mixed lists: Tests of the associative link hypothesis. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 59, 219-227.
- Saint-Aubin, J., & Poirier, M. (2000). Immediate serial recall of words and nonwords: Tests of the retrieval-based hypothesis. *Psychonomic Bulletin & Review*, 7(2), 332-340.
- Schweppe, J., Rummer, R., Bormann, T., & Martin, R. C. (2011). Semantic and phonological information in sentence recall: Converging psycholinguistic and neuropsychological evidence. *Cognitive Neuropsychology*, 28(8), 521-545.
- Semel, E., Wiig, E. & Secord, W. (1994). *Clinical Evaluation of Language Fundamentals - Revised*. San-Antonio, TX: Psychological Corporation.
- Stokes, S. F., Wong, A. M.-Y., Fletcher, P., & Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: the case of Cantonese. *Journal of Speech, Language, and Hearing Research*, 49(2), 219-236.
- Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of Primary Language Impairment at age 5. *Journal of Speech, Language, and Hearing Research*, 54(2), 580-597.
- Van Daal, J., Verhoeven, L., van Leeuwe, J., & van Balkom, H. (2008). Working memory limitations in children with severe language impairment. *Journal of Communication Disorders*, 41(2), 85-107.
- Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(5), 1256-1271.

Wallan, A., Chiat, S., & Roy, P. (2011). *Evaluation of an Arabic Sentence-Repetition Test for Preschool Children*. Poster presented at the ASHA Convention 2011, San Diego.

**Appendix**

- A. Examples of semantically plausible (SP) and semantically implausible (SI) sentences matched on length (in words and syllables) and syntactic structure used in Study 1.

SP: The young dancer was looking at the shiny mirror.

SI: *The wet pencil was resting at the sunny hammer.*

SP: The funny clown was jumping on the big sofa.

SI: *The rusty house was dancing on the round ladder.*

SP: I saw my black cat in a tall tree.

SI: *I read my green dog in a sad glass.*

- B. Examples of semantically plausible (SP) and semantically implausible (SI) sentences matched on length (in words and syllables) and syntactic structure used in Study 2.

SP: The grumpy waiter took the soup to the kitchen staff.

SI: *The fluffy candle showed the soap to the dinner time.*

SP: The red bus was late so we drove by car.

SI: *The red grass was brave so we spoke to jam.*

SP: We asked for a new fork because this one is dirty.

SI: *We baked for the black moon because this one is thirsty.*

Table 1. Percentage of words repeated correctly according to the plausibility and presentation conditions in the child and adult studies

| Presentation condition       | Plausible sentences (SD) | Implausible sentences (SD) |
|------------------------------|--------------------------|----------------------------|
| Adult study                  |                          |                            |
| Immediate                    | 96.35 (2.61)             | 86.76 (7.89)               |
| Delay with backward counting | 86.64 (6.71)             | 60.10 (12.62)              |
| Delay with naming            | 76.60 (11.64)            | 46.58 (14.85)              |
| Child study                  |                          |                            |
| Immediate                    | 93.61 (7.08)             | 88.33 (6.67)               |
| Silent delay                 | 92.64 (5.96)             | 79.17 (13.62)              |
| Delay with forward counting  | 84.93 (12.01)            | 65.49 (17.04)              |

Figure 1. Boxplots for word recall according to semantic plausibility and recall condition

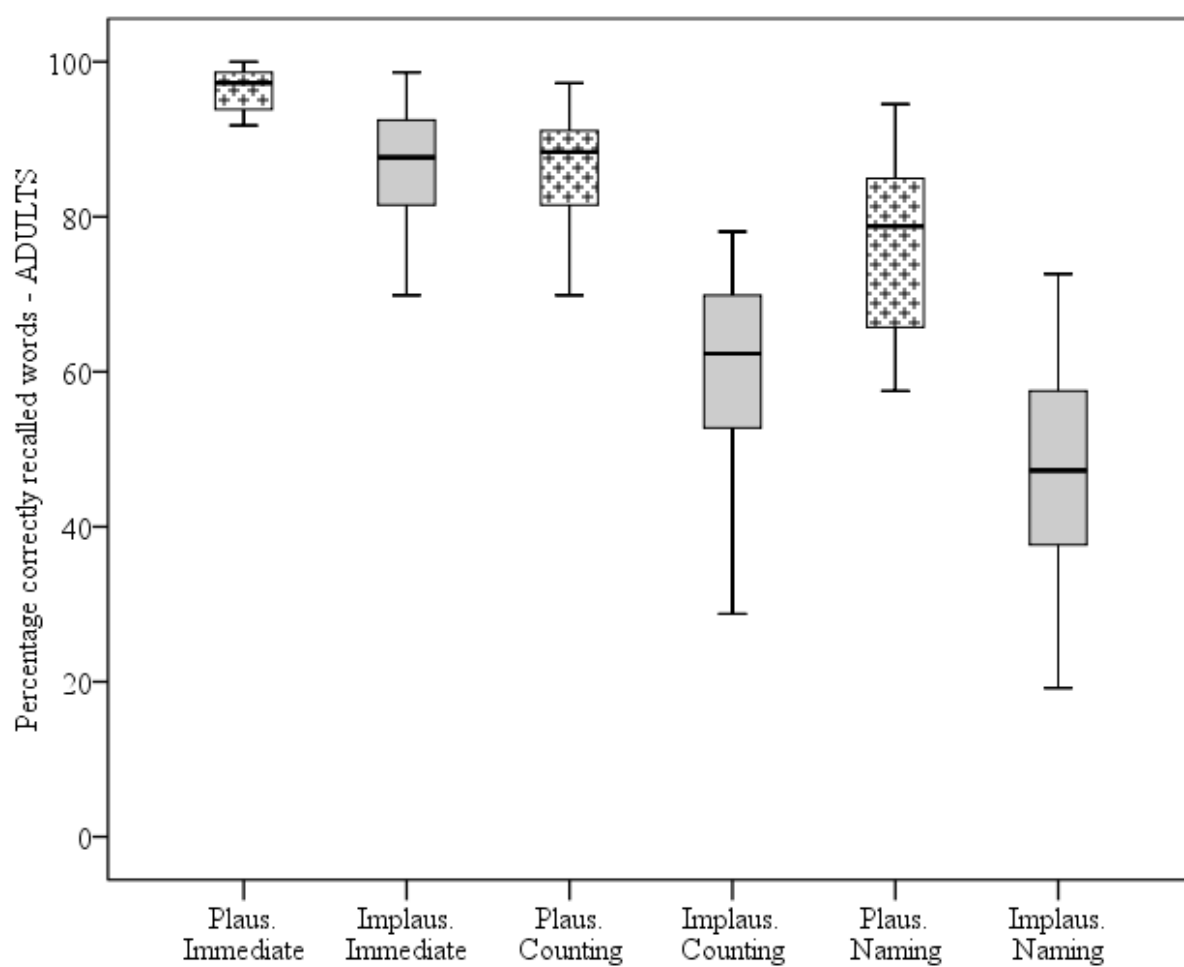


Figure 2. Boxplots for word recall according to semantic plausibility and recall condition

