



Verbal short-term memory capacities and executive function in semantic and syntactic interference resolution during sentence comprehension: Evidence from aphasia

Yingying Tan, Randi C. Martin*

Rice University, Department of Psychology, 6100, Main Street, Houston, TX 77005, USA



ARTICLE INFO

Keywords:

Interference
Sentence comprehension
Semantic short-term memory
Executive function
Cue-based retrieval

ABSTRACT

This study examined the role of verbal short-term memory (STM) and executive function (EF) underlying semantic and syntactic interference resolution during sentence comprehension for persons with aphasia (PWA) with varying degrees of STM and EF deficits. Semantic interference was manipulated by varying the semantic plausibility of the intervening NP as subject of the verb and syntactic interference was manipulated by varying whether the NP was another subject or an object. Nine PWA were assessed on sentence reading times and on comprehension question performance. PWA showed exaggerated semantic and syntactic interference effects relative to healthy age-matched control subjects. Importantly, correlational analyses showed that while answering comprehension questions, PWA' semantic STM capacity related to their ability to resolve semantic but not syntactic interference. In contrast, PWA' EF abilities related to their ability to resolve syntactic but not semantic interference. Phonological STM deficits were not related to the ability to resolve either type of interference. The results for semantic interference are consistent with prior findings indicating a role for semantic but not phonological STM in sentence comprehension, specifically with regard to maintaining semantic information prior to integration. The results for syntactic interference are consistent with the recent findings suggesting that EF is critical for syntactic processing.

1. Introduction

Given the ubiquitous presence of long distance dependencies, successful sentence comprehension would seem to necessarily draw on working memory (WM) or certain types of short-term memory (STM) capacity (e.g., phonological STM, semantic STM) as a means of maintaining earlier constituents to link with later ones. However, there is considerable debate on the nature of the critical memory resources involved (Caplan et al., 2013; Caplan and Waters, 1999, 2013; Daneman and Carpenter, 1980; Daneman and Merikle, 1996; Evans et al., 2015; Fedorenko et al., 2006, 2007; Johns et al., 2014; Just and Carpenter, 1992; MacDonald and Christiansen, 2002; McElree, 2000; Misyak and Christiansen, 2012). In fact, there is debate over whether memory capacity per se plays any role in comprehension or whether other cognitive abilities, such as the general executive function (EF) involved in resolving interference (Novick et al., 2005), plays a more crucial role. The present study addressed these issues by examining sentence processing for brain damaged PWA with varying degrees of STM and EF deficits.

1.1. Role of STM in sentence comprehension

Prior investigations of the role of STM or WM capacity in comprehension have typically focused on the consequences of the number of intervening words between elements in a sentence that have to be integrated (e.g., Gibson, 1998). According to prominent sentence processing models, the greater the distance across which such integrations occur (for instance for object vs. subject relatives), the higher the WM demand (Gibson, 1998, 2000). Research on healthy individuals has typically focused on complex span measure like reading span and whether those with greater WM capacities on such measures have more difficulty with sentences with a greater integration distance. Complex span measures have been used as earlier research suggested that simple span measures (such as digit or word span) had lower relations to sentence processing than did reading span (Daneman and Carpenter, 1980). Just and Carpenter (1992) argued that the complex reading span measure reflected a domain-general capacity for carrying out all aspects of language processing. Some findings have supported a relation between complex span and the comprehension of sentences with longer distance integrations (Fedorenko et al., 2006, 2007) though there is

* Correspondence to: Department of Psychology, Rice University, MS-25, P.O. Box 1892, Houston, TX 77251-1892, USA.
E-mail addresses: Yingying.tan@mpi.nl (Y. Tan), rmartin@rice.edu (R.C. Martin).

contradictory evidence as well (Caplan and Waters, 2013, 1999; Evans et al., 2015; Reali and Christiansen, 2007; Van Dyke et al., 2014).

In the study of the relation of WM and comprehension for persons with aphasia (PWA), the evidence is mixed as well. Most such studies have focused on STM performance as measured by simple span tasks, given that the dual task nature of complex WM tasks presents a considerable challenge for these individuals (but see Miyake et al. (1994, 1995) and Sung et al. (2009) for evidence supporting the role of verbal WM in PWA sentence processing). Here also, evidence has suggested that STM measures tapping primarily phonological retention do not relate to comprehension deficits (Caplan and Waters, 1999, 2013; Friedmann and Gvion, 2003; Hanten and Martin, 2000; Martin and Romani, 1994), though there is some recent contradictory evidence here as well (Caplan et al., 2013; Pettigrew and Hillis, 2014). On the other hand, STM measures tapping the ability to retain semantic information have been shown to relate to sentence comprehension deficits for particular sentence structures where several words have to be maintained prior to integration. These sentence types include sentences with several adjectives preceding a noun or several nouns preceding a verb (Hanten and Martin, 2000; Martin et al., 1994; Martin and He, 2004). However, PWA with semantic STM deficits did not show sensitivity to the distance between two words signaling a grammatical error (Martin and Romani, 1994; Martin and He, 2004). Thus, Martin and colleagues proposed a multiple capacities approach in which there are separable systems for maintaining phonological and semantic information, with semantic storage but not phonological storage being relevant for comprehension. In addition, the findings from the grammaticality judgments suggested that a different capacity was involved in maintaining syntactic information. Recent results from Wright et al. (2007) study also supported the multiple capacities approach by showing that PWA's deficits in comprehending syntactically complex sentences (i.e., passive and objective-relative clauses) was predicted by their performance in the *n*-back task requiring syntactic level processing, but not the *n*-back task requiring either semantic or phonological level processing.

1.2. Role of EF in sentence comprehension

A separate body of research, with more recent origins, has postulated a role for general executive abilities in sentence comprehension – specifically, an ability to exert cognitive control in resolving conflict between different representations within a level (e.g., with the syntactic level) or across levels (between semantic and syntactic levels). Those examining the role of EF in sentence processing have often employed garden path sentences, where the initially favored interpretation of an ambiguous syntactic structure has to be overridden (e.g., for the subject/object ambiguity in “While Anna bathed the baby played in the crib”), under the assumption that EF is necessary to suppress the inappropriate interpretation to allow for consideration of a less frequent alternative structure. Individual differences in EF have been shown to relate to the ability to make the required revision following a garden path for healthy individuals (Engelhardt et al., 2017; Hsu and Novick, 2016; Novick et al., 2014; Thothathiri et al., 2012) and PWA (Novick et al., 2005; Vuong and Martin, 2011, 2015). Case studies of PWA with EF deficits have documented an impaired ability to comprehend sentence with lexical or syntactic ambiguities together with a better ability to process matched unambiguous sentences (Vuong and Martin, 2011, 2015).

While most studies of the role of executive function in comprehension have focused on the processing of ambiguous sentences where a preferred interpretation has to be overridden, recently, some have also started examining comprehension of sentences with other types of conflict (Alexander, 2006; Dickey and Thompson, 2009; Thothathiri et al., 2012, 2017), or in naturalistic conversation rather than in lab experiments (Frankel et al., 2007; Nicholas and Connor, 2017; Penn et al., 2010). Some researchers have suggested that executive function

deficits may underlie the comprehension difficulty observed for more complex sentences like passives or object relatives, because initial assumptions about a sentence interpretation may have to be inhibited. For instance, for a relative clause sentence, individuals may assume that the head noun will be the subject of the relative clause. This expectation will be met for subject relatives but for object relatives, this expectation will not be fulfilled and predicted grammatical assignments will have to be undone via processes such as executive function in order to arrive at the appropriate interpretation. There is evidence for such revision for healthy individuals (Traxler et al., 2002), but it appears that so far few studies with brain damaged PWA have examined this issue.

1.3. Cue-based parsing, working memory, and executive function

As indicated above, prominent theories of sentence comprehension have focused on the working demands assumed to derive from distance – that is, from integrating sentence elements across intervening constituents. By contrast, the relatively recent cue-based parsing theory offers a different perspective, focusing on the semantic and syntactic interference among sentence elements as the source of comprehension difficulty (Lewis and Vasishth, 2005; McElree et al., 2003; Segaert et al., 2012; Van Dyke, 2007; Van Dyke and Johns, 2012; Van Dyke and Lewis, 2003; Van Dyke and McElree, 2006, 2011). According to the cue-based parsing approach, as each word in a sentence is processed, semantic and syntactic features are generated and these bundles of features are stored in memory. When a later element is processed that should be integrated with an earlier one, retrieval cues are generated in order to access the relevant target. Consider, for example, the processing of the sentence, “Toys from Lisa arrived yesterday.” The first word “toys” would be encoded as having the grammatical features of being a plural subject noun. Semantic features such as “definite,” “inanimate object,” and “used for play” would also be encoded. When the main verb “arrived” is processed, retrieval cues would be generated to locate the subject noun so that it could be integrated with this verb. These retrieval cues would include grammatical features such as “noun”, “subject”, and semantic features consistent with an entity that can arrive. These retrieval cues are used to access in parallel memory representations that fit these cues. In this example, “toys” would provide the best match, but “Lisa” would also provide a partial match as “Lisa” is a noun that can arrive, but not a subject. According to the cue-based parsing approach, comprehension difficulty results from the interference engendered by such partial matches. Semantic or syntactic interference occurs when the distractor(s) partially matches the semantic or syntactic retrieval cues, respectively. When there are two or more possible candidates for integration, some interference resolution mechanism must come on line to select the appropriate target.

Given the emphasis on retrieval from memory, one might have expected that WM or STM capacity would play a critical role in maintaining elements that would later be retrieved. However, many studies have demonstrated that it is the degree of interference from non-target elements during integration, rather than the distance across which integrations occur, that determines comprehension difficulty (Glaser et al., 2013; Lewis and Vasishth, 2005; McElree, 2000, 2015; Segaert et al., 2012; Van Dyke and Lewis, 2003; Van Dyke and McElree, 2011). To accommodate these findings, the cue-based parsing approach adopts recent general models of WM termed embedded processes models, that assume WM consists of the activated portion of long-term memory, together with a very limited capacity for the focus of attention (i.e., from 1 to 2 items). Access to items outside the focus depends on parallel access using cue-based retrieval (Cowan, 1999; Lewis, 1996; McElree, 2001, 2006; Oberauer, 2002).

Because of the very limited capacity of the focus of attention, researchers who support the cue-based parsing account assume that everyone has the requisite capacity, and consequently, individual differences in capacity would not be a relevant factor in comprehension success. Instead, what is critical are individual differences in the

richness of the encoding of the semantic and syntactic features of items, which would depend on the individual's language expertise derived from linguistic experience (Lau et al., 2006; MacDonald and Christiansen, 2002; Perfetti, 2007; Van Dyke et al., 2014; Wells et al., 2009) and biological factors such as phonological precision (MacDonald and Christiansen, 2002). According to this experience-based approach, one would not expect measures of WM or STM capacity to relate to sentence comprehension performance. Researchers taking the cue-based parsing approach have emphasized individual differences in the richness of the linguistic representations – with those with better linguistic abilities and verbal knowledge showing smaller interference effect (Van Dyke et al., 2014). Beyond the effects of linguistic knowledge, one might also predict that measures of EF would relate to comprehension ability, given the emphasis on interference and interference resolution in comprehension.

Recently, Tan et al. (2017) addressed these assumptions through a large scale study of healthy young adults, examining the relation of individuals' sensitivity to semantic and syntactic interference during sentence comprehension to individual differences in measures of working memory, language ability and executive function. The WM measures included two complex span tasks (reading span, Daneman and Carpenter, 1980; operation span, Unsworth et al., 2005) and two simple span tasks (digit span reflecting phonological retention and category probe reflecting semantic retention, Martin and Romani, 1994). The additional cognitive measures consisted of a measure of verbal knowledge (i.e., WAIS vocabulary) and a measure of the ability to resolve interference (i.e., the Stroop task). WAIS vocabulary was included as a measure of verbal ability. Even after controlling for vocabulary, subjects with better general WM capacity (as indexed by a composite of two complex WM span tasks) showed smaller syntactic interference effects in self-paced reading and smaller semantic interference effects in question answering. In addition, subjects with better semantic retention ability showed smaller semantic but not syntactic interference effects in question answering. However, performance on the Stroop task and on the phonological STM task was unrelated to semantic and syntactic interference resolution for all dependent measures. Taken together, these results support a role of general WM capacity and semantic STM in sentence processing, which partially contradict the predictions from cue-based parsing account.

Although Tan et al.'s (2017) study shed some new light on the cognitive mechanisms underlying sentence comprehension, it also opened some areas of further investigation. First, this study was conducted with healthy college students at a highly selective university who had relatively limited variation in their cognitive abilities. Thus, it is possible that a more extensive role for working memory and executive function might be established in a population with brain damage, displaying a wider range of individual differences on both WM/STM and other cognitive abilities and on interference in sentence comprehension. Testing such individuals would provide a means of obtaining evidence converging with that for healthy young subjects from a sample showing a wider range of abilities. Second, given prior reports of a significant relation between performance on the Stroop task and performance on complex WM tasks (Kane and Engle, 2003; Tan et al., 2017), it is somewhat unexpected that subjects' syntactic interference resolution ability in Tan et al. (2017) was related to WM but not Stroop. The lack of a Stroop-language relation might have been caused by the fact that only a response time but not error rate measure for Stroop was available, as most subjects' performed near ceiling. Previous findings suggest that RT and error rate-interference effects in the Stroop task reflect different EF mechanisms – specifically, goal maintenance vs. competition resolution, respectively (Kane and Engle, 2003). As interference resolution is the aspect of sentence comprehension under investigation, it may be more informative to assess the role of competition resolution as indexed by error rate effects in the Stroop task in the individuals with aphasia, who make significantly more errors than do controls (Vuong and Martin, 2011). Finally, even though numerous

Table 1
Example stimuli with syntactic and semantic interference.

Sentence type	Example
LoSyn/LoSem	The jockey who had challenged the unbeatable record yesterday will win.
LoSyn/HiSem	The jockey who had challenged the unbeatable champion yesterday will win.
HiSyn/LoSem	The jockey who claimed that the record was unbeatable yesterday will win.
HiSyn/HiSem	The jockey who claimed that the champion was unbeatable yesterday will win.
Question	<i>Will the jockey win?</i>

Note. “Lo-” and “Hi-” refer to low and high interference condition, while “-Syn” and “-Sem” refer to syntactic interference and semantic interference condition.

studies have been completed with healthy populations, prior investigations of sentence comprehension deficits in brain damaged populations have not investigated the role of semantic and syntactic interference resolution during comprehension. However, studies from Dickey et al. (2007) and Dickey and Thompson, (2009) have suggested that PWA's comprehension deficits might occur because they were more vulnerable to interference from other competitors in the story context. The possible role of EF ability in determining patient comprehension deficits has not been widely investigated, except primarily for a limited range of ambiguous sentences with garden path structures (Friedmann and Gvion, 2007; Novick et al., 2009; Vuong and Martin, 2015). Also, prior studies assessing the role of semantic STM in comprehension for brain damaged individuals have concentrated on comprehension of sentences with list-like structures, that is, with several prenominal adjectives preceding a noun (e.g., “rusty old red wagon”) or several nouns preceding a verb (e.g., “rugs, vases, and mirrors cracked during the move”). Tan et al. (2017) findings of a relation between category probe and semantic interference for healthy young adults opens up the possibility of investigating the relation of semantic STM to the comprehension of a wider range of sentence types for brain damaged PWA.

1.4. Current study

The present study examined the relation between PWA' STM and EF and their ability to resolve interference during comprehension. The sentences employed here were unambiguous sentences based on those from prior studies investigating the cue-based parsing approach (Van Dyke, 2007; see Table 1), which included a 2 × 2 manipulation of semantic and syntactic interference in which each sentence had a high or low interference value on semantic and syntactic interference. All sentences included an embedded clause. For the semantic interference manipulation, this embedded clause contained a noun that was either semantically plausible (e.g., “champion”) or implausible (e.g., “record”) as the subject of the main verb (e.g., “win”). Based on the results from previous studies (Lewis and Vasishth, 2005; Tan et al., 2017; Van Dyke, 2007), we expected that the semantically plausible nouns would cause greater processing difficulty because they partially matched the semantic retrieval cues, and thus reduce the amount of activation spread from the verb (e.g., “win”) to the real target (e.g., “jockey”).¹ For the syntactic interference manipulation, the embedded clause contained a noun that was either the subject or direct object of the embedded clause verb. Similarly, we expected that the syntactically plausible noun

¹ It should be noted that semantically plausible noun could also lead to less parsing difficulty if there were a full match between the distractor and the semantic retrieval cues, e.g., in the sentence as “The key to the cabinets are on the table”. Jäger et al. (2017) have conducted a Bayesian meta-analysis and shown that in the partially cue-matching condition, such as in our current experiment, the partially plausible word would generally increase sentence processing difficulty, while in the fully cue-matching condition, such as in Wagers et al. (2009), a facilitation effect is expected.

would cause greater processing difficulty because it partially matched the syntactic retrieval cues. As discussed earlier, prior research with healthy individuals has shown slower processing times and greater comprehension errors for the high interference conditions for both syntactic and semantic manipulations. Moreover, in these materials, the distance across which integrations were made was matched for the high and low interference conditions. The use of non-garden path sentences allows us to investigate whether interference resolution ability plays a role in structures where the conflict is more subtle than in difficult garden path sentences.

According to the cue-based parsing approach, PWA' STM capacity measures should be unrelated to their sensitivity to semantic and syntactic interference once language expertise (e.g., verbal knowledge) is taken into account. However, this approach would allow for EF abilities involved in interference resolution to play a role in resolving both semantic and syntactic interference. That is, those with better EF abilities should show smaller semantic and syntactic interference effects. In contrast, according to the multiple capacities approach (Martin and Romani, 1994; Martin and He, 2004), we predict that semantic capacity would be relevant to resolving semantic interference. That is, to the extent that semantic representations decay rapidly (Martin et al., 2006) or suffer from excessive interference (Hamilton and Martin, 2007) for those with reduced semantic capacity, we predict greater difficulty in resolving semantic interference for those with smaller semantic capacities. Also, to the extent that a semantic capacity is distinct from a syntactic capacity, we would predict that semantic STM capacity would not relate to the degree of syntactic interference. If semantic STM is distinct from semantic representations per se, this relation to the resolution of semantic interference should hold, even after controlling for semantic knowledge. In contrast, phonological STM capacity should not relate to either semantic or syntactic interference resolution abilities.

2. Methods

2.1. Participants

Nine PWA (six male; mean age = 64.1 years, $SD = 9.2$) were tested in the experiment. Participants were selected from the PWA subject pool at the Brain and Language Lab at Rice University on the basis of not having major difficulties with single word processing, including word comprehension, production and reading, and relatively preserved sentence comprehension ability (the screening tests are described in detail below). It should be noted that participants' ability to comprehend single words and simple sentences was controlled for in later analyses. In the beginning, we screened 25 PWA and 16 were eliminated due to severe deficits in the basic sentence comprehension test (i.e., scoring below 70% correct in sentence comprehension tasks consisting of active/passive transitive sentences and subject relative clauses), leaving 9 that completed the full battery of tasks. Demographic and lesion information of the nine PWA is provided in Table 2 (see Appendix D for aphasia classification and severity scores from clinical assessment and word production measures). The participants were all right-handed, native English speakers who had suffered a left hemisphere stroke, and had from 12 to 20 years of education (mean = 14.8 years, $SD = 2.8$). For seven out of the nine participants, hearing sensitivity was assessed with an audiometer (Dial and Martin, 2017), and they performed within the range for age-matched healthy control participants. The two who were unavailable for this testing, ML and DA, had been tested on a minimal pairs speech discrimination task and ML scored 96% correct (control range: 96–100%) and DA scored slightly below the control range (93% correct). Thus, all appeared to have adequate hearing to perceive speech input. They were diagnosed with aphasia by speech-language pathologists or as assessed by the Western Aphasia Battery (Kertesz, 1982) or the Boston Diagnostic Aphasia Exam (Goodglass et al., 2001). The overall aphasia quotient is shown in Appendix D. All the testing was carried out at least 18 months post

Table 2
PWA' biographical and clinical information.

Patient	Age	Education (years)	Gender	Years post-stroke	Lesion site
ML	70	14	M	21	L IFG, MFG, lateral parietal
EV	65	14	F	11	L frontal, MFG, occipital (minor)
NLA	54	14	M	2	L STG
SJ	63	13	F	5	L PPG
DA	57	19	F	1.5	L IFG
MB	64	13	M	5	L parietal lobe
SQ	69	14	M	11	L IFG, parietal, temporal
SH	85	12	M	6	L temporal, posterior parietal
QO	67	20	M	2	L STG, insular

Note: M = male; F = female; L = left.

L = left; IFG = inferior frontal gyrus; MFG = middle frontal gyrus; STG = superior temporal gyrus; PPG = posterior parietal gyrus.

stroke. Eight healthy age-matched (mean age = 70.3 years, $SD = 6.1$; $t(1, 15) = -1.13, p = .276$) and education level-matched² (mean years of education = 16.8, $SD = 3.0$; $t(1, 14) = -1.15, p = .269$) individuals from a participant pool maintained by the Brain and Language Laboratory were tested as a control group.

2.2. General procedure

All PWA were tested for the sentence comprehension task and several screening assessments in multiple sessions lasting about 1–1.5 h each, typically with one week separating sessions. It took approximately 8 sessions to complete testing each participant. Control participants were tested in four 1 h sessions for the sentence comprehension experiment, and were not tested on the screening assessments as it was assumed they would perform very well (i.e., 100% correct) in these tests.

2.3. Background language assessments

All PWA were assessed on single word and sentence processing to insure sufficient ability to carry out the sentence comprehension task. The screening criteria were specified for each task in the following section.

2.3.1. Single word processing

Psycholinguistic assessments of language processing in aphasia (PALPA): Two tests from the PALPA battery (Kay et al., 1992) were used in the current experiment. First, a visual lexical decision task that manipulated imageability and frequency (Set 25) was included. A list of 60 real words and 60 non-words were presented. Second, a word reading task was used. Participants were instructed to read aloud sets of 20 nouns, adjectives, verbs and functors matched on word frequency and length in order to verify that their word reading abilities were sufficient for processing the written sentences used in this study. Mean accuracy was calculated. Performance above 90% correct in each test was required.

Single picture-word matching (PWM): In this task (Martin et al., 1999), Participants saw a picture and had to answer the question that "Is this a ___?" Each word (e.g., "cat") was presented four times in different conditions, including a correct condition (e.g., a picture of cat), a semantically related condition (e.g., a picture of a dog), a phonological related condition (e.g. a picture of a hat), and an unrelated condition (e.g. a picture of nail clippers). Fifty-four words were tested, and the test was administered over four sessions, with each word only occurring

² One of the control participants' years of education was not available.

once per session. The mean accuracy in each condition and overall accuracy across all trials was calculated. Overall performance above 90% correct was required.

Peabody picture vocabulary test (PPVT): The Peabody Picture Vocabulary Test (Dunn and Dunn, 2007, 1981) test was used, in which participants heard a spoken word and had to point to the picture that matched it from four picture alternatives. Testing stopped at which point the participants made six errors over eight consecutive trials. Participants' raw score were converted into standard scores based on normative data for age-matched controls ($\mu = 100$ and $\sigma = 15$). Performance within 2 s.d. from the age-matched controls was required (i.e., score > 70).

2.3.2. Sentence processing

Sentence-picture matching task: This task was used to assess comprehension for reversible sentences (Martin and Feher, 1990). The experimenter read the sentences to the PWA and they selected from two pictures the one matching the sentence. PWA' accuracy was assessed on simple reversible active/passive transitive sentences ($n = 16$; e.g., "The boy pulled the girl" or "The ball was hit by the boy") and subject relative clause sentences ($n = 48$; e.g., "The boy who pulled the girl wore black shoes" or "The boy who wore black shoes pulled the girl"). For the simple active/passive sentence, the incorrect picture reversed the role of agent and theme (e.g., a picture displaying "The girl pulled the boy"). For the subject relative clause sentence, the incorrect picture either reversed the role of agent and theme around the action verb (e.g., "pulled") or reversed who was described by the descriptive phrase (e.g., "wore black shoes"). Subject relative clause structures were included because they appeared in the experimental materials whereas object relatives did not. Overall performance above 75% correct was required.

Quantitative production analysis: The QPA developed by Saffran and colleagues (Rochon et al., 2000; Saffran et al., 1989) was included to provide an objective measurement of grammatical aspects of production in PWA. This task was not used for screening but instead to provide an indication of PWA' sentence production abilities. In this task, PWA were asked to narrate a fairy tale, usually "Cinderella", while their production was recorded. If the Cinderella narrative yielded fewer than 150 narrative words, the PWA was asked to tell another well-known story. After transcription, PWA's performance was evaluated based on the first 150 (± 10) narrative words they produced. The number of narrative words produced per minute, and the proportion of closed class words and verbs are reported in Table 2. As shown in the table, the PWA did not exemplify typical agrammatic speech as all of them were within control range on proportion of closed class words and proportion verbs. However, four were below control range on proportion of well-formed sentences and most (6/9) had a median utterance length below the range of controls.

2.3.3. Background testing results

As shown in Table 3, all the PWA performed very well in the visual lexical and word reading tasks from PALPA. In addition, they had a standard score at or above 89 on the PPVT test and scored above 95% correct across conditions in the single picture-word matching test. Although EV scored 83% correct on the semantically related condition in the PWM test (whereas other PWA scored > 90%), which might indicate some semantic knowledge deficit, EV's overall accuracy was high on the picture-word matching task (95%) and she scored less than one standard deviation below the control mean on the PPVT. Regarding sentence processing ability, although some PWA had sentence production deficits as compared to age-matched controls as indexed by the QPA (e.g., 6 out of the 9 PWA had a reduced median utterance length score), all the PWA showed relatively preserved sentence comprehension ability as indexed by the accuracy of the sentence-picture matching task for simple structures (averaged across the active/passive sentence and subject relative clause sentences: mean = 89.6%, $SD = 10.2\%$, range = 71–100%).

2.4. Sentence comprehension task

2.4.1. Materials

This experiment used sentence materials similar to those in Van Dyke (2007) and Tan et al. (2017) sentence comprehension experiment with healthy younger subjects, which varied the degree of semantic interference (high vs. low) and syntactic interference (high vs. low). Eighty sets of experimental sentences were created with four different types of sentence in each set (see examples in Table 1 and full set of materials in Appendix A). The four sentences in each set were assigned to four different lists. Each subject received one full or half list in each testing session, depending on his or her rate of completing the task. Additionally, eighty sentences without relative clause structure were added to each list as fillers (e.g., "The maid cleaned the floors after the children made it dirty again"). In all, each participant saw 320 critical sentences (mean length = 15.7 words) and 320 filler sentences (mean length = 13.1 words) during the experiment.

2.4.2. Procedure

Participants silently read one sentence at a time at their own pace. All trials began with a fixation point appearing in the center of the screen for 500 ms, followed by a sentence. Participants pressed a button when they had finished reading the sentence. Sentence reading times were recorded as the time between sentence onset and button press. After each sentence, a comprehension question requiring a yes/no answer was presented and subjects responded by pressing a button corresponding to "yes" or "no." For the experimental sentences, the question probed the critical subject-verb integration (e.g., for the example sentences, "Will the jockey win?"). For the filler sentences, the comprehension questions probed other phrases in the sentence (e.g. for the filler example, "Did the children make the floor dirty again?"). In addition, there were an equal number of yes and no answers to the questions in each list. The correct answer to two-thirds ($n = 214$) of the critical sentences was "yes", while the correct answer to two-thirds ($n = 214$) of the filler sentences was "no". Question answering speed was recorded as the time between question onset and button press. After answering the question, participants (or the experimenter for some PWA) pressed a button to bring up the next sentence. Ten practice sentences with similar structures as the experimental sentences followed by comprehension questions were presented prior to each testing session to familiarize participants with the procedure.

2.5. Short-term memory tests

PWA' STM capacity was accessed via category probe recognition, synonymy judgment, rhyme probe recognition, digit span, and digit matching tasks. We included multiple tasks in order to compute a composite variable for each STM capacity, which provides more reliable and robust measure than a single measure for a given construct (Nunnally et al., 1967; Saffran and Martin, 1997). Based on the results from previous study (Allen et al., 2012), category probe recognition and synonymy judgment tasks were used to tap semantic STM capacity; rhyme probe recognition, digit span, and digit matching tasks were used to tap phonological STM capacity. Rather than dichotomously classifying the PWA into having either semantic or phonological STM deficits, we used these continuous measurements to reflect the degree of deficits in either or both types of capacity. Most of the PWA had varying degrees of deficits on both types of tasks. However, the variation in PWA' each STM capacity and the lack of correlation between PWA' phonological and semantic STM capacity (see detailed results in Table 6) allowed us to investigate the relationship between the magnitude of semantic and syntactic interference effect and different verbal STM capacities.

Category probe task: The category probe task (Martin et al., 1994) was included to tap PWA' semantic STM. In this task, participants were presented with an auditory word list followed by a probe word. They

Table 3
Language profiles of the PWA.

Tasks			ML	NLA	EV	SJ	DA	MB	SQ	SH	QO	Controls mean (Range)	
Word	Visual lexical decision (PALPA) (Accuracy)	Words and nonwords	n/a	1.0	.94	.99	n/a	.99	.92	.93	.99	.98	
	Reading (PALPA)	Grammatical Class	.96	1.0	.96	.98	1.0	.99	.90	.95	.97	99.6 (.36)	
	PPVT		107	91	89	94	89	120	97	115	111	100 (std dev. 15)	
	Single picture-word matching (Accuracy)	Semantically related		.96	.93	.83	.94	.91	.98	.96	.93	1.0	
		Phonologically related		1.0	.98	.96	.98	1.0	.98	1.0	.94	1.0	
		Unrelated		1.0	1.0	1.0	1.0	1.0	1.0	1.0	.96	1.0	
		Correct		1.0	.96	1.0	.96	1.0	.96	1.0	.96	.98	
Overall		.99	.97	.95	.97	.98	.98	.99	.95	1.0			
Sentence	Sentence-picture matching (Accuracy)	Active/passive and Subject relative clause sentence	.85	.76	.81	.75	.93	.95	.91	.91	.98		
	QPA	Words per minute	36	88	145	75	68	59	69	67	121	161 (107–232)	
		Prop. closed class words	.48	.52	.56	.52	.51	.59	.54	.58	.42	.54 (.47–.61)	
		Prop. verbs	.48	.45	.54	.47	.51	.49	.33	.48	.39	.48 (.35–.63)	
		Prop. Well-formed sentences	.61	.81	.95	.73	.53	.50	.93	.79	.86	.95 (.75–1.0)	
		Median utterance length	5.0	8.5	5.0	5.0	6.0	9.0	5.57	6.0	8.9	8.2 (6.5–10.5)	

Note: a) PWA' visual lexical decision performance was measured as the mean score on PALPA 25, which manipulated the frequency and imageability of words.

b) For PPVT, Mean = 100, SD = 15.

c) n/a = not applicable.

d) Control data on the QPA test were reported in [Rochon et al. \(2000\)](#) (N = 12). The proportion of closed class words was calculated by dividing the number of closed class words by the number of total narrative words. The proportion of verbs was calculated by dividing the number of verbs by the sum of verbs and nouns. The proportion of well-formed sentence was calculated by dividing the number of well-formed sentences by the total number of sentences produced.

e) The controls' data was not included on single picture-word matching and sentence-picture matching assessments as controls would perform at a ceiling level (e.g. 100% correct) in these tests.

had to make a Yes/No judgement on whether the probe word was in the same category as any of the words in the list by pressing corresponding button. Before testing, participants were shown a list of all nine categories (e.g. animals, clothing, fruits, etc.) that would be presented in the experiment as well as all the words belonging to each category. Each word in one list was drawn from different categories. The list length started from one-item, and there were 20 – 28 trials at each list length. Each word was presented at approximately 1 s per word with a 500 ms inter-stimulus interval. During the practice trials, the loudness of the stimuli was adjusted to a comfortable level for each participant. Testing stopped at which point the participants' accuracy dropped below 75% on a given list length. Span was calculated for each patient by using linear interpolation to find the list length at which they would score 75% accurate.

Synonymy judgment: In the synonymy judgment task ([Martin et al., 2006](#)), participants saw three written nouns presented on a diagonal on a page that were read aloud by the experimenter. They had to indicate which two of the three words were most similar in meaning by pointing to the correct pair. Half of the 48 trials consisted of three concrete words, while the other half consisted of three abstract words. All three words in each list were semantically related (e.g. *jelly, jam, honey*), and thus participants had to consider all three possible pairs of relations simultaneously to determine which two were most related (that is, no word could be easily ruled out). Thus, participants had to hold the relations between pairs of words in mind while determining which of the three possible pairs were most similar, putting demands on semantic STM. Participants' overall accuracy was calculated.

Rhyme probe task: The rhyme probe task ([Martin et al., 1994](#)) was included to measure PWA' phonological STM. Similar to the category probe task, participants were presented with an auditory word list followed by a probe word, but they had to judge whether the probe word rhymed with any of the words in the list by pressing a corresponding button. List length started from one-item, and there were 20 – 24 trials at each list length. This task was administrated and scored in the same fashion as the category probe task.

Digit span and digit matching span: Two standard memory span tasks (i.e., forward digit span and digit matching span tasks) were also included as the index of phonological STM capacity. In the forward digit span task from the WAIS-R ([Wechsler, 1997](#)), participants heard a list of

digits and were required to repeat the numbers aloud in order at the end of each list. The list length started from two-item, and there were 2 trials at each list length. Testing stopped at which point the participants missed both trials at a given list length. PWA' digit span was calculated as the longest list length they could complete (e.g. 3, if they repeated both trials on 3-item list correctly). Their span was a half number if they only repeated one trial correctly at the longest list length level (e.g. 2.5, if they only repeated one trial on 3-item list and missed both trials on 4-item list). In the digit matching task ([Martin et al., 1999](#)), participants heard two lists containing the same digits and were asked to judge whether the order of the items was the same or different across the two lists. In the non-matching condition, the second list reversed the order of two adjacent items in the first list. List length started at two-items, and there were 20 trials at each list length (10 matching and 10 non-matching trials). Testing stopped at the list length at which participants' accuracy dropped below 75% on a given list length, or at the six-item list length. Digit matching span was calculated for each PWA by using linear interpolation to find the list length at which they would score 75% accurate.

2.6. Executive function tests

Verbal Stroop task: The standard verbal Stroop task ([Stroop, 1935](#)) was used to measure PWA's resistance to automatic or prepotent responses. Participants were presented a series of words or strings of asterisks, and were required to name the ink color aloud as quickly and accurate as possible. There were three conditions, including a congruent condition (e.g., the word "blue" presented in blue ink; N = 12), an incongruent condition (e.g., the word "blue" presented in red ink; N = 65), and a neutral condition (e.g. a string of asterisks; N = 77). In order to compare to the results of [Tan et al. \(2017\)](#) study on healthy subjects, the number of trials in each condition was kept consistent. However, different from [Tan et al. \(2017\)](#), in which Stroop interference was calculated as response-time interference, for each participant, Stroop interference was calculated as the differences between error rates for incongruent and neutral trials.

Picture-word interference (PWI) task: In the picture-word interference task ([Lupker, 1979](#); [Schriefers et al., 1990](#)), a picture was presented in the center of the screen with a word superimposed in the

center of that picture. PWA was required to name the picture aloud while ignoring the word. There were two conditions - a related condition (i.e., the picture and the word were semantically related) and an unrelated condition. Each picture was presented in both conditions in different sessions. There were two sessions with 90 trials in each. Before the experiment, to make sure participants could name the picture correctly, they were presented with the experiment pictures with their correct name. For each participant, the PWI score was calculated as the mean error rate differences between related and unrelated trials.

2.7. Data analysis

The sentence comprehension experiment produced three dependent variables: sentence reading time (RT-sentence), question answering speed (RT-question), and accuracy for comprehension questions. For both RT analyses, only data from correctly answered trials were included. Outliers were calculated by condition for each subject and RTs greater than 2.5 standard deviations away from the mean were removed from the analyses. This trimming removed 2% of the data in PWA's sentence reading times and 3% of the data in their question answering times.

A logarithmic transformation was performed on the RT data in order to remove the effects of speed on the size of effects (Verhaeghen and De Meersman, 1998). Some studies on healthy populations have found that variation in processing speed could account for the correlations between working memory capacity and performance on complex cognitive tasks to a large extent (Fry and Hale, 1996; Salthouse et al., 1998), and some researchers have argued that PWA's sentence comprehension deficits are due to slowed processing (Haarmann and Kolk, 1991; Kolk, 1995; Swinney and Zurif, 1995). Using the logarithmic transformation removed the effect of overall speed in calculating differences between RTs in the high and low interference conditions.

For the STM, EF, and semantic knowledge tasks, some data from patient ML, EV, SJ, MB, and SH (Allen et al., 2012; Hamilton and Martin, 2007; Vuong and Martin, 2011), and all the control data had been reported in previous studies from Brain and Language Lab (e.g., Martin and Romani, 1994; Martin et al., 1994). The STM scores of patient ML, NLA, SJ, and SH were based on STM testing conducted during the sentence comprehension experiment sessions. The scores of patient EV, SI, MB, DA, and QO, whose category and rhyme probe spans were not tested in conjunction with the sentence comprehension sessions, were based on the average scores of two closest tests conducted before and after the sentence comprehension experiment.

3. Results

3.1. Sentence comprehension task

We first examined sentence comprehension performance for controls, followed by an assessment of PWA's performance relative to controls. We wished to verify that the semantic and syntactic interference effects demonstrated with healthy young adults in prior studies would be replicated in the current study for our older control participants and to determine the extent to which PWA demonstrated effects similar to or outside the range of controls. The mean performance of control participants and individual PWA data are shown in Table 4. The magnitude of semantic and syntactic interference effects was calculated as the difference in accuracy and RT between high semantic/syntactic interference and low semantic/syntactic interference conditions, respectively. Internal reliability for semantic and syntactic interference effect in accuracy, speed of question answering, and sentence reading times was calculated as split-half correlation adjusted with the Spearman-Brown prophecy formula (Cronbach, 1951). The reliability for all dependent measures was high ($r_s > .87$).

3.1.1. Control participants

Participants' accuracy and reaction times for comprehension questions were modelled separately in linear mixed-effects models using R (3.1.1) implemented within the *lme4* packages, with a logistic linking function for accuracy data (Baayen et al., 2008; Baayen and Milin, 2010; Jaeger, 2008). Each of the independent variables was mean-centered prior to analysis. The semantic and syntactic interference was coded in an ANOVA-style with -1 for low interference condition and 1 for high interference condition. Thus, negative coefficients for each main effect of log RT and positive coefficients for each main effect of accuracy signify worse performance (i.e., longer RT and lower accuracy) in the high interference conditions. In the mixed-effect model, fixed effects included the main effect of semantic interference, the main effect of syntactic interference, and the interaction of semantic and syntactic interference. In addition, sentence length (number of words) was included as a control factor in all the models, and question length was also included in the models with question RT as dependent variables. We initially started with the maximal random-effects structures (Barr et al., 2013) which included both by-participant and by-item random intercepts and random slopes for the main effects and interaction of semantic and syntactic interference effects, but simplified the models if they failed to converge through iteratively removing the term that explained the least variance.

3.1.1.1. Comprehension questions. The control participants maintained a high level of accuracy on comprehension questions in all conditions (mean = 96%, SD = 1%, range = 94–98%). The mixed-effects model with a maximal random effects structure only converged for the question answering times. For accuracy, the model converged when there were just by-subject and by-item adjustment to the intercepts. The modelling results on accuracy data might be considered anti-conservative (Barr et al., 2013).³ An analysis of the accuracy data revealed a significant main effect of syntactic interference (LI: 98%, HI: 94%; $\beta = -.41510$, $SE = .13400$, $z = -3.097$, $p = .002$), a significant main effect of semantic interference (LI: 97%, HI: 95%; $\beta = -.27150$, $SE = .13320$, $z = -2.038$, $p = .042$) as well as an interaction between semantic and syntactic interference effects ($\beta = -.36360$, $SE = .13340$, $z = -2.725$, $p = .006$). Further comparisons demonstrated a syntactic interference effect only in the high-semantic interference condition (LI: 98%, HI: 92%; $\beta = -.76450$, $SE = .17710$, $z = -4.317$, $p < .001$), but not in the low-semantic interference condition (LI: 98%, HI: 96%; $\beta = -.07496$, $SE = .22857$, $z = -.328$, $p = .743$). No significant effects were observed for controls for question RTs ($p_s > .35$).

3.1.1.2. Sentence reading times. Analysis of the log-transformed sentence reading times via mixed-effects model showed that control participants read the sentences more slowly in the high syntactic (HI: 6.25 s, LI: 5.55 s; $\beta = .02250$, $SE = .00519$, $t = 4.336$, $p < .001$) and high semantic interference conditions (HI: 6.08 s, LI: 5.72 s; $\beta = .01252$, $SE = .00421$, $t = 2.974$, $p = .011$). The interaction of semantic and syntactic interference was not significant in sentence reading times ($\beta = .00276$, $SE = .00351$, $t = .787$, $p = .433$).

3.1.1.3. Control summary. In sum, these control participants demonstrated similar effects to those reported earlier for healthy young subjects (Tan et al., 2017; Van Dyke, 2007). They showed both

³ Regarding the anti-conservative concern of the mixed-effects model on accuracy data, we compared the results from the mixed-effects model and from the 2 (high vs. low semantic interference) \times 2 (high vs. low syntactic interference) factorial repeated measures ANOVAs using error terms based on participant (F_1) and on item (F_2) on all the dependent variables (i.e., question accuracy, question RT, sentence RT), because Barr et al. (2013) have suggested that ANOVA is a better method than the random-intercepts-only mixed-effects model. The pattern of control participants' results was the same (see ANOVA results in Appendix C).

Table 4
Performance of PWA and controls on sentence comprehension task.

		ML	NLA	EV	SJ	DA	MB	SQ	SH	QO	Control Mean (SD) (N = 8)	Control (Range) (N = 8)
Accuracy	LoSyn/LoSem	.76	.93	.85	.94	.89	.90	.80	.96	.95	.98 (.02)	.95–1.0
	LoSyn/HiSem	.60	.88	.83	.93	.86	.91	.67	.93	.93	.98 (.02)	.95–1.0
	HiSyn/LoSem	.78	.95	.92	.95	.88	.86	.64	.94	.93	.96 (.03)	.91–1.0
	HiSyn/HiSem	.57	.85	.60	.85	.82	.87	.55	.96	.83	.92 (.02)	.89–.95
	Sem Interf effect	.19	.08	.17	.05	.05	-.01	.11	.00	.06	.02 (.02)	(-.01)–.05
	Syn Interf effect	.00	.01	.08	.04	.03	.04	.14	.00	.06	.04 (.02)	.0–.07
Question RT (ms)	LoSyn/LoSem	4025	3075	3972	3048	3135	1788	3540	4343	2082	1689 (468)	1259–2732
	LoSyn/HiSem	4179	3037	4583	3350	3701	1862	3748	4064	2299	1705(450)	1304–2688
	HiSyn/LoSem	3167	3478	4135	3933	3723	1805	3695	4441	2082	1738 (476)	1282–2809
	HiSyn/HiSem	4740	3379	5857	3663	3426	1883	3725	4669	2123	1746 (413)	1293–2585
	Sem Interf effect (log)	.08	-.01	.09	-.01	.01	.01	.01	.00	.02	0 (.02)	(-.03)–.03
	Syn Interf effect (log)	-.03	.04	.07	.05	.02	.00	.01	.03	-.02	.01 (.02)	(-.02–.04)
Sentence RT (ms)	LoSyn/LoSem	20,303	21,377	10,618	15,694	20,093	9318	11,022	2,1478	10,636	5436 (1645)	3733–8668
	LoSyn/HiSem	19,592	25,265	11,199	15,947	19,054	10,710	12,330	22,188	11,943	5672 (1797)	3754–8895
	HiSyn/LoSem	20,777	25,335	12,065	15,601	21,287	10,984	11,018	24,804	11,497	6013 (2044)	3758–9765
	HiSyn/HiSem	20,059	27,235	11,651	16,341	20,626	10,796	10,721	25,434	12,490	6484 (2274)	4432–10825
	Sem Interf effect (log)	-.02	.05	.00	.01	-.01	.02	.01	.01	.04	.02 (.02)	0–.06
	Syn Interf effect (log)	.00	.05	.04	.00	.03	.03	-.02	.05	.02	.04 (.02)	.02–.07

Note: a) The comprehension question accuracy was calculated as the unweighted averaged accuracy of “yes”-answer questions and “no”-answer questions.
 b) “Sem Interf” refers to semantic interference; “Syn Interf” refers to syntactic interference. Interference effect was calculated by subtracting subjects’ mean RT in the low interference conditions from that in the high interference conditions; while for accuracy data, interference effect was calculated in the opposite way.
 c) The interference effects were calculated from log transformed RTs. For illustration purposes only, the table used raw RTs for subjects’ performance in each condition, despite the fact that all analyses were conducted using log-transformed RTs.

Table 5
Analyses for PWA’ sentence comprehension performance.

DV	Measure	Effect	Mixed-effects model			
			β	SE	z/t	p
Question	Accuracy	Sem	-.35037	.05770	-6.072	.000**
		Syn	-.17097	.06058	-2.822	.005*
		Sem × Syn	-.10285	.05759	-1.786	.074
	RT (log)	Sem	.01180	.00625	1.888	.095
		Syn	.00733	.00553	1.325	.206
		Sem × Syn	.00032	.00581	.056	.957
Sentence	RT (log)	Sem	.00882	.00400	2.207	.049*
		Syn	.01178	.00488	2.415	.028*
		Sem × Syn	-.00667	.00474	-1.407	.165

Note. “-Syn” and “-Sem” refer to syntactic interference and semantic interference condition.

* p < .05.

** p < .01.

semantic and syntactic interference effects in question answering accuracy and in sentence reading times. In question accuracy, they showed an interaction between semantic and syntactic interference, with greater syntactic interference in the high than low semantic interference condition. Such an interaction has been reported in some (Van Dyke, 2007) but not all (Glaser et al., 2013; Segaert et al., 2012; Tan et al., 2017) prior studies with young subjects. However, in the current study, the interaction is hard to interpret given the near ceiling accuracy in the low syntactic interference conditions.

3.1.2. PWA

3.1.2.1. Comprehension questions. The PWA showed a large range of overall comprehension performance (mean = 84%, SD = 10%, range = 67–95%). Binomial tests were used to compare each PWA’s performance to chance (50%). For overall accuracy across the four conditions, all the PWA performed significantly above chance (ps < .001). For accuracy in each condition, all PWA were above chance for all conditions except for SQ whose performance did not differ from chance for the most difficult condition (HiSem/HiSyn condition: mean accuracy = 55%, p = .06). However, age-matched

control participants scored nearly at ceiling overall (mean = 97%) and all PWA fell below the range of controls’ performance in at least one condition.

The mixed-effects model with a maximal random effects structure only converged for the sentence RTs. For the question answering times, the model converged after omitting by-item adjustments to the intercept, while for accuracy, the model converged when there were just by-subject adjustment to the intercepts. As mentioned above, the modeling results on accuracy data might be considered anti-conservative (Barr et al., 2013), although it was overall consistent with what we obtained from ANOVA analysis.⁴ Results of PWA from the mixed-effects models are shown in Table 5. For question accuracy, there were significant effects of both semantic (HI: 80%, LI: 88%), and syntactic (HI: 82%, LI: 86%) interference with subjects being less accurate in the high

⁴ The PWA results from mixed-effects model and ANOVA analysis were consistent, except that semantic interference effect in the sentence RT was marginally significant in the by-subject analysis in the ANOVA analysis, $F_1(1, 8) = 4.62, p = .06, F_2(1, 79) = 6.38, p = .01$. (See ANOVA results in Appendix C).

than the low semantic and syntactic interference conditions. Analysis of question answering speed showed a marginal main effect of semantic interference (HI: 3.57 s, LI: 3.30 s), while neither syntactic interference (HI: 3.55 s, LI: 3.32 s) nor the interaction of semantic and syntactic interference reached significance. The results for PWA overall were similar to those for controls in showing robust semantic and syntactic interference effects in accuracy.

Individual patient data were analysed to determine whether each PWA's semantic and syntactic interference effects were significantly different than those for controls using the revised Standardized Difference Test (RSDT; two-tailed tests; see Appendix B for detailed results) (Crawford and Garthwaite, 2002, 2005; Crawford and Howell, 1998).

In the accuracy data, a considerable range of effects was obtained for both types of interference (semantic: $-.01$ to $.18$; syntactic: 0 – $.14$). 4/9 PWA showed effects that were significantly greater than those for controls for semantic interference ($ps < .05$; ML, NLA, EV, and SQ) whereas only 1/9 showed effects greater than controls for syntactic interference (SQ, $p = .002$). For question reaction time, 2/9 showed effects significantly greater than those for controls for semantic interference ($ps < .001$; ML & EV) and 1/9 showed an effect greater than controls for syntactic interference (EV, $p = .025$), when using log RT.

3.1.2.2. Sentence reading times. For sentence RTs, there was large variation in means (overall mean = 16.60 s, SD = 5.69 s, range = 10.45 – 24.80 s). Generally, all the PWA read much more slowly than controls (control mean = 5.89 s, SD = 1.93 s, range = 3.92 – 9.53 s) when combining across all four conditions. As shown in Table 5, PWA showed the same overall pattern as the controls. They were slower in the high semantic interference (HI: 16.87 s, LI: 16.32 s), and high syntactic interference conditions (HI: 17.15 s, LI: 16.04 s) as compared to the low semantic/syntactic interference conditions, respectively. The interaction between semantic and syntactic interference manipulations was not significant.

In examining individual patient effects using the RSDT on log latency RTs, none of the PWA showed either semantic or syntactic interference effects that were greater than controls. Thus, the long overall RTs did not relate to larger interference effects in sentence reading times. However, when looking at the relation between overall sentence reading times, which is the averaged log RT collapsing over the four conditions, and the magnitude of semantic and syntactic interference effects in accuracy, an interesting pattern emerged. Mean log RT for sentence reading was negatively related to the size of the syntactic interference effect in accuracy ($r = -.79$, $p = .01$) but unrelated to semantic interference effect in accuracy ($r = -.09$, $p = .82$). This suggests that PWA' reading time was devoted more to deriving an appropriate syntactic structure and less to resolving semantic interference, perhaps because subjects were less aware of an error when a semantically plausible but incorrect integration was made in the high semantic interference condition.

3.1.2.3. PWA summary. As a group, the PWA showed a similar pattern to controls in that they demonstrated semantic and syntactic interference effects in question answering accuracy and sentence reading times. However, all PWA' accuracy was below the range of controls in at least one condition and many showed semantic and syntactic interference effects that were significantly greater than those for controls. While all PWA also read much more slowly than controls, their effects in log sentence reading time generally did not differ significantly from controls.

The greater interference effects for the PWA on question accuracy indicate that several had difficulty resolving interference despite the fact that these sentences were lexically and syntactically unambiguous. Thus, the findings go beyond prior demonstrations of PWA's difficulty in comprehending garden path sentences (Friedmann and Gvion, 2007; Novick et al., 2009; Vuong and Martin, 2015), showing that they have

difficulty resolving interference even when there is no need to override a strong tendency to interpret a sentence incorrectly based on lexical or syntactic biases. The next section deals with whether these difficulties in resolving interference relate to PWA's working memory or general executive functions.

3.2. Relating sentence processing to STM and executive function

3.2.1. PWA performance on STM and executive function tasks

PWA's performance on STM and EF tests is reported in Table 6. Composite scores were calculated by averaging z -scores across tasks tapping each ability to improve the reliability of the measures (Nunnally et al., 1967). These composites consisted of: 1) phonological STM capacity: rhyme probe, digit span, and digit matching span, 2) semantic STM: category probe and synonymy judgment tasks, and 3) executive function (i.e., inhibition): Stroop and picture-word interference tasks. In addition, to control for PWA' semantic knowledge, a composite variable was computed from averaging z -scores for the PPVT and single picture word matching (mean accuracy in the semantically related condition) tasks (see Table 3 for PWA's performance on each task).

STM tasks: As shown in Table 6, all PWA had different degrees of deficits in STM capacities, and varied in terms of the relative deficit for semantic and phonological STM. For example, PWA' category probe span ranged from 1.5 to 3.4 items and their rhyme probe span ranged from 1.3 to 4.3 items, with all scores being at or below the bottom of the range for age-matched controls' span (category probe: overall mean = 5.4 items, range = 3.4–7 items; rhyme probe: overall mean = 7 items, range = 5.8–9 items). Moreover, as can be seen by comparing the category and rhyme probe spans of various PWA, some PWA had relatively better phonological STM capacity as compared to their semantic STM capacity (e.g. EV, DA), while some others showed the opposite pattern (e.g. SH).

Unfortunately, patient DA had a second stroke before we could test her on the synonymy judgment task. In order to include DA's data in the investigation of the relation between sentence comprehension and STM capacity, we estimated DA's performance in the synonymy judgment task through imputation based on her performance in the category probe span task. That is, based on a regression using the other eight PWA' performance on category probe span and synonymy judgment task DA's accuracy on synonymy judgments was predicted to be 91%.

EF tasks: Table 6 presents PWA' performance on the two EF tasks. Generally, most PWA performed within the range of age-matched controls in the verbal Stroop task (all except EV and SQ) and the PWI task (all except SQ), although they showed some variation (for Stroop, SD = .13; for PWI, SD = .05). In addition, PWA' composite score for lexical knowledge, which was included as a control variable in the correlational analysis, is presented in Table 6 as well.

3.2.2. Relations between individual differences measures

The correlations among the different composite measures are shown in Table 7. All of the variables going into the composite measures have been widely used in previous studies and have been demonstrated to have good validity and reliability (Allen et al., 2012; Martin et al., 2006; Pettigrew and Martin, 2014; Segaert et al., 2012).

Most critically, with respect to the correlations among the IDs measures, as shown in Table 7, the correlation between the phonological and semantic STM composites was close to zero ($r = -.08$, $p = .85$). The absence of a correlation between these two variables is consistent with previous findings that semantic and phonological retention capacities are separable (Hamilton et al., 2009; Martin and Romani, 1994; Martin et al., 1994). Such a difference between the two STM capacities as well as the variability in each STM measures (as shown in Table 6) enabled us to investigate the memory-language relation by relating sentence comprehension performance to the continuum of different STM capacities.

Table 6
PWA performance on STM, EF, and semantic knowledge tasks.

	Semantic STM			Phonological STM			EF (Inhibition)			Semantic knowledge ^b Composite semantic knowledge (PPVT + PWM)	
	Category probe	Synonymy judgment	Sem. STM Composite	Rhyme probe	Digit span	Digit matching span	Pho. STM Composite	Stroop (% error)	PWI (% error)		EF Composite
ML	1.5	.75	-3.07	1.3	3.5	3.6	-3.11	.06	-.03	-1.26	0.92
NLA	1.5	.98	-0.51	2.3	3.5	4.7	-1.08	.10	.00	-.40	-1.03
EV	1.9	.81	-1.78	3.7	5	6	2.87	.36	.03	2.12	-3.24
DA	2.1	(.91) ^a	(-0.29) ^a	4.0	6	6	4.10	-.01	.05	-.33	-1.61
SJ	2.2	1.0	0.85	2.7	3.5	3.7	-1.77	.08	.01	-.37	-0.58
MB	2.5	.96	0.78	4.3	3.5	3.3	-0.61	.05	-.03	-1.33	2.41
SQ	2.6	.85	-0.26	3.0	5.5	5.2	1.87	.24	.15	3.40	0.08
SH	3.0	.98	1.82	2.0	3	4	-2.54	-.08	.00	-1.77	0.97
QO	3.4	.98	2.47	4.2	3.5	4.3	0.28	.05	.04	-.05	2.07
Controls^c	5.4	.95		7.0	5.7	6		.06 ^f	.03 ^f		
Range	3.4 – 7 ^c	(SD = .05) ^d		5.8 – 9 ^c	3–7.5 ^c	5.8 – 6 ^d		SD = .06 ^f	SD = .03 ^f		

Note:
^a As discussed in the paper, patient DA was not tested on the synonymy test and her performance in this task was estimated based on a regression imputation.
^b The composite semantic knowledge measure was computed from PPVT and PWI (semantic related condition only) tasks, which data were presented in Table 1.
^c Control data were reported in Allen and Martin (2012).
^d Control data were reported in N. Martin et al. (2006).
^e Control data were reported in Martin, Shelton, & Yaffee (1994).
^f Control data were reported in Vuong and Martin (2011).

Table 7
Pearson correlation coefficients for all individual differences measures.

	2	3	4	5	6	7	I	II	III	IV
1. Category probe	.49	.43	-.14	-.17	-.31	.31	.86**	.05	.002	.51
2. Synonymy judgment		.27	-.42	-.30	-.60	-.05	.86**	-.18	-.38	.34
3. Rhyme probe			.41	.33	.28	.14	.41	.70*	.25	-.03
4. Digit span				.85**	.65	.50	-.32	.91**	.68*	-.55
5. Digit matching span					.70*	.37	-.27	.88**	.63	-.78*
6. Stroop						.43	-.53	.66	.85**	-.67*
7. PWI							.15	.41	.85*	-.15
I. Sem. STM composite								-.08	-.22	.49
II. Pho. STM composite									.63*	-.55
III. EF composite										-.48
IV. Sem. knowledge composite										

Note: a) * $p < .05$; ** $p < .01$; *** $p < .001$. For the reliability test, it marks the values significant different from zero by within-subjects t -test.
 b) n/a = not applicable.

Table 8
Correlations between individual differences measures and semantic and syntactic interference resolution ability in question answering performance.

DV	Effect	Semantic STM	Phonological STM	EF
Accuracy	Semantic	-.817* [-.966, -.114]	.100 [-.728, .856]	.482 [-.566, .965]
	Syntactic	-.007 [-.661, .713]	.568 [-.191, .918]	.920* [.451, .987]
RTs	Semantic	-.770 [-.921, .374]	.202 [-.602, .882]	.243 [-.544, .896]
	Syntactic	-.037 [-.913, .824]	.310 [-.573, .787]	.264 [-.379, .839]

Note: * $p < .05$ in the Pearson product-moment tests. The values in the [] are the 95% confidence interval of the correlation using a bootstrap technique.

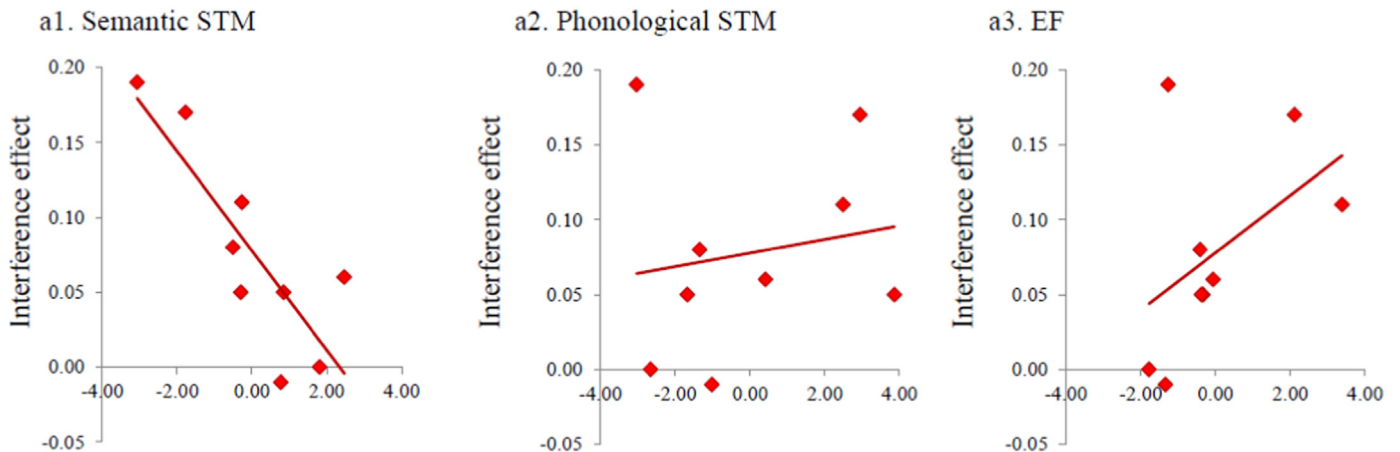
3.2.3. Relating sentence processing effects to STM and EF

We examined the relations between semantic/syntactic interference resolution ability during sentence comprehension and the semantic/phonological STM and EF using Pearson product-moment correlations. Given the very long times for sentence reading and the fact that log latency effects for semantic and syntactic interference in sentence reading time for individual PWA were generally not significantly greater than those for controls, we did not assess correlations for sentence reading times. Correlational tests were thus conducted on question answering accuracy and response time for semantic and syntactic interference with the three individual differences measures (i.e., semantic STM, phonological STM, and EF). To treat our results with

greater caution given our small sample size, we reported 95% confidence interval around the correlation coefficient with a non-parametric bootstrap method (Efron, 1988; Efron and Tibshirani, 1993; Lunneborg, 1985) using R (3.1.1) implemented within the *boot* packages ($n = 10,000$ bootstrapped samples). The bootstrap technique makes no assumption about the distribution of the data, and provides a valid estimation of the CI for small sample size (Chernick, 2008). We reported the CIs from bias-corrected and accelerated bootstrap (BCA) method for the correlations between comprehension question performance and other cognitive abilities.

The correlational results are presented in Table 8 (see also Fig. 1). With respect to the STM composites, the semantic composite had a

A. Semantic interference



B. Syntactic interference

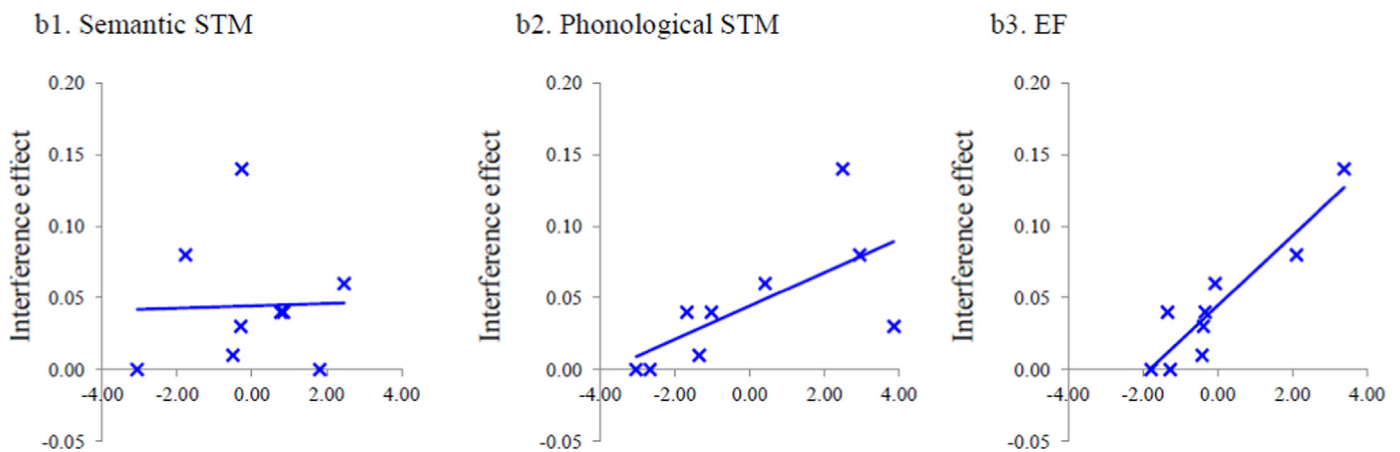


Fig. 1. Scatter plots of the correlations between semantic (A)/syntactic (B) interference effects and semantic STM (1) and phonological STM (2) and EF (3) in question answering accuracy. The correlations between interference effects and individual differences were not presented as none of them was significant.

strong and significant negative correlation with semantic interference in accuracy, $r = -.817$, CI $[-.966, -.114]$, but no significant relation to syntactic interference, $r = -.007$, CI $[-.661, .713]$, whereas the phonological STM composite did not correlate with either type of interference (semantic interference, $r = -.100$, CI $[-.728, .856]$; syntactic interference $r = .568$, CI $[-.191, .918]$). The EF composite had a strong and significant correlation with the syntactic interference effect ($r_2 = .920$, CI $[.451, .987]$) but not with the semantic interference effect ($r_2 = -.482$, CI $[-.566, .965]$). There was no significant correlation between any individual differences measures and either semantic or syntactic interference effect in question RT.

Given that there was some degree of correlation (though non-significant) between the semantic knowledge composite and both the semantic STM composite ($r = .492$, CI $[-.391, .921]$) and the EF composite ($r = .482$, CI $[-.970, .049]$), we wished to determine whether the two significant correlations reported above would remain so when semantic knowledge was partialled out. Therefore, we constructed two regression models to further test these two relations with semantic knowledge composite included as a control variable in each model and used a bootstrap technique to assess the significance of the beta-weights (Freedman, 1981; Fox, 2002; Jones and Waller, 2013). The semantic STM composite remained a significant predictor for the semantic interference effect size, $B = -.03$, $[-.05, -.02]$, and the EF composite

remained a significant predictor for the syntactic interference effect size, $B = .03$, $[.02, .03]$. Another potentially confounding factor is that generally poor syntactic abilities might have led PWA' to rely on executive function abilities in order to handle the high syntactic interference condition. To address this issue, we controlled for PWA' performance in comprehending active/passive and subject relative clause sentences from the relation between EF and syntactic interference. However, the beta-weight remained significant as well, $B = .03$, $[.02, .03]$. Therefore, these results demonstrated that during question answering, smaller semantic STM capacity was associated with larger semantic interference effects and lesser executive function ability was associated with larger syntactic interference effects. These correlations could not be simply accounted for by PWA' deficits in semantic knowledge or their overall syntactic abilities.

3.3. Summary of main results

Overall, the result for these older control subjects and the PWA replicated prior findings with healthy young adults (e.g., Glaser et al., 2013; Tan et al., 2017; Van Dyke, 2007), showing syntactic and semantic interference effects in the processing of grammatically unambiguous sentences. With respect to the major questions of interest regarding the relation of these interference effects to STM and EF, PWA

with more severe semantic STM deficits showed greater deficits in semantic but not syntactic interference resolution while answering comprehension question as shown in the accuracy data, even after controlling for their semantic knowledge. The lack of relation between phonological STM and semantic and syntactic interference effects indicated that phonological retention ability did not play a critical role in interference resolution during sentence comprehension. Additionally, our investigation into the EF-language relation demonstrated that PWA with more severe deficits in executive functions had more difficulty in resolving syntactic interference effects during sentence comprehension. The lack of any significant correlation in the question RT might be due to the less variance in the RT relative to accuracy data. Most of the PWA showed comparable magnitude of semantic and syntactic interference as the controls in their log-transformed question RT.

4. Discussion

In the present study, we explored an important and controversial topic in psycholinguistic research: what is the role of verbal STM and EF mechanisms in sentence processing? Through testing PWA with varying degrees of STM or EF deficits on sentences with semantic and syntactic interference, first, we found that many of the PWA showed exaggerated semantic and syntactic interference effects, while answering comprehension questions. These results are consistent with some neuropsychological findings from serial recall or item recognition tasks showing that when multiple competing representations are active in memory, PWA may show more difficulty than controls in settling on the appropriate one (Barde et al., 2010; Hamilton and Martin, 2005, 2007; Martin and Lesch, 1996). More importantly, as summarized above, our results demonstrated a specific relation between semantic STM and the resolution of semantic interference and between EF and the resolution of syntactic interference. To some degree, these findings replicate those from Tan et al. (2017) for healthy young subjects as that study also found a specific relation between semantic STM capacity and semantic interference resolution in question answering. However, in that study, the Stroop effect, which was the only measure of EF, failed to correlate with any measure of interference resolution whereas a complex WM span measure was related to syntactic interference resolution. The relation of the current findings to those of Tan et al. (2017) and the implications of both for theories of sentence processing are elaborated below.

4.1. Relations of verbal STM to sentence comprehension

The finding of a specific relation between semantic STM and semantic interference effects, but no relation between semantic STM and syntactic interference effects nor between phonological STM and either semantic or syntactic interference effects, is in line with previous studies of both PWA and healthy young subjects - that is, the smaller the semantic STM span, the more difficulty in resolving semantic interference (Barde et al., 2010; Hamilton and Martin, 2007; Kush et al., 2015; Segaert et al., 2012) or in maintaining semantic representations during sentence processing (Martin and He, 2004; Martin and Romani, 1994; Martin et al., 1994). These results are most consistent with the multiple capacities model proposed by Martin et al. (1994, 1999) and Martin (2005). They cannot be explained by a domain-general WM account, which assumes a unitary WM capacity underlying all types of language processing (Fedorenko et al., 2006, 2007; Just and Carpenter, 1992; King and Just, 1991). That is, domain-general WM capacity accounts cannot explain the lack of correlation between phonological STM and interference resolution, nor the lack of relation between semantic STM and syntactic interference resolution.

How might the multiple capacities approach account for the current findings? This model suggests that it is the connections between linguistic representations stored in LTM and STM buffers that keep this information activated (Martin, 2005). PWA with semantic STM deficits

have difficulty in maintaining a precise semantic representation of the correct target information, due either to decay (Saffran and Martin, 1997) or to interference between multiple representations (Hamilton and Martin, 2007). Therefore, when cues generated at the verb are used to access these memory representations, the degree of match of cue-target and cue-distractors may be closer for these degraded representations than would be the case for controls with more precise representations. Hence, deciding between possibilities will take longer. On the other hand, given that the cues generated at the verb have no relation to phonology (Kush et al., 2015) and phonological features have no bearing on determining semantic and syntactic fit to the retrieval cues, no relation to phonological STM would be expected. The assumption of discrimination difficulty between semantic representations for PWA with semantic STM deficits is consistent with finding from Barde et al. (2010). They found on a probe recognition task a specific relation between semantic STM deficits and difficulty discriminating current list items from semantically related prior list items and a relation between phonological STM deficits and difficulty discriminating current items from phonologically related prior list items.

A critical finding in the current experiment is that even after controlling for PWA' semantic verbal knowledge, measured through the PPVT and picture-word matching tasks, the specific relation between semantic STM and semantic interference persisted. In the literature on verbal STM deficits, some have argued that reduced semantic capacity derives from a specific deficit in maintaining semantic features (Martin et al., 1994; Vallar et al., 1997) whereas others have argued that these result from damage to long-term lexical representations per se (Dell et al., 1997). In line with the latter view, some researchers taking a cue-based parsing approach of language processing have argued that the observed correlation between WM capacity and individual differences in interference resolution ability during sentence comprehension derives from variation in the strength of long-term knowledge representation in different domains (e.g. semantic, syntactic) – due to linguistic experience or genetic factors (Van Dyke et al., 2014). These same factors contribute to performance on WM/STM tasks, resulting in the correlation between the two. However, in the current study, although there was a moderate correlation between PWA' semantic knowledge and semantic STM ($r = .49, p = .18$), PWA' semantic STM composite score was a significant predictor for the magnitude of their semantic interference effect even after controlling for their semantic knowledge. These results provide strong support for the multiple capacities approach. However, it should be noted that we used participants' semantic knowledge as an approximation for their linguistic experience because previous studies have shown a correlation between these two factors (Acheson et al., 2008; Van Dyke et al., 2014). However, it is possible that other components of linguistic experience (e.g., grammatical knowledge) could predict individuals' sentence processing performance. This requires future investigation.

4.2. Relations of EF to sentence comprehension

The significant positive relation between PWA' EF and their ability to resolve syntactic interference and the non-significant relation to semantic interference requires some explanation. According to the multiple capacities view (Martin and Romani, 1994; Tan et al., 2017), there is a separable capacity for maintaining syntactic information (see also, Caplan and Waters, 1999, 2013; Friedmann and Gvion, 2003). The results for the semantic STM measure are consistent with this view as the measure correlated with semantic but not syntactic interference effects. However, we did not include a measure of syntactic capacity as it is difficult to construct one that is independent of syntactic knowledge. The specific relation reported here is to EF ability rather than to a syntactic capacity. Our finding of a relation to EF is consistent with several studies on individuals with Parkinson's disease which have also demonstrated a correlation between impairment in syntactic processing and their inability to inhibit irrelevant information (Colman et al.,

2011; Hochstadt, 2009; Hochstadt et al., 2006). Our finding is also consistent to some extent with prior studies showing that those with EF deficits have difficulty resolving incompatible relationships during language processing (Novick et al., 2005; Thothathiri et al., 2012; Vuong and Martin, 2011, 2015). However, for these later studies, it is not the case that the findings were restricted to resolving incompatible syntactic representations. For example, Vuong and Martin (2011) found that PWA with EF deficits had difficulty using sentence context to access the less frequent meaning of a biased ambiguous noun. On the basis of these findings, one might have expected EF abilities to relate to both semantic and syntactic processing.

One possible reason for the lack of relation of EF abilities to semantic interference is that the PWA was less aware of an error when an incorrect but semantically plausible noun was selected as the subject. This supposition is in line with our finding of a correlation between overall sentence reading times and accuracy effects for syntactic interference but not for semantic interference. Thus, PWA may have been less motivated to resolve conflict on the semantically interfering trials as the conflict was not detected. It should be acknowledged that in the present study there was a moderate correlation between EF and semantic interference in question answering ($r = .39$, $p = .306$); however, this was far from significance with this small sample size. It is possible that with a larger sample size, and perhaps therefore with a larger subset of PWA who were more aware of the semantic conflict, a relation to EF would be found for semantic interference as well.

As noted in the introduction, the study of Tan et al. (2017) with healthy young subjects did not find a relation of the Stroop effect to either semantic or syntactic interference. However, as discussed earlier, this may have resulted from the fact that the Stroop effect for these subjects was in terms of RTs rather than accuracy, given their ceiling effects on accuracy as prior results suggest that RT effects reflect goal maintenance whereas accuracy effects reflect conflict resolution (Kane and Engle, 2003). Tan et al. (2017) did find a relation between a complex WM measure and syntactic interference during sentence processing and semantic interference during question answering. Given the many findings implicating EF abilities in complex WM span performance (Kane and Engle, 2002; Unsworth et al., 2014), those results are also consistent with a role for EF in interference resolution.

4.3. Conclusions and future directions

The current study demonstrated semantic and syntactic interference effects during sentence comprehension for both PWA and controls, as predicted by the cue-based parsing approach. For the PWA, several showed effects outside the range of controls, particularly for semantic interference. Importantly, our results provide strong evidence consistent with the multiple capacities approach of language processing, which emphasizes the separable retention capacities of verbal STM for phonological, semantic, and maybe syntactic information. We found that PWA with more severe semantic STM deficits had more difficulty in resolving semantic interference during sentence comprehension, even after controlling for their semantic knowledge deficits. However, phonological STM was unrelated to neither semantic nor syntactic interference resolution during sentence processing. Additionally, the observation of a link between EF and the syntactic interference effect supports recent claims that EF plays an important role in language processing.

Our finding of a relation between semantic STM and semantic interference resolution even after controlling for language expertise contradicts predictions from the cue-based parsing approach that variation in sentence parsing and interpretation should be attributed to verbal knowledge or EF ability but not STM capacity. To accommodate these finding with an embedded processes approach to WM/STM, one would have to assume that activated information outside the focus of attention decays at a differential rate for different types of information (e.g., semantic, syntactic, phonological). It is not entirely clear,

however, how such an approach could be distinguished from one postulating modality specific STM buffers, though perhaps one could test the assumption of unlimited capacity at very short delays that is inherent in the embedded processes account. In any event, the current findings suggest that both executive function and modality specific retention ability, at least for semantic information, are critical for sentence processing.

Acknowledgements

This work was partially supported by the Lodieska Stockbridge Vaughn Fellowship at Rice University. The authors are grateful to Julie Van Dyke for help with creating the sentence materials and thoughtful discussions of these results.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.neuropsychologia.2018.03.001>.

References

- Acheson, D.J., Wells, J.B., MacDonald, M.C., 2008. New and updated tests of print exposure and reading abilities in college students. *Behav. Res. Methods* 40, 278–289.
- Alexander, M.P., 2006. Impairments of procedures for implementing complex language are due to disruption of frontal attention processes. *J. Int. Neuropsychol. Soc.* 12, 236–247.
- Allen, C.M., Martin, R.C., Martin, N., 2012. Relations between short-term memory deficits semantic processing and executive function. *Aphasiology* 26, 428–461.
- Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278.
- Baayen, R.H., Davidson, D.J., Bates, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412.
- Baayen, R., Milin, P., 2010. Analyzing reaction times. *Int. J. Psychol. Res.* 3, 12–28.
- Barde, L.H.F., Schwartz, M.F., Chrysikou, E.G., Thompson-Schill, S.L., 2010. Reduced short-term memory span in aphasia and susceptibility to interference: contribution of material-specific maintenance deficits. *Neuropsychologia* 48, 909–920.
- Chernick, M.R., 2008. *Bootstrap Methods: A Guide for Practitioners and Researchers*, 2nd ed. Wiley, Hoboken, NJ.
- Caplan, D., Michaud, J., Hufford, R., 2013. Short-term memory WM and syntactic comprehension in aphasia. *Cogn. Neuropsychol.* 30, 77–109.
- Caplan, D., Waters, G.S., 1999. Verbal WM and sentence comprehension. *Behav. Brain Sci.* 22, 77–94.
- Caplan, D., Waters, G., 2013. Memory mechanisms supporting syntactic comprehension. *Psychon. Bull. Rev.* 20, 243–268.
- Colman, K.S., Koerts, J., Stowe, L.A., Leenders, K.L., Bastiaanse, R., 2011. Sentence comprehension and its association with executive functions in PWA with Parkinson's disease. *Park. Dis.*
- Cowan, N., 1999. An embedded-processes model of working memory. *Models Work. Mem. Mech. Act. Maint. Exec. Control* 62–101.
- Crawford, J.R., Garthwaite, P.H., 2002. Investigation of the single case in neuropsychology: confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia* 40, 1196–1208.
- Crawford, J.R., Garthwaite, P.H., 2005. Testing for suspected impairments and dissociations in single-case studies in neuropsychology: evaluation of alternatives using Monte Carlo Simulations and revised tests for dissociations. *Neuropsychology* 19, 318–331.
- Crawford, J.R., Howell, D.C., 1998. Comparing an individual's test score against norms derived from small samples. *Clin. Neuropsychol.* 12, 482–486.
- Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334.
- Daneman, M., Carpenter, P., 1980. Individual differences in working memory and reading. *J. Verbal Learn. Verbal Behav.* 19, 450–466.
- Daneman, M., Merikle, P.M., 1996. Working memory and language comprehension: a meta-analysis. *Psychon. Bull. Rev.* 3, 422–433.
- Dell, G.S., Schwartz, M.F., Martin, N., Saffran, E.M., Gagnon, D.A., 1997. Lexical access in aphasic and nonaphasic speakers. *Psychol. Rev.* 104, 801–838.
- Dial, H., Martin, R., 2017. Evaluating the relationship between sublexical and lexical processing in speech perception: evidence from aphasia. *Neuropsychologia* 96, 192–212.
- Dickey, M.W., Choy, J.J., Thompson, C.K., 2007. Real-time comprehension of wh-movement in aphasia: evidence from eyetracking while listening. *Brain Lang.* 100, 1–22.
- Dickey, M.W., Thompson, C.K., 2009. Automatic processing of wh-and NP-movement in agrammatic aphasia: evidence from eyetracking. *J. Neurolinguist.* 22, 563–583.
- Dunn, L.M., Dunn, D.M., 2007. *Peabody Picture Vocabulary Test*, 4 ed. Pearson, San Antonio, Texas.
- Dunn, L., Dunn, L., 1981. *Peabody Picture Vocabulary Test - Revised*. Guidance Service,

- Circle Pines, MN.
- Efron, B., 1988. Bootstrap confidence intervals: good or bad? *Psychol. Bull.* 104, 293–296.
- Efron, B., Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*. Chapman & Hall, Boca Raton, FL.
- Engelhardt, P.E., Nigg, J.T., Ferreira, F., 2017. Executive function and intelligence in the resolution of temporary syntactic ambiguity: an individual differences investigation. *Q. J. Exp. Psychol.* 70, 1263–1281.
- Evans, W.S., Caplan, D., Ostrowski, A., Michaud, J., Guarino, A.J., Waters, G., 2015. Working memory and the revision of syntactic and discourse ambiguities. *Can. J. Exp. Psychol.* 69, 136–155.
- Fedorenko, E., Gibson, E., Rohde, D., 2006. The nature of working memory capacity in sentence comprehension: evidence against domain-specific working memory resources. *J. Mem. Lang.* 54, 541–553.
- Fedorenko, E., Gibson, E., Rohde, D., 2007. The nature of working memory in linguistic, arithmetic and spatial integration processes. *J. Mem. Lang.* 56, 246–269.
- Fox, J., 2002. *An R and S-Plus Companion to Applied Regression*. Sage, Thousand Oaks, CA.
- Frankel, T., Penn, C., Ormond-Brown, D., 2007. Executive dysfunction as an explanatory basis for conversation symptoms of aphasia: a pilot study. *Aphasiology* 21, 814–828.
- Freedman, D.A., 1981. Bootstrapping regression models. *Ann. Stat.* 9, 1218–1228.
- Friedmann, N., Gvion, A., 2003. Sentence comprehension and working memory limitation in aphasia: a dissociation between semantic-syntactic and phonological reactivation. *Brain Lang.* 86, 23–39.
- Friedmann, N., Gvion, A., 2007. As far as individuals with conduction aphasia understood these sentences were ungrammatical: garden path in conduction aphasia. *Aphasiology* 21, 570–586.
- Fry, A.F., Hale, S., 1996. Processing speed, working memory and fluid intelligence: evidence for a developmental cascade. *Psychol. Sci.* 7, 237–241.
- Gibson, E.A., 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68, 1–76.
- Gibson, E.A., 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In: In: Miyashita, Y., Mirantz, A., O'Neil, W. (Eds.), *Image, language, brain* MIT Press, Cambridge, MA, pp. 95–126.
- Glaser, Y.G., Martin, R.C., Van Dyke, J.A., Hamilton, A.C., Tan, Y., 2013. Neural basis of semantic and syntactic interference in sentence comprehension. *Brain Lang.* 126, 314–326.
- Goodglass, H., Kaplan, E., Barresi, B., 2001. *The Assessment of Aphasia and Related Disorders*, 3 ed. Lippincott Williams & Wilkins, Philadelphia.
- Haarmann, H.J., Kolk, H.H.J., 1991. Syntactic priming in Broca's aphasics: evidence for slow activation. *Aphasiology* 5, 247–263.
- Hamilton, A.C., Martin, R.C., 2005. Dissociations among tasks involving inhibition: a single-case study. *Cogn. Affect. Behav. Neurosci.* 5, 1–13.
- Hamilton, A.C., Martin, R.C., 2007. Proactive interference in a semantic short-term memory deficit: role of semantic and phonological relatedness. *Cortex* 43, 112–123.
- Hamilton, A.C., Martin, R.C., Burton, P.C., 2009. Converging functional magnetic resonance imaging evidence for a role of the left inferior frontal lobe in semantic retention during language comprehension. *Cogn. Neuropsychol.* 26, 685–704.
- Hanten, G., Martin, R.C., 2000. Contributions of phonological and semantic STM to sentence processing: evidence from two cases of closed head injury in children. *J. Mem. Lang.* 43, 335–361.
- Hochstadt, J., 2009. Set-shifting and the on-line processing of relative clauses in Parkinson's disease: results from a novel eye-tracking method. *Cortex* 45, 991–1011.
- Hochstadt, J., Nakano, H., Lieberman, P., Friedman, J., 2006. The roles of sequencing and verbal working memory in sentence comprehension deficits in Parkinson's disease. *Brain Lang.* 97, 243–257.
- Hsu, N.S., Novick, J.M., 2016. Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychol. Sci.* 27, 572–582.
- Jäger, L.A., Engelmann, F., Vasishth, S., 2017. Similarity-based interference in sentence comprehension: literature review and Bayesian meta-analysis. *J. Mem. Lang.* 94, 316–339.
- Jaeger, T.F., 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446.
- Jones, J.A., Waller, N.G., 2013. Computing confidence intervals for standardized regression coefficients. *Psychol. Methods* 18, 435–453.
- Johns, C.L., Gordon, P.C., Long, D.L., Swaab, T.Y., 2014. Memory availability and referential access. *Lang. Cogn. Neurosci.* 29, 60–87.
- Just, M.A., Carpenter, P.A., 1992. A capacity theory of comprehension: individual differences in working memory. *Psychol. Rev.* 99, 122–149.
- Kane, M.J., Engle, R.W., 2002. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychon. Bull. Rev.* 9, 637–671.
- Kane, M.J., Engle, R.W., 2003. Working-memory capacity and the control of attention: the contributions of goal neglect response competition and task set to Stroop interference. *J. Exp. Psychol. Gen.* 132, 47–70.
- Kay, J., Coltheart, M., Coltheart, M., Lesser, R., 1992. *Psycholinguistic assessment of language processing in aphasia*. Lawrence Erlbaum Associates, Hove, UK.
- Kertesz, A., 1982. *Western Aphasia Battery*. Psychological Corp, San Antonio, TX.
- King, J., Just, M.A., 1991. Individual differences in syntactic processing: the role of working memory. *J. Mem. Lang.* 30, 580–602.
- Kolk, H., 1995. A time-based approach to agrammatic production. *Brain Lang.* 50, 282–303.
- Kush, D., Johns, C.L., Van Dyke, J.A., 2015. Identifying the role of phonology in sentence-level reading. *J. Mem. Lang.* 79, 18–29.
- Lau, E., Stroud, C., Plesch, S., Phillips, C., 2006. The role of structural prediction in rapid syntactic analysis. *Brain Lang.* 98, 74–88.
- Lewis, R.L., 1996. Interference in short-term memory: the magical number two (or three) in sentence processing. *J. Psycholinguist. Res.* 25, 93–115.
- Lewis, R.L., Vasishth, S., 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cogn. Sci.* 29, 375–419.
- Lunneborg, C.E., 1985. Estimating the correlation coefficient: the bootstrap approach. *Psychol. Bull.* 98, 209–215.
- Lupker, S.J., 1979. The semantic nature of response competition in the picture-word interference task. *Mem. Cogn.* 7, 485–495.
- MacDonald, M.C., Christiansen, M.H., 2002. Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychol. Rev.* 109, 35–54.
- Martin, N., Schwartz, M.F., Kohen, F.P., 2006. Assessment of the ability to process semantic and phonological aspects of words in aphasia: a multi-measurement approach. *Aphasiology* 20, 154–166.
- Martin, R.C., 2005. Components of short-term memory and their relation to language processing evidence from neuropsychology and neuroimaging. *Curr. Dir. Psychol. Sci.* 14, 204–208.
- Martin, R.C., Feher, E., 1990. The consequences of reduced memory span for the comprehension of semantic versus syntactic information. *Brain Lang.* 38, 1–20.
- Martin, R.C., He, T., 2004. Semantic short-term memory and its role in sentence processing: a replication. *Brain Lang.* 89, 76–82.
- Martin, R.C., Lesch, M.F., 1996. Associations and dissociations between language impairment and list recall: implications for models of short-term memory. In: Gathercole, S.E. (Ed.), *Models of Short-term Memory*. Psychology Press, East Sussex, UK, pp. 149–178.
- Martin, R.C., Lesch, M.F., Bartha, M.C., 1999. Independence of input and output phonology in word processing and short-term memory. *J. Mem. Lang.* 41, 3–29.
- Martin, R.C., Romani, C., 1994. Verbal working memory and sentence comprehension: a multiple-components view. *Neuropsychology* 8, 506–523.
- Martin, R.C., Shelton, J.R., Yaffee, L.S., 1994. Language processing and working memory: neuropsychological evidence for separate phonological and semantic capacities. *J. Mem. Lang.* 33, 83–111.
- McElree, B., 2000. Sentence comprehension is mediated by content-addressable memory structures. *J. Psycholinguist. Res.* 29, 111–123.
- McElree, B., 2001. Working memory and focal attention. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 817–835.
- McElree, B., 2006. Accessing recent events. *Psychol. Learn. Motiv.* 46, 155–200.
- McElree, B., 2015. Memory processes underlying real-time language comprehension. In: Lindsay, D.S., Kelley, C.M., Yonelinas, A.P., Roedinger, H.L.L.I.I. (Eds.), *Remembering: Attributions, Processes, and Control in Human Memory: Essays in honor of Larry Jacoby*. Psychology Press, New York, NY, pp. 133–152.
- McElree, B., Foraker, S., Dyer, L., 2003. Memory structures that subserve sentence comprehension. *J. Mem. Lang.* 48, 67–91.
- Miyake, A., Carpenter, P.A., Just, M.A., 1994. A capacity approach to syntactic comprehension disorders: making normal adults perform like aphasic patients. *Cogn. Neuropsychol.* 11, 671–717.
- Miyake, A., Carpenter, P.A., Just, M.A., 1995. Reduced resources and specific impairments in normal and aphasic sentence comprehension. *Cogn. Neuropsychol.* 12, 651–679.
- Misyak, J.B., Christiansen, M.H., 2012. Statistical learning and language: an individual differences study. *Lang. Learn.* 62, 302–331.
- Nicholas, M., Connor, L.T., 2017. People with aphasia using AAC: are executive functions important? *Aphasiology* 31, 819–836.
- Novick, J.M., Kan, I.P., Trueswell, J.C., Thompson-Schill, S.L., 2009. A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychol.* 26, 527–567.
- Novick, J.M., Hussey, E., Teubner-Rhodes, S., Harbison, J.I., Bunting, M.F., 2014. Clearing the garden-path: improving sentence processing through cognitive control training. *Lang. Cogn. Neurosci.* 29, 186–217.
- Novick, J.M., Trueswell, J.C., Thompson-Schill, S.L., 2005. Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. *Cogn. Affect. Behav. Neurosci.* 5, 263–281.
- Nunnally, J.C., Bernstein, I.H., Berge, J.M.F. t., 1967. *Psychometric Theory*. 226 McGraw-Hill, New York.
- Oberauer, K., 2002. Access to information in working memory: exploring the focus of attention. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 411–421.
- Penn, C., Frankel, T., Watermeyer, J., Russell, N., 2010. Executive function and conversational strategies in bilingual aphasia. *Aphasiology* 24, 288–308.
- Perfetti, C.A., 2007. Reading ability: lexical quality to comprehension. *Sci. Stud. Read.* 11, 357–383.
- Pettigrew, C., Hillis, A.E., 2014. Role for memory capacity in sentence comprehension: evidence from acute stroke. *Aphasiology* 28, 1258–1280.
- Pettigrew, C., Martin, R.C., 2014. Cognitive declines in healthy aging: evidence from multiple aspects of interference resolution. *Psychol. Aging* 29, 187–204.
- Reali, F., Christiansen, M.H., 2007. Processing of relative clauses is made easier by frequency of occurrence. *J. Mem. Lang.* 57, 1–23.
- Rochon, E., Saffran, E.M., Berndt, R.S., Schwartz, M.F., 2000. Quantitative analysis of aphasic sentence production: further development and new data. *Brain Lang.* 72, 193–218.
- Saffran, E.M., Berndt, R.S., Schwartz, M.F., 1989. The quantitative analysis of agrammatic production: procedure and data. *Brain Lang.* 37, 440–479.
- Saffran, E.M., Martin, N., 1997. Language and auditory-verbal short-term memory impairments: evidence for common underlying processes. *Cogn. Neuropsychol.* 14, 641–682.
- Salthouse, T.A., Fristoe, N., McGuthry, K.E., Hambrick, D.Z., 1998. Relation of task switching to speed age and fluid intelligence. *Psychol. Aging* 13, 445–461.
- Segaert, K., Menenti, L., Weber, K., Petersson, K.M., Hagoort, P., 2012. Shared syntax in

- language production and language comprehension—an fMRI study. *Cereb. Cortex* 22, 1662–1670.
- Schriefers, H., Meyer, A.S., Levelt, W.J., 1990. Exploring the time course of lexical access in language production: picture-word interference studies. *J. Mem. Lang.* 29, 86–102.
- Stroop, J.R., 1935. Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662.
- Sung, J.E., McNeil, M.R., Pratt, S.R., Dickey, M.W., Hula, W.D., Szuminsky, N.J., Doyle, P.J., 2009. Verbal working memory and its relationship to sentence-level reading and listening comprehension in persons with aphasia. *Aphasiology* 23, 1040–1052.
- Swinney, D., Zurif, E., 1995. Syntactic processing in aphasia. *Brain Lang.* 50, 225–239.
- Tan, Y., Martin, R.C., Van Dyke, J.A., 2017. Semantic and syntactic interference in sentence comprehension: a comparison of working memory models. *Front. Psychol.* 8, 198. <http://dx.doi.org/10.3389/fpsyg.2017.00198>.
- Thothathiri, M., Kim, A., Trueswell, J.C., Thompson-Schill, S.L., 2012. Parametric effects of syntactic–semantic conflict in Broca's area during sentence processing. *Brain Lang.* 120, 259–264.
- Thothathiri, M., Rattinger, M., Trivedi, B., 2017. Cognitive control during sentence generation. *Cogn. Neurosci.* 8, 39–49.
- Traxler, M.J., Morris, R.K., Seely, R.E., 2002. Processing subject and object relative clauses: evidence from eye movements. *J. Mem. Lang.* 47, 69–90.
- Unsworth, N., Fukuda, K., Awh, E., Vogel, E.K., 2014. Working memory and fluid intelligence: capacity attention control and secondary memory retrieval. *Cogn. Psychol.* 71, 1–26.
- Unsworth, N., Heitz, R.P., Schrock, J.C., Engle, R.W., 2005. An automated version of the operation span task. *Behav. Res. Methods* 37, 498–505.
- Vallar, G., Di Betta, A.M., Silveri, M.C., 1997. The phonological short-term store-rehearsal system: patterns of impairment and neural correlates. *Neuropsychologia* 35, 795–812.
- Van Dyke, J., 2007. Interference effects from grammatically unavailable constituents during sentence processing. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 407–430.
- Van Dyke, J., Johns, C.L., 2012. Memory interference as a determinant of language comprehension. *Lang. Linguist. Compass* 6, 193–211.
- Van Dyke, J., Johns, C.L., Kukona, A., 2014. Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition* 131, 373–403.
- Van Dyke, J., Lewis, R.L., 2003. Distinguishing effects of structure and decay on attachment and repair: a cue-based parsing account of recovery from misanalyzed ambiguities. *J. Mem. Lang.* 49, 285–316.
- Van Dyke, J., McElree, B., 2006. Retrieval interference in sentence comprehension. *J. Mem. Lang.* 55, 157–166.
- Van Dyke, J., McElree, B., 2011. Cue-dependent interference in comprehension. *J. Mem. Lang.* 65, 247–263.
- Verhaeghen, P., De Meersman, L., 1998. Aging and the negative priming effect: a meta-analysis. *Psychol. Aging* 13, 435–444.
- Vuong, L.C., Martin, R.C., 2011. LIFG-based attentional control and the resolution of lexical ambiguities in sentence context. *Brain Lang.* 116, 22–32.
- Vuong, L.C., Martin, R.C., 2015. The role of LIFG-based executive control in sentence comprehension. *Cogn. Neuropsychol.* 32, 243–265.
- Wagers, M.W., Lau, E.F., Phillips, C., 2009. Agreement attraction in comprehension: representations and processes. *J. Mem. Lang.* 61, 206–237.
- Wechsler, D., 1997. WAIS-III: Wechsler Adult Intelligence Scale. Psychological Corporation, San Antonio, TX.
- Wells, J.B., Christiansen, M.H., Race, D.S., Acheson, D.J., MacDonald, M.C., 2009. Experience and sentence processing: statistical learning and relative clause comprehension. *Cogn. Psychol.* 58, 250–271.
- Wright, H.H., Downey, R.A., Gravier, M., Love, T., Shapiro, L.P., 2007. Processing distinct linguistic information types in working memory in aphasia. *Aphasiology* 21, 802–813.