Published in final edited form as:

Cortex. 2019 March ; 112: 58–68. doi:10.1016/j.cortex.2018.11.017.

Independent Contributions of Semantic and Phonological Working Memory to Spontaneous Speech in Acute Stroke

Randi C. Martin¹ and Tatiana T. Schnur²

¹Rice University

²Baylor College of Medicine

Abstract

Patients with left hemisphere stroke often have language deficits which impair their ability to produce phrases and sentences. One possible source of these speech impairments is the disruption of verbal working memory (WM). Single-case studies of chronic stroke have suggested the existence of a WM capacity specific to maintaining semantic information that is critical for preparing multiple words in phrases prior to speech onset (Martin & Freedman, 2001; Martin & He, 2004; Martin, Miller, & Vu, 2004; Freedman, Martin, & Biegler, 2004). The current study tested this hypothesis by examining spontaneous narrative language production and working memory capacities in a large sample of individuals at the acute stage of stroke (N=36), prior to the reorganization of function or strategy development. Here we show using a multiple regression approach that patients' semantic but not phonological WM capacity had an independent contribution in predicting phrasal elaboration and increasing utterance length whereas patients' phonological but not semantic WM capacity had an independent contribution in predicting a more rapid speech rate. Importantly, neither WM capacity independently predicted grammatical abilities in speech, implying that the other relations did not result from overall severity. These results indicate that separable semantic and phonological WM components exist that support different aspects of narrative speech. To our knowledge, this is the first study to examine spontaneous speech in a large group of acute stroke patients demonstrating a critical relationship between working memory and the ability to produce more words in phrases and longer utterances.

Keywords

language production; working memory; acute stroke; narrative

Corresponding author: Randi C. Martin, Psychology Dept.- MS25, PO. Box 1892; Rice University, Houston, Texas 77251, rmartin@rice.edu, Tatiana T. Schnur, Dept. of Neurosurgery, Baylor College of Medicine, 1 Baylor Plaza, Houston, TX, 77030, tschnur@gmail.com.

Conflict of Interest

The authors declare no conflict of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Models of working memory (WM) for verbal information typically emphasize the retention of phonological (i.e., speech sound) information (e.g., Baddeley, 2000; Gupta & Tisdale, 2009; Page and Norris, 1998). Some evidence, however, suggests a separate capacity for the retention of semantic information (Shivde & Anderson, 2011; N. Martin & Saffran, 1997; Nishiyama, 2013), which plays a more critical role than the phonological capacity in language comprehension (Haarmann, Davelaar, & Usher, 2003; Tan, Martin, & Van Dyke, 2017) and production (Martin & Freedman, 2001; Martin et al., 2004). Some of the most compelling evidence on language production has come from case studies of individuals with specific disruptions of this semantic capacity following stroke and their associated speech production deficits. However, only a few case studies of individuals at the chronic stage of stroke have been reported and, moreover, these studies focused on the elicitation of simple phrases and sentences in constrained laboratory tasks (e.g., producing “short hair” to describe a picture; Martin & Freedman, 2001; Martin, Miller, & Vu, 2004; Freedman et al., 2004). The present study sought to investigate whether these claims regarding the role of semantic vs. phonological WM capacities would extend to more naturalistic speech involved in narrative language production (i.e., story telling) for a larger sample (N=36) at the acute stage of stroke (most within 72 hrs. of stroke). Testing at the acute stage allows one to observe language behavior prior to the reorganization of brain function and prior to the development of patient strategies for task performance.

In order to address theories regarding the WM resources in language production, we obtained quantitatively defined features of narrative language production using the Quantitative Production Analysis methodology (QPA; Rochon, Saffran, Berndt, & Schwartz, 2000) and related these features to phonological and semantic WM capacities. Addressing these issues in brain damaged patients has advantages given the wide range of narrative production and WM abilities in this group, providing a sensitive test of any relation. Relatively few studies of language production abilities have been carried out at the acute stage of stroke (within 3–5 days of stroke, e.g., Deleon et al., 2007; Hillis et al., 2006) and most of these have examined production at the single word level. The few studies which have examined fluency during language production in the acute or sub-acute phase of stroke used global measures of production derived from standard aphasia batteries (e.g., category or phonemic fluency or subjective assessments of narrative fluency; Corbetta et al., 2015; Forkel et al. 2014; Geranmayeh, Chau, Wise, Leech, & Hampshire, 2017; Kummerer et al., 2013). Thus, the present study represents the first investigation of narrative language production at the acute stage where precise, quantitative measures of production have been obtained and where these measures have been used to test theories of the role of WM in language production.

During sentence production, speakers carry out some degree of advance planning of the structure and words in an upcoming utterance prior to the onset of articulation (e.g., Allum & Wheeldon, 2007, 2009; Bock & Levelt, 1994). The degree to which this planning draws on WM depends on the scope of planning - that is, the number of content (e.g., nouns, verbs, adjectives) or grammatical words (e.g., pronouns, determiners, conjunctions) planned in advance of articulation. Models of language production typically assume a first stage of

planning at the message level, involving a conceptual, non-linguistic representation of the ideas to be expressed. The next stage consists of planning at both a lexical-semantic (i.e., word meaning) level and a grammatical level, involving the selection of words to express concepts and the choice of a syntactic structure for ordering the lexical items (Chang, Dell, & Bock, 2006; Allum & Wheeldon, 2007, 2009). The subsequent stage involves phonological planning where phonological forms are retrieved and joined together. Prior to this stage, words are represented in terms of their semantic and syntactic features. The final stage involves articulation of the phonological representations. The standard assumption is that an entire clause is planned at the conceptual level, whereas smaller units (i.e., phrases or words) may be planned at the subsequent levels (Garrett, 1975, 1980). Planning is assumed to be carried out in a cascaded fashion such that planning at a subsequent level can begin even though planning at an earlier level is not complete (Bock & Levelt, 1994). Furthermore, while a portion of an utterance is being articulated, planning can proceed for subsequent sentence elements at earlier levels of representation. To the extent that several words are planned simultaneously at the levels of semantic or phonological encoding, one would assume that this planning depends on WM resources to keep the word representations active until articulation can begin. If semantic and phonological capacities are independent, planning at the lexical-semantic level could draw on the semantic capacity whereas planning at the phonological level could draw on the phonological capacity. Given the separation between the stages involving lexical-semantic representations and phonological representations, it is possible that a different scope of planning applies to each.

Proposals regarding the scope of planning at the lexical-semantic/grammatical encoding stages range from the single word (e.g., Griffin & Bock, 2000) to the phrase (e.g., Allum & Wheeldon, 2007, 2009) to the entire clause (e.g., Garrett, 1975, 1980; Meyer, 1996). Prior case studies of four chronic stroke patients on constrained production tasks (Freedman et al., 2004; Martin & Freedman, 2001; Martin et al., 2004) provided evidence supporting a phrasal scope of planning at a lexical-semantic level. These studies contrasted individuals who had deficits in semantic vs. phonological WM (Hanten & Martin, 2000; Martin, Shelton & Yaffee, 1994; Martin & He, 2004). For instance, those with phonological WM deficits failed to show standard phonological effects on word list recall (i.e., phonological similarity and word length effects for visually presented words), whereas those with semantic WM deficits did show such effects. In contrast, those with phonological WM deficits showed the standard advantage of word over nonword recall, whereas those with semantic WM deficits did not, suggesting that they did not benefit from the semantic information in the words (Martin & He, 2004). In addition, patients with the two types of deficits showed contrasting patterns on two tasks designed to emphasize semantic (category probe task) or phonological retention (rhyme probe) (Martin et al., 1994; Martin & He, 2004). On the category probe task, individuals heard a list of words followed by a probe word and judged whether the probe word was in the same semantic category as one of the list words. For the rhyme probe task, they judged whether the probe word rhymed with any of the list words. The patients with the semantic WM deficit did better on the rhyme probe task than the patients with the phonological WM deficit whereas the reverse was the case for those with the phonological WM deficit. On tasks involving the production of phrases or sentences to describe pictures, patients with the semantic WM deficit struggled to produce utterances containing more than

one content word in a phrase (e.g., “short blonde hair” or “the ball and the block moved above the faucet”). For example, for the target utterance “small leaf,” patient AB, with a semantic WM deficit said, “It’s a leaf. It’s small. I can’t get it” (Martin & Freedman, 2001). Another patient with a semantic WM deficit, ML, had an utterance onset latency over 1000 ms longer for sentences beginning with a conjoined noun phrase (e.g., “the ball and block moved above the faucet”) than for matched sentences beginning with a single noun (e.g., “the ball moved above the block and the faucet”) whereas control subjects showed a much smaller effect (56 ms; Martin et al., 2004). In contrast, those with a phonological WM deficit showed a normal level of performance and patterns of effects of the number of content words on latencies. The patients with the semantic WM deficit did better in producing sentences (e.g., “The hair is blonde”) than phrases (e.g., “the blonde hair”) to describe the same picture, which we attributed to the fact that the sentences had fewer words per phrase (Martin & Freedman, 2001). Thus, the conclusion was that planning occurs for an entire phrase at a semantic level. We argued that this semantic capacity was specific to lexical representations, rather than to domain-general conceptual representations. If the capacity were conceptual, then one would not have expected the advantage for sentences over phrases, as presumably the same conceptual representation would be involved in producing both types of utterances. Instead, the capacity appeared specific to holding word meanings within a phrase. Restrictions in this capacity cause difficulties in producing utterances with more than one content word per phrase. The good performance of the patients with a phonological WM deficit suggested that the scope of planning at the phonological level is smaller, perhaps encompassing a single word and accompanying grammatical words (Wheeldon & Lahiri, 2002).

The current study examined whether these findings observed across four single case studies in constrained empirical tasks would extend to spontaneous language production in a relatively large group of individuals (N=36). Rather than focusing on specific individuals showing the most striking differences between phonological and semantic WM deficits, the present study took a case series approach (Schwartz & Dell, 2010), in which continuous variation on the two WM capacities was used to predict the degree to which subjects are able to elaborate their speech by using larger numbers of words in each phrase when producing sentences (sentence elaboration). Such an approach allowed us to consider a wider range of individuals. Also, examining more spontaneous speech allowed us to determine whether the patterns uncovered in constrained tasks, with a limited range of targeted phrase types (i.e., adjective-noun phrases and conjoined noun phrases), extend to speech contexts more like those encountered in everyday life where individuals are free to select phrase and sentence structures to convey their intended meaning.

To elicit narrative language production, patients were asked to tell the Cinderella story after being reminded of the story by viewing a picture book with the words concealed. Their narratives were scored according to the QPA system (Rochon, et al., 2000), which provides a detailed analysis of types of words used (e.g., content vs. grammatical) and structural aspects of narratives (e.g., the proportion of sentences that were grammatically well formed). The present study focused on two language production measures where we expected WM to have specific influence, sentence elaboration and mean length of utterance. Sentence elaboration is based only on utterances (primarily defined by pauses) which qualify as

sentences in this scoring system - that is, utterances with a subject and main verb (whether these utterances are grammatical or not) - and consists of the mean number of content words plus pronouns in the subject and main verb phrases excluding the subject and main verb.

¹Table 1 provides examples of patients' sentence utterances, highlighting the content words plus pronouns in the subject phrase and the main verb phrase. Mean length of utterance is computed across all types of utterances and is simply the mean of the number of words per utterance.

We generated the following predictions concerning the relationships between WM capacities and abilities to produce multiword speech. If there is a separate semantic capacity and this capacity supports planning the content words in a phrase, we predicted that subjects with greater semantic WM capacity would show greater sentence elaboration and a longer mean length of utterance. If phonological WM capacity is not involved in this planning, then there should be no relation between phonological capacity and these narrative measures. As a control variable, we examined proportion of closed class words out of all narrative words. Closed class words are grammatical words such as prepositions (e.g., for, to) and determiners (e.g., a, the) conveying the structure of the sentence and are termed closed-class because they are a fixed set² This measure is thought to be connected to patients' ability to structure a phrase grammatically (Gordon, 2006; Gordon & Dell, 2003). As this ability may be separate from the ability to plan content words within a phrase (Martin & Freedman, 2001), we predicted no relations between this measure and the two WM capacities. If we find that our semantic WM measure does not relate to the proportion of closed class words but does to the other measures, such would indicate that the observed correlations do not arise simply from overall severity of deficits. Beyond these three primary measures, we also examined the relation between these WM capacities and speech rate (i.e., word per minute). We reasoned that if patients planned several words in a phrase simultaneously prior to the onset of an utterance, then while this phrase was being uttered, they could plan the subsequent phrase, which could be uttered at the end of the preceding phrase, resulting in more rapid, fluent speech. Also, they could avoid stops and restarts that might arise for individuals who could not plan all the intended words in a phrase, thus resulting in several attempts at a given phrase.

As measures of WM capacity, we used the category probe task described earlier to measure semantic WM capacity (Martin et al., 1994; Martin & He, 2004) and a digit matching span task to measure phonological capacity (Allport, 1984; Martin et al., 1994). Both tasks required a simple yes-no judgment and thus avoided any difficulties the patients might have with overt articulation of the lists. In the digit matching span task, participants hear two lists of digits and judge whether they are the same or different, where on the different trials, the order of two adjacent digits is switched (e.g., 2 1 7 3 – 2 7 1 3)³. The use of this task for tapping phonological WM is justified on the grounds that random digit lists carry little

¹In the QPA scoring system, the subject noun and main verb are not included in these elaboration counts since utterances must have a subject and main verb to qualify as sentences and every subject noun phrase requires a noun and every main verb phrase requires a verb.

²Closed class words contrast with content words which are open class words since new content words (e.g., nouns, verbs) are continuously added to the vocabulary

³Digit matching span was used rather than the rhyme probe measure used in prior case studies because preliminary data indicated that many of the patients at the acute stage had difficulty understanding the concept of a rhyming relation. Unpublished data from our lab

semantic information and thus WM capacity derived from this task should depend primarily on the retention of phonological codes. To rule out the possibility that any relation between the semantic and phonological WM measures and the narrative production measures derived from semantic or phonological deficits per se, we partialled out of any correlations performance on a task designed to assess the integrity of semantic and phonological information related to words (picture-word matching task; Breese & Hillis, 2004; Martin, Lesch, & Bartha, 1999).

2. Methods

2.1 Participants

As part of an ongoing study, a total of 69 subjects were recruited from Memorial Hermann and The Houston Methodist Hospitals in the Texas Medical Center in Houston, Texas within 72 hours of the onset of an acute ischemic left hemisphere stroke. Informed consent was obtained from either the subject or a legally authorized representative of the subject as approved by the institutional review boards of the Baylor College of Medicine, Rice University, the University of Texas Health Science Center, and the Houston Methodist Hospital. Inclusion criteria required that s/he suffered a left hemisphere cerebrovascular accident in the preceding 72 hours, was a native monolingual speaker of English, had sufficient cognitive and language function to understand task instructions, was between 18 and 85 years of age, could verbally consent or indicate that a family member could consent for them. Subjects had no previous history of symptomatic neurological disorder that introduced a separate cause of language or cognitive deficits apart from the stroke and no significant sensory deficits that rendered them unable to follow task instructions or see task items to be named. We did not include subjects who presented with significant impairment in semantic processing or spoken word perception (i.e., a composite z-score less than -2.5 relative to other brain damaged subjects on domain-specific tasks, leading to the elimination of one subject (see below). Of the 69 subjects, 37 (17 female) completed the tasks relevant to the goal of this paper. Six subjects were pre-morbidly left-handed. However, one subject was excluded (resulting in $N=36$) because he scored more than 2.5 standard deviations below the patient mean on a measure of spoken word perception (i.e., a composite z-score for proportion correct on auditory lexical decision and on picture-word matching with phonologically related distractors; the next lowest score was -1.68). All subjects completed the tasks in the same order across subjects, within a median of three days after onset of stroke symptoms (range 1–12 days, $SD=2.4$, average = 3.5), with the exception of two cases who were tested at eight and 12 days post-stroke, respectively. The mean age of subjects was 63.5 years (range 34–85; $SD=12.7$) and the mean years of education (information available for 25 subjects) was 13.5 years (range 10–20; $SD=2.3$).

Patients were assessed for the degree of apraxia of speech in order to determine if that contributed to their narrative production measures. Most directly, apraxia could contribute to reduced speech rate, but could potentially affect other aspects of production measures as well (e.g., Code, Ball, Tree, & Dawe, 2013). Twenty-eight patients were administered

for individuals at the chronic stage of stroke show a high correlation between digit matching span and rhyme probe ($r = .72$, $p = .0002$, $N=20$).

subtest 5 of the Second Edition of the Apraxia Battery for Adults (Dabul, 2000). For the remainder, speech samples from a picture description (either Cookie Theft or Picnic) and from the Cinderella narrative were analyzed for speech errors. A speech pathologist scored all patients on a four-point scale: 1 no apraxia (no speech sound errors), 2 mild apraxia (speech errors on <25% of words), 3 moderate apraxia (speech errors on 25–50% of words), and 4 severe apraxia (speech errors on > 50% of words). Thirty of the 36 patients had no apraxia of speech. The remaining six had scores as follows: 1.5 (n=1), 2.0 (n=3), 2.5 (n=1), and 3.0 (n=1).

2.1 Design

We collected data at the subject's bedside in the hospital or at the subject's home if the subject was discharged before testing was completed. Responses were recorded both by hand where possible and by digital recording. Responses were scored either by a post- BA or undergraduate research assistant. Subjects were tested on a range of tasks. Below we describe those relevant to the current study.

2.2 Tasks

2.2.1 Spoken Word Perception.—We used a shortened form of the auditory lexical decision from the PALPA task 5 (Kay, Lesser & Coltheart, 1992). Subjects were read a list of words and identified whether the word was a real word (n=20) or not (n=20). The dependent measure was the proportion of words and nonwords correctly identified.

2.2.2 Semantic & Phonological Input Processing.—We assessed deficits in semantic and phonological input processing using a single word single picture matching task (Breese & Hillis, 2004; Martin et al., 1999). The experimenter presented 17 pictures, once with a matching word (e.g., DUCK/duck), once with a semantic foil (e.g., DUCK/swan) once with a phonological foil (e.g., DUCK/truck) and once with an unrelated foil (e.g., DUCK/razor) for a total of 68 trials divided into four presentation sets. Subjects were asked “Is this a “ and responded “yes” if the picture matched the object and “no” otherwise, either verbally, or nonverbally by pointing to the word “yes” or “no” or by nodding/shaking the head. Accuracy for the different trial types is shown in Table 2. The mean accuracy across all trials was 95.0% correct with a 3.7% standard deviation and a range of 84% to 100%. Patients were generally quite accurate on the matching, unrelated, and phonologically related trials. There was more variability on the semantically related trials, with a mean of 86% correct and a range from 59% to 100%. Only 6 subjects scored greater than 90% correct on the semantically related trials (the mean for controls) and the remainder scored between 59% and 90% correct. In order to measure the participants' ability to discriminate the semantic and phonological foils from matching words, we calculated d' measures for performance on the semantically and phonologically related trials in comparison to matching trials. Only one subject was more than 2.5 SDs away from the mean accuracy on semantically or phonologically related trials (semantic: d' range = 1.62 – 3.76, mean = 2.96, cut-off criterion $d' < 1.33$; phonological: d' range = 1.99 – 3.76, mean = 3.39, cut-off criterion $d' < 1.73$) and this subject was eliminated from further analyses. We included partial performance scores for two subjects who completed only three of four presentation

sets. These d' measures were used to control for phonological and semantic processing abilities in relating the QPA measures to WM abilities.

2.2.3 WM Tasks.

2.2.3.1 Semantic WM (category probe task).: We used the category probe task to assess semantic WM (Martin et al., 1994). Subjects judged whether a spoken probe word was in the same category as any word in a preceding spoken list. The list items were sampled from the categories of animals, body parts, clothing, fruit, and kitchen equipment. Subjects responded “yes” if the probe word was in the same category and “no” otherwise, either verbally, or non-verbally by pointing or nodding of the head. Subjects were familiarized with the five categories to which the items belonged, and the instructions for the task prior to test administration. The list began with one item and went up to four items, with the items being presented at approximately one word per second. The number of trials was 8 for the one- and two-item lists, 12 for the three-item lists, and 16 for the four-item lists. For each list length, half of the lists had an item matching the category of the probe item and half did not. The position of the matching item was equated across serial position. Testing was discontinued when accuracy reached or dropped below 75% at a given list length. The dependent measure was the estimated span length for 75% correct, using linear interpolation between the two list lengths that span 75%. If a subject did not reach or dropped below 75% accuracy on the last list, linear interpolation was completed assuming the subject would have scored 50% on a list length of five items.

2.2.3.2 Phonological WM (Digit Span/Digit Match Span).: Due to changes in task design across the course of subject recruitment, four subjects were administered a digit span task (Wechsler, 1981) and the other 32 were administered the digit matching span task (Allen, Martin, & Martin, 2012). The digit span task was administered using the standard procedure from the WAIS-R (Wechsler, 1981). In the digit matching span task, subjects heard two digit lists, with the items in the list presented at approximately one word per second, and responded “yes” if they were the same and “no” if different. On non-matching trials, the second list reversed the order of two adjacent items. List lengths varied from two to six digits. The number of lists was 6 for the two-item lists, 8 for the 3-item lists, 6 for the 4-item lists, 8 for the 5-item lists and 10 for the 6-item lists. Half of the lists matched and half did not. For the non-matching lists, the position of the exchanged digits was approximately equally sampled across possible list positions. The task was discontinued once the subject reached or dropped below 75% percent accuracy on a given list. The dependent measure was estimated span length for 75% percent correct, calculated in the same fashion as the category probe task. If a subject did not reach or drop below 75% accuracy on the final list, the linear interpolation was calculated under the assumption the subject would have scored 50% on a list length of seven digits.

2.2.4 Narrative Production task.—Subjects recounted the Cinderella story after viewing the picture-book “Cinderella” (Jeffers, 2004) with the text occluded. Subjects took as long as needed viewing the modified picture-book, subsequently retelling the tale without visual aid. We asked subjects to be as descriptive as possible and use full sentences. If subjects produced little, the experimenter prompted subjects to continue speaking, typically

by asking the subject to elaborate further. We scored narratives using the procedures in the *Quantitative Production Analysis Training Manual* (Berndt, Wayland, Rochon, Saffran, & Schwartz, 2000) for all utterances produced (Gordon, 2006). For each subject, we calculated four dependent measures of narrative production. We included sentence elaboration (a combined index of subject noun and verb phrase elaboration), mean utterance length, words per minute, and proportion of the words spoken that were closed class words. We established reliability of the transcription and scoring across two raters using a randomly selected sub-sample of 15 patients who produced more than 10 utterances. For transcription reliabilities, we sampled the middle 30 seconds of recording from each patient with slight adjustment as not to break any complete utterances. We compared raters on the number of narrative words, the identity of narrative words, and the number of utterances into which identical narrative words were segmented. On average, across samples, raters reached 95% agreement on the number of narrative words (range 78% - 100%) where an average 93% of narrative words were identical (range 78% - 100%). When segmenting identical narrative words into utterances, raters showed an average agreement of 98% (range 75% - 100%). To establish reliability for scoring following Gordon (2006), 10 utterances were randomly selected from each of the 15 patients based on the utterances segmented by one rater. Across the four dependent measures of interest, inter-rater reliability was on average 95% (range 86%-100%).

In the current study, we employed the following narrative production measures from the QPA

2.2.4.1 Sentence elaboration.: We categorized utterances as sentences when they included a subject and main verb. The sentence elaboration score is a combined score of the number of content words plus pronouns in subject noun phrases and the number of content words plus pronouns in the main verb phrases divided by the number of subject noun or verb phrases produced.

2.2.4.2 Mean Utterance length.: The mean number of words per utterance, across all types of words and all types of utterances produced.

2.2.4.3 Words per minute.: The number of complete words produced per minute.

2.2.4.4 Proportion of closed-class words produced.: Narrative words were divided into closed-class (pronouns, auxiliary verbs, conjunctions, prepositions, adverbs not ending in -ly, and determiners) and other words (open-class words, including nouns, verbs, adjectives, and adverbs ending in -ly). The proportion of closed-class words was the number of closed class words as a proportion of all narrative words produced.

3. Results

3.1 Summary statistics for patient performance on the target QPA measures for the narrative language task and on WM and single word processing measures are presented in Table 2. As indicated by the means and minimum and maximum values shown there, considerable variability was obtained on all measures. Interestingly, while patients' means were more

than one standard deviation lower than those for controls for nearly all measures, the exception was proportion of closed class words, where the mean, standard deviation, and range were very similar for patients and controls. This suggests that this group of patients were preserved in their ability to structure a sentence in terms of the selection of closed class items. Also shown in Table 2 is the percentage of patients who scored below the control range for each measure. These proportions were highest for the two WM measures, moderately high for the narrative measures for the QPA, and lowest for the single word processing measures.

Pairwise correlations among these measures are shown in Table 3. As might have been expected given that narrative production involves single word semantic and phonological processing, several significant correlations were obtained between the word processing measures (d' semantic and d' phonological) and the narrative measures. For instance, the d' semantic measure correlated significantly with sentence elaboration, mean length of utterance, and proportion of closed class words and the d' phonological measure correlated significantly with mean length of utterance. In general, these correlations were larger for the semantic than the phonological WM measure and larger for the narrative measures related to the content of the utterances than for speech rate (word per minute). For the WM measures, significant correlations were obtained between category probe span and two of the narrative measures predicted to be related to semantic WM - that is, sentence elaboration and mean length of utterance, whereas smaller correlations were obtained between these narrative measures and digit matching span, with the correlation non-significant for sentence elaboration. Also in line with predictions, neither WM measure correlated significantly with proportion of closed class words. The lack of correlation of proportion closed class words with the WM measures could not be attributed to a lack of sensitivity of the measure (given patients' good performance) as this measure correlated significantly with two other measures - mean utterance length and d' semantic. However, for words per minute, the pattern was opposite than predicted with a stronger correlation for digit matching span than category probe. The correlation between the two WM measures was of marginal significance, $r(34) = .32, p = .06$.

3.2. Assessment of independent contributions of semantic and phonological WM

Given the correlation between the two WM measures, it is possible that a significant correlation between a narrative measure and both WM measures might be obtained, but not because both phonological and semantic WM contribute, but due to the overlap in variance between the two. Consequently, we wished to determine whether an independent contribution of the semantic WM measure could be demonstrated when the phonological WM and the two control word processing measures were included as predictors. Thus, we carried out multiple regressions of the four narrative measures on category probe span, digit matching span, d' semantic, and d' phonological. In a multiple regression, the significance of the coefficient (i.e., the beta weight) for each variable reflects the significance of the contribution of that variable to the outcome variable that is independent of the other predictors (Darlington, 1990). That is, in this application, the significance of the co-efficient for category probe span reflects the significance of the contribution of this variable in predicting the outcome narrative measure that is independent of digit matching span and the

two measures of semantic and phonological processing ability (d' semantic and d' phonological). Similarly, the significance of the co-efficient for digit matching reflects the significance of its contribution independent of category probe, d' semantic and d' phonological. In carrying out these regressions, we first tested for outliers, using the criteria of a studentized residual greater than 2.5 and a Cook's d greater than 3 times the mean. No outliers were identified and thus all 36 subjects were included in all regressions.

The results of the multiple regressions are shown in Table 4. Leverage plots showing the influence of category probe span (semantic WM) and digit matching span (phonological WM) when controlling for the other variables are shown in Figure 1. As predicted, for sentence elaboration and mean length of utterance, the category probe measure had a significant independent contribution whereas digit matching span did not. For proportion of closed class words, neither WM measure had a significant independent contribution. The one outcome not in line with predictions was that for words per minute, digit matching span had a significant independent contribution whereas category probe span did not. These analyses were also carried out including the apraxia of speech rating as a control variable. The pattern of significant and non-significant weights was unaffected (see Supplementary Table S1).⁴

4. Discussion

To summarize, the results confirmed three of the four predictions regarding a role for semantic WM, separate from phonological WM, in spontaneous language production in a large group of subjects during the acute phase of stroke. Based on the results from the multiple regressions, we found that semantic WM had a significant contribution in predicting the two measures sensitive to the number of content words per phrase (sentence elaboration and mean length of utterance) that was independent of phonological WM whereas phonological WM did not have a significant contribution that was independent of semantic WM. In contrast, neither WM measure had an independent contribution in predicting a grammatical property of the utterance (i.e., proportion of closed class words). The one unanticipated outcome was that speech rate had a significant independent relation to phonological WM but not semantic WM. While this outcome was not in line with predictions, the contrasting patterns for speech rate vs. sentence elaboration and mean length of utterance serve to support the contention that there are two separate WM capacities that play different roles in language production.

The findings for sentence elaboration and mean length of utterance are consistent with prior case studies of four individuals showing that the two with semantic WM deficits, but not the two with phonological WM deficits, had difficulty producing phrases containing more than one content word (Freedman et al., 2004; Martin & Freedman, 2001; Martin et al., 2004). The present results extend these findings to spontaneous language production across a large group of subjects in the acute phase of stroke before the development of behavioral

⁴We also analyzed the data using only those subjects who scored within the control range on both the d' semantic and d' phonological measures, as another approach to controlling for single word processing abilities. For this smaller sample ($n=18$), the pattern of regression coefficients was very similar to that for the whole group, though the coefficients were of marginal significance for category probe as a predictor of sentence elaboration ($p=.05$) and for digit matching as a predictor for words per minute ($p=0.10$).

strategies and neural reorganization could occur. The results support the contention that in spontaneous speech as well as in constrained production, semantic WM, but not phonological WM, is used in planning the content words in sentences. That is, semantic WM is used to maintain the content words within a phrase prior to phonological retrieval. Neither WM capacity is relevant for planning the grammatical structure of the sentence as conveyed through the use of closed class words. Thus, sentence production depends on WM resources to keep semantic representations within a phrase active until articulation can begin. Other evidence suggests that this semantic WM capacity plays a general role in language processing, supporting comprehension when word meanings are maintained across some distance prior to their integration (Hanten & Martin, 2001; Martin & Romani, 1994; Martin & He, 2004), though it would be valuable to replicate these patient findings as well at the acute stage of stroke.

One issue that often arises regarding the relation of semantic WM to sentence processing is the extent to which such relations derive from semantic processing abilities per se rather than a separable retention capacity (see Martin et al., 1999, for discussion). In the present study, we addressed this issue by controlling for patients' performance on picture-word matching, specifically for trials in which the word was semantically related to the picture. In this task, subjects judged whether a spoken word matched a single picture, making a yes-no judgment. The semantically related non-matching trials including words from the same semantic category (e.g., fruits, clothing, birds), which were quite similar in meaning (e.g., pear for apple, boot for shoe, duck for swan). Prior evidence suggests that such a yes/no task requires more detailed semantic information than does a task in which the patient is presented with the correct object and a semantic foil simultaneously and must choose the one matching the picture (Hillis & Caramazza, 1995). In the present study, patient performance on the semantically related trials had a mean of 87% and a range of 59% to 100% correct, with four patients scoring less than 75% correct. Performance on the picture word matching task significantly correlated with three of the four narrative measures, indicating that there was sufficient sensitivity in the measure to predict performance on a production task. Moreover, considerable evidence suggests that patients with semantic processing deficits often lose specific details of concepts that are needed for fine discriminations between closely related members of the same category (required for the picture-word matching task), while preserving more general superordinate category information (required for the category probe task) (Rogers & Patterson, 2007; Warrington, 1975). Thus, there would appear to be little grounds for arguing that the semantic processing requirements of the category probe task were more challenging than those for the picture-word matching task. Of course the category probe task requires more than accessing the category representation of the list and probe words, but the major additional requirement is the working memory requirement - that is, maintenance of the semantic information for the list items so that a category match or non-match decision could be made.

As suggested in the introduction, the absence of a relation between phonological WM capacity and the sentence elaboration and mean length of utterance measures might be attributed to a much smaller scope of planning at the phonological level (Wheeldon & Lahiri, 2002; Garrett, 1980), which is within all of the patients' capacities. If so, then how does one explain the fact that phonological WM had a greater independent relation than did

semantic WM to speech rate? Of course, in producing speech, it is necessary to access a phonological representation in order to begin articulation. The speed and accuracy with which phonological representations can be accessed would thus influence speech rate. One might speculate that performance on the digit matching span task reflects not just storage of phonological forms derived from perception of the digits, but also reflects the ability to access output phonological forms, perhaps because of an involvement of subvocal rehearsal in the task (see Allen & Hulme, 2006; Martin et al., 1999). Certainly the rate at which the digits were presented would allow for rehearsal to be employed. It should be noted that the speech rate measure from the QPA reflects speed of retrieval for all words, including single word utterances, rather than structural aspects of the utterance, and thus might be less influenced by phrasal planning than was the case for sentence elaboration and mean length of utterance. Consistent with this line of reasoning, sentence elaboration and mean length of utterance had a strong positive correlation ($r(35) = .85, p < .0001$), whereas the correlations between words per minute and the two other measures were much smaller and of marginal significance ($r(35) = .310, p = .061$, for sentence elaboration and $r(35) = .287, p = .085$, for mean length of utterance).

In conclusion, this study provides strong support for the contention that a WM capacity specific to maintaining lexical-semantic representations is critical for producing multiword utterances. This lexical-semantic capacity is distinct from knowledge of semantic representations and also distinct from a capacity for maintaining phonological representations, which had no independent contribution in predicting measures related to phrasal planning. In contrast, phonological WM capacity but not semantic WM had an independent role in predicting speech rate, perhaps due to a role for the speed of accessing phonological forms in both measures. In addition to their theoretical significance, the present results potentially have important clinical implications as they suggest that improving semantic and phonological WM after stroke could have a direct effect on patients' ability to produce connected speech.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors wish to thank Miranda Brenneman, Cris Hamilton, and Danielle Rossi for data collection and analysis. We thank Erica Johns, Bowie Lin, and Hao Yan for transcription and analysis of narrative speech samples. We thank Lynn Maher for her help in diagnosing apraxia of speech in our subject population. Finally, we thank our subjects who agreed to participate in this study. This work was supported by an R01DC014976 award to the Baylor College of Medicine from the National Institute on Deafness and Other Communication Disorders and an award from the Moody Endowment to Rice University. Some of this work was presented at the Academy of Aphasia in Baltimore, MD (2017) and in Wales, UK (2016).

References

- Allen CM, Martin RC, & Martin N (2012). Relations between short-term memory deficits, semantic processing and executive function. *Aphasiology*, 26, 428–461. [PubMed: 22736889]
- Allen R, & Hulme C (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language*, 55, 64–88.

- Allport DA (1984). Auditory verbal short-term memory and conduction aphasia In Bouma H & Bouwhuis DG (Eds.), *Attention and performance X: Control and language processes* (pp. 351–364). Hillsdale, NJ: Erlbaum.
- Allum PH, & Wheeldon LR (2007). Planning scope in spoken sentence production: The role of *grammatical units*. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 33, 791–810.
- Allum PH, & Wheeldon LR (2009). Scope of lexical access in spoken sentence production: Implications for the conceptual-syntactic interface. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 35, 1240–1255.
- Baddeley AS (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Science*, 4, 417–423.
- Berndt RS, Wayland S, Rochon E, Saffran E, & Schwartz M (2000). Quantitative production analysis: A training manual for the analysis of aphasic sentence production. Hove, U.K: Psychology Press.
- Bock JK, & Levelt W (1994). Language production: Grammatical encoding In Gernsbacher MA (Ed.), *Handbook of psycholinguistics* (pp. 945–984). New York: Academic Press.
- Breese EL, & Hillis AE (2004). Auditory comprehension: is multiple choice really good enough? *Brain and Language*, 89, 3–8. [PubMed: 15010231]
- Corbetta M, Ramsey L, Callejas A, Baldassarre A, Hacker CD, Siegel JS, et al. (2015). Common behavioral clusters and subcortical anatomy in stroke. *Neuron*, 85(5), 927–941. [PubMed: 25741721]
- Chang F, Dell GS, & Bock K (2006). Becoming syntactic. *Psychological Review*, 113, 234–272. [PubMed: 16637761]
- Dabul B (2000). *Apraxia battery for adults* (2nd ed). Austin, TX: PRO-ED Inc.
- Darlington RB (1990). *Regression and linear models*. New York, NY: McGraw-Hill College.
- DeLeon J, Gottesman RF, Kleinman JT, Newhart M, Davis C, Heidler-Gary J, Lee A, & Hillis AE (2007). Neural regions essential for distinct cognitive processes underlying picture naming. *Brain*, 130(5), 1408–1422. [PubMed: 17337482]
- Forkel SJ, Thiebaut de Schotten M, Dell'Acqua F, Kalra L, Murphy DG, Williams SC, & Catani M (2014). Anatomical predictors of aphasia recovery: a tractography study of bilateral perisylvian language networks. *Brain*, 137(7), 2027–2039. [PubMed: 24951631]
- Freedman M, Martin RC, & Biegler K (2004) Semantic relatedness effects in conjoined noun phrase production: Implications for the role of short-term memory. *Cognitive Neuropsychology*, 21, 245–265. [PubMed: 21038203]
- Garrett MF (1975). The analysis of sentence production In Bower G (Ed.), *Psychology of Learning and Motivation*, (pp. 133–177). New York, NY: Academic Press.
- Garrett M (1980). Levels of processing in sentence production In Butterworth B (Ed.), *Language production*, Vol. 1: Speech & talk (pp.177–220). London: Academic Press.
- Gordon JK (2006). A quantitative production analysis of picture description. *Aphasiology*, 20, 188–204.
- Gordon JK, & Dell GS (2003). Learning to divide the labor: an account of deficits in light and heavy verb production. *Cognitive Science*, 27, 1–40.
- Griffin Z, & Bock K (2000). What the eyes say about speaking. *Psychological Science*, 11, 274–279. [PubMed: 11273384]
- Geranmayeh F, Chau TW, Wise RJ, Leech R, & Hampshire A (2017). Domain-general subregions of the medial prefrontal cortex contribute to recovery of language after stroke. *Brain*, 140(7), 1947–1958. [PubMed: 29177494]
- Gupta P, & Tisdale J (2009). Word learning, phonological short-term memory, phonotactic probability, and long-term memory: towards an integrated framework. *Philosophical Transactions of the Royal Society B*, 364, 3755–3771.
- Hanten G, & Martin R (2000). Contributions of phonological and semantic short-term memory to sentence processing: Evidence from two cases of closed head injury in children. *Journal of Memory and Language*, 43, 335–361.

- Haarmann HJ, Davelaar EJ, & Usher M (2003). Individual differences in semantic working memory and sentence comprehension. *Journal of Memory and Language*, 48, 320–345.
- Hillis AE, Kleinman JT, Newhart M, Heidler-Gary J, Gottesman R, Barker PB, Aldrich E, Llinas R, Wityk R, & Chaudhry P (2006). Restoring cerebral blood flow reveals neural regions critical for naming. *Journal of Neuroscience*, 26, 8069–8073. [PubMed: 16885220]
- Jeffers S, Perrault C, & Ehrlich A (2004). *Cinderella*. New York: Dutton Children's Books.
- Kay J, Lesser R, & Coltheart M (1992). *The Psycholinguistic Assessment of Language Processing in Aphasia (PALPA)*. Hove, U.K: Erlbaum.
- Kümmerer D, Hartwigsen G, Kellmeyer P, Glauche V, Mader I, Klöppel S, Suchan J, Karnath H, Weiller C, & Saur D (2013). Damage to ventral and dorsal language pathways in acute aphasia. *Brain*, 136(2), 619–629. [PubMed: 23378217]
- Martin N, & Saffran EM (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, 14, 641–682.
- Martin RC, & Freedman ML (2001). Short-term retention of lexical-semantic representations: Implications for speech production. *Memory*, 9, 261–280.
- Martin RC, & He T (2004). Semantic short-term memory deficit and language processing: A replication. *Brain and Language*, 89, 76–82. [PubMed: 15010239]
- Martin R, Lesch M, & Bartha M (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, 41, 3–29.
- Martin RC, Miller M, & Vu H (2004). Working memory and sentence production: Evidence for a phrasal scope of planning at a lexical-semantic level. *Cognitive Neuropsychology*, 21, 625–644. [PubMed: 21038225]
- Martin RC, & Romani C (1994). Verbal working memory and sentence processing: A multiple components view. *Neuropsychology*, 8, 506–523.
- Martin R, Shelton J, & Yaffee L (1994). Language processing and WM: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language*, 33, 83–111.
- Meyer AS (1996). Lexical access in phrase and sentence production: Results from picture- word interference experiments. *Journal of Memory and Language*, 35, 477–496.
- Nishiyama R (2013). Dissociative contributions of semantic and lexical-phonological *information to immediate recognition*. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 642–648.
- Page MPA. , & Norris D. (1998;). The primary model: A new model of immediate serial recall. *Psychological Review*. ,105,;761–781.-. [PubMed: 9830378]
- Rochon E, Saffran E, Berndt R, & Schwartz M (2000). Quantitative analysis of aphasic sentence production: Further development and new data. *Brain and Language*, 3, 193–218.
- Rogers TT, & Patterson K (2007). Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, 136(3), 451–469. doi: 10.1037/0096-3445.136.3.451 [PubMed: 17696693]
- Schwartz MF, & Dell GS (2010). Case series investigations in cognitive neuropsychology. *Cognitive Neuropsychology*, 27, 477–494. [PubMed: 21714756]
- Shivde G. , & Anderson MC. (2011;). On the existence of semantic working memory: Evidence for direct semantic maintenance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. ,37,;1342–1370.-.
- Tan Y, Martin RC, Van Dyke JA (2017). Semantic and syntactic interference in sentence comprehension: A comparison of working memory models. *Frontiers in Psychology*. 8:198, doi: 10.3389/fpsyg.2017.00198 [PubMed: 28261133]
- Warrington EK (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, 27, 635–657. [PubMed: 1197619]
- Wechsler D (1981). *Wechsler Adult Intelligence Scale-Revised (WAIS-R)*. New York: Harcourt Brace Jovanovich.
- Wheeldon LR, & Lahiri A (2002). The minimal unit of phonological encoding: Prosodic or lexical. *Cognition*, 85, B31–B41. [PubMed: 12127702]

Figure 1.

Multiple regression scatterplots demonstrating the independent contribution of semantic WM (panels A-D), and phonological WM (panels E-H) to spontaneous language production (sentence elaboration (NP + VP), mean utterance length, words per minute, and the proportion closed class words produced) while controlling for both measures of input word processing (measured by accuracy (d') on word-picture matching with semantically related (for semantic WM) and phonologically related foils (for phonological WM)).

Examples of patients' sentence utterances and their respective phrase elaboration counts for: A) subject noun phrase and B) main verb phrases. The relevant phrases are italicized and the words contributing to the elaboration score are in bold.

Table 1.

A) Subject noun phrases	Number of content words (excluding subject noun)
<i>She</i> ran the rain.	0
<i>The young prince</i> took her away	1
<i>Cinderella</i> dreamt about the nice dance and visions of the prince	0
<i>Her fairy godmother</i> told her she must be in by midnight	2

B) Main verb phrases	Number of content words (excluding the main verb)
<i>She ran the rain.</i>	1
The young prince <i>took her away</i>	2
<i>Cinderella dreamt about the nice dance and visions of the prince</i>	4
Her fairy godmother <i>told her she must be in by midnight</i>	4

Table 2.

Descriptive statistics for narrative, working memory, and lexical processing measures

Variable	Patients (N=36)					Controls (N=13) ^a				
	Mean	Std Dev	Min	Max	Percent BCR ^b	Mean	Std Dev	Min	Max	
Sentence Elaboration	2.1	0.7	0.5	3.6	33	3.2	0.82	1.8	4.5	
Mean Length of Utterance	7.2	1.9	2.3	11.5	33	9.8	2.11	6.1	12.9	
Words per Minute	107.9	42.1	16.6	220	56	152	23.92	112.3	199	
Proportion Closed Class	0.54	0.05	0.43	0.63	8	0.54	0.03	0.46	0.57	
Category Probe Span	2.5	1.3	0.5	4.5	69	4.7	0.76	3.4	7	
Digid Matching Span	5	1.3	2	6.5	75	6.1	0.51	5.8	9	
Word-Picture Matching										
Matching	98	3.7	87	100	19	100	0	100	100	
Semantically related	86	9.6	59	100	8	90	7.3	76	100	
Phonologically related	95	4.7	85	100	19	99	2.6	94	100	
Unrelated	99	3.3	82	100	11	100	0	100	100	
d' semantic	3.0	0.5	1.6	3.8	19	3.2	0.4	2.6	3.8	
d' phonological	3.4	0.5	2	3.8	28	3.7	0.14	3.4	3.8	

^a 12 of 13 participants were tested on word-picture matching^b Percent BCR - percentage below control range

Table 3.
Correlations among all measures (N=36). Correlations in bold are significant at $p < .05$.

	words/min	sentence elaboration	prop. closed class	mean utterance length	category probe	digit matching span	d' sem
sentence elaboration	0.29						
proportion closed class	0.10	0.30					
mean utterance length	0.27	0.85	0.35				
category probe span	0.29	0.51	0.25	0.49			
digit matching span	0.38	0.29	-0.07	0.34	0.32		
d' semantic	0.18	0.42	0.33	0.35	0.35	0.15	
d' phonological	0.05	0.32	0.23	0.33	0.32	0.42	0.60

Table 4.

Results of multiple regression for each language production variable (p-value for significant coefficients shown in bold).

Production Measure	Co-efficients				Whole Model		
	<i>beta</i>	<i>t</i>	<i>Std error</i>	<i>p</i>	<i>F</i>	<i>df</i>	<i>p</i>
Sentence elaboration					4.06	4,31	0.009
Category probe	0.373	2.30	0.162	0.028			
Digit matching	0.148	0.88	0.168	0.384			
d' semantic	0.294	1.55	0.190	0.132			
d' phonological	-0.042	-0.21	0.200	0.837			
Mean length of utterance					3.54	4,31	0.017
Category probe	0.360	2.17	0.166	0.038			
Digit matching	0.186	1.08	0.172	0.288			
d' semantic	0.179	0.92	0.195	0.365			
d' phonological	0.032	0.15	0.213	0.879			
Words per minute					2.53	4,31	0.060
Category probe	0.169	0.97	0.174	0.338			
Digit matching	0.438	2.44	0.180	0.021			
d' semantic	0.264	1.30	0.203	0.204			
d' phonological	-0.345	-1.60	0.216	0.120			
Proportion closed class words					1.56	4,31	0.21
Category probe	0.204	1.12	0.182	0.204			
Digit matching	-0.217	-1.15	0.189	-0.217			
d' semantic	0.223	1.04	0.214	0.306			
d' phonological	0.119	0.52	0.229	0.119			