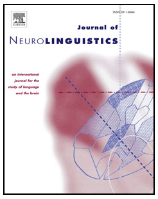
Journal of Neurolinguistics 48 (2018) 90–103

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling



Working memory and discourse production in people with aphasia

Dalia Cahana-Amitay[a,*], Theodore Jenkins[b]

[a] Boston University School of Medicine, Department of Neurology, United States
[b] University of Connecticut, Department of Speech, Language, and Hearing Sciences, United States

ARTICLE INFO

Keywords:
Aphasia
Narrative
Microlinguistic
Macrolinguistic
Span

ABSTRACT

This study explored the relationship between Working Memory (WM) and discourse production in people with aphasia, based on data from the AphasiaBank. The dataset comprised the children's story “Cinderella” and basic WM measures of span, collected from 45 participants (15 people with nonfluent Broca's aphasia, 15 people with anomic aphasia, 15 people with fluent Wernicke's aphasia). Discourse samples were coded for and analyzed in terms of content, micro- (words and sentences) and macro- (groups of sentences) linguistic components, known to demonstrate multi-level discourse ability. Comparisons were made among the different participant groups to identify differences and/or commonalities in performance. Results showed that WM, as measured by reduced word and sentential span, influences macrolinguistic narrative components and may be sensitive to aphasia type. Findings were interpreted in terms of a potential deficit to Baddeley's episodic buffer, affecting sequential and hierarchical narrative information processing.

1. Introduction

Discourse is the communicative foundation upon which humans typically manage their day-to-day tasks (Davidson, Worrall, & Hickson, 2003). It is commonly defined in terms of linguistic units that extend beyond a simple clause, expressed for a specific communicative function (Armstrong, 2000; Halliday & Christian, 2004). Although many Persons with Aphasia (PWA) retain the ability to communicate basic wants and needs, most seek to expand their discourse repertoire to capture a greater range of communicative purposes (Worrall et al. 2011). Indeed, clinical research has attempted to address this priority, with a noticeable increase in the study of discourse patterns among PWA (Armstrong, Ferguson, & Simmons-Mackie, 2013), especially over the past 40 years (Bryant, Ferguson, & Spencer, 2016). However, a clear characterization of discourse problems that PWA exhibit remains a challenge, largely due to methodological differences among studies, such as participant demographics and medical history, sample size, methods of discourse elicitation, and absence of normative data against which evaluation of discourse samples can be done (Whitworth, Claessen, Leitao, & Webster, 2015).

Most studies of discourse production in aphasia have focused on the analysis of the linguistic features underpinning narratives produced by people with aphasia (Bryant et al., 2016). Based on a review of 165 studies, Bryant et al. (2016) identified 536 different discourse measures, which they divided into three broad groupings: (i) language productivity (sample length, lexical diversity, speech fluency, word finding), (ii) information content (efficiency, cohesion, lexical, semantic/conceptual, schema-related), and (iii) grammatical complexity (morphological, word class, syntactic). This division suggests that a comprehensive analysis of narrative discourse in aphasia should involve multiple discourse levels (Ellis, Henderson, Wright, & Rogalski, 2016; Linnik, Bastiaanse, & Höhle, 2016; Sherratt, 2007).

---

[*] Corresponding author. Language in the Aging Brain, Veterans Affairs Boston Healthcare System, 150 South Huntington Avenue, Boston, MA 02130, United States.
E-mail address: dcamitay@bu.edu (D. Cahana-Amitay).

https://doi.org/10.1016/j.jneuroling.2018.04.007
Received 17 July 2017; Received in revised form 20 April 2018; Accepted 21 April 2018

Armstrong (2000) proposed a useful theoretical framework to capture the multi-level nature of discourse production patterns in aphasia. In this framework, discourse problems PWA experience can be represented from three main perspectives:

(i) *Structuralist*, which is focused primarily on microstructural discourse components, such as lexical items, morpho-syntactic structures, word classes, word-finding problems, grammatical complexity, grammatical errors
(ii) *Functionalist*, which is focused primarily on macro-structural discourse components, such as overall text structure (e.g., story grammar), conversational components (e.g., turn-taking), communicative effectiveness/success
(iii) *Macro-micro*, which combines (i) and (ii) and considers components such as content units, correct information units, main concepts

Evidence suggests that PWA have the greatest challenges with microstructural discourse components, as reflected in omission of obligatory arguments (e.g., Byng & Black, 1989; Whitworth, 2010), problems with discourse cohesion (Glosser & Deser, 1991), and difficulties using referential devices such as pronouns (e.g., Armstrong, 2000; Nicholas, Obler, Albert, & Helm-Estabrooks, 1985). Reports have been mixed, though, with some studies noting deficits (Ulatowska, North, & Macaluso-Haynes, 1981), and others not (Bloom, Borod, Santschi-Haywood, Pick, & Obler, 1996; Glosser & Deser, 1991). Difficulties with the production of micro-level discourse components manifest differently, based on aphasia type (Code, 2010). For example, in a recent study of narratives produced by 22 aphasic participants, Manning and Franklin (2016) found that fluent participants had more pronoun errors and produced qualitatively different articles than their nonfluent counterparts. The nonfluent participants, on the other hand, had a higher number of omissions than fluent ones.

PWA have a somewhat easier time with macro-structural discourse components (Armstrong et al., 2013; Lemme, Hedberg, & Bottenberg, 1984; Lock & Armstrong, 1997), although some studies have found that PWA produce impoverished information (Doyle et al., 2000), show reduced relevance and accuracy (Nicholas & Brookshire, 1993), and have trouble with story grammar (Whitworth, 2010). Distinctions between aphasia types are less evident in such analyses, as PWA appear to be more impaired than adults without aphasia but oftentimes show no statistical difference between aphasia groups (e.g., Manning & Franklin, 2016).

The differentiation between production of micro-vs. macro-level discourse components among PWA is not clear-cut, however, recent studies demonstrate an interdependency between the two (for a detailed description of discourse measures in people with neurogenic disorders, see Kong, 2016). Andreetta, Cantagallo, and Marini (2012), examined the effect of lexical deficits in people with anomic aphasia on their production of macro-linguistic narrative components. In comparison to people without aphasia, discourse impairments among PWA manifested as slower speech rate, shortened Mean Length of Utterance, fewer grammatical sentences, greater semantic paraphasic errors, as well as greater cohesion and global coherence errors, and lesser lexical information units. The authors found that cohesion errors and sentence completeness were strongly correlated, as were global coherence errors and production of lexical information units, and concluded that certain aspects of lexical retrieval possibly affect production of macrolinguistic components in narratives of PWA. They proposed that impaired utterance production among people with anomia adversely affects sentence completeness and inter-utterance cohesion. They further argued that to handle lexical deficits, patients produce lexical fillers and repetitions that impede level of global coherence.

Despite the gains in understanding discourse production in aphasia, characterization of the discourse impairments in this population remains a challenge. Findings suggest that there is considerable interindividual variability among PWA, in terms of the amount of content they provide in their narratives vis-à-vis the intactness of the linguistic structures they produce to express that information. That is, a PWA may demonstrate mild linguistic impairment but, at the same time, show limited expression of discourse content, and vice versa (Pritchard, Hilari, Cocks, & Dipper, 2017). To account for such variability, it might be useful to consider the potential contribution of nonlinguistic cognitive factors to performance, such as those associated with working memory, which are known to be part and parcel of linguistic behaviors observed in aphasia (for a comprehensive review, see Cahana-Amitay & Albert, 2015). Examination of the production of discourse in aphasia through such a lens captures the idea that discourse embodies cognitive-linguistic interlinks that make up the unfolding of natural language in a communicative context (Fergadiotis & Wright, 2011; Manning & Franklin, 2016).

Working memory (WM) has been conceptualized as a limited capacity system, designed to maintain, process, and manipulate information over short spans of time (Baddeley, 2003). In this multicomponent model, a central executive system allots attention resources to three slave components: (1) a phonological loop, which governs the rehearsal and maintenance of verbal information; (2) a visuospatial sketchpad, which stores visual and spatial information; and (3) an episodic buffer, which integrates information from the phonological loop, the visuospatial sketchpad, and long-term memory. In such a system, WM could be considered the workspace in which information is stored and manipulated, while the central executive system would be responsible for the ability to organize WM representations to enhance efficiency of information processing (e.g., Carpenter, Just, & Reichle, 2000; Connor, MacKay, & White, 2000).

Baddeley's WM model is assumed to interact with language performance, suggesting that the integrity of WM capacity can impact a person's ability to carry out language tasks (Baddeley, 2003; Carpenter & Just, 1989; Murray, 2012). Indeed, there is ample evidence that many PWA have difficulties with tests of WM and that these impairments adversely affect their language performance (e.g., Baldo & Dronkers, 2006; Beeson, Bayles, Rubens, & Kaszniak, 1993; Caplan & Waters, 1999; Caspari, Parkinson, LaPointe, & Katz, 1998; Christensen & Wright, 2010; Friedmann & Gvion, 2003; Gvion & Friedmann, 2012; Laures-Gore, Marshall, & Verner, 2011; Rönnberg et al., 1996; Sung et al., 2009; Ween, Verfaellie, & Alexander, 1996; Wright, Downey, Gravier, Love, & Shapiro, 2007). Most reports, describe WM deficits influencing tasks involving phonological, semantic, phrasal, and sentential elements (for a review, see Cahana-Amitay & Albert, 2015), with little to no mention of effects on discourse-level components. This gap is important to address, as the more open-ended a language task is, the more cognition-reliant it becomes (Alexander, 2006).

A notable attempt to answer the question concerning the relationship between WM and discourse in aphasia is a recent treatment study by Henderson, Kim, Kintz, Frisco, and Wright (2017), which examined the effects of narrative treatment on WM among 4 PWA. The narrative treatment they administered spanned three weeks and involved three presentations of sequential pictures of the same narrative, with each iteration containing new informational content (4, 6, 8 pictures). Treatment was followed by a maintenance phase immediately after treatment and an additional evaluation after a month. Changes in participants' performance were assessed in terms of number of thematic units. Improved performance, as measured by increased number of thematic units, was taken as an indication of activation of verbal WM and the episodic buffer, where multimodal information from the phonological loop, the visuospatial sketchpad, and long-term memory is processed. It is difficult to tell, though, whether the treated participants experienced measurable changes in their WM following discourse treatment, as the authors did not include cognitive measures in their post-treatment assessments, and if they did, they did not report the participants' WM scores.

Despite the inconclusive picture that emerges from the Henderson et al. (2017) study, their idea that Baddeley's episodic buffer plays an important role in discourse production is compelling. The buffer functions as a temporary store of up to four “episodes” of a multimodal code combing different types of information (e.g., visual with phonological) (Baddeley, 2007). With such a chunking mechanism in place, the buffer offers an efficient way to store complex information, accounting for word recall effects that extend beyond simple word lists (Baddeley, 2007). Thus, of the different components of Baddeley's WM model, the episodic buffer seems to provide the most straightforward locus for examining potential effects on discourse production, where narrative information is conveyed episodically. A deficit to the buffer might, therefore, interfere with a person's ability to integrate information from different sources or to temporarily store the complex information. However, the literature offers limited insight into how the episodic buffer functions or how to measure its integrity, making it difficult to predict specific patterns of impairment in aphasia.

According to Henderson et al. (2017), impact of an impairment to the episodic on discourse performance should be observed at both micro- and macro-level components. For microlevel elements, they note potential challenges with lexical access, based on the multimodal nature of the semantic system that underpins meaning-making. Dysfunction could involve difficulties combining enough semantic information to derive meaning, and/or problems combining semantic information from LTM with phonological information. The authors contend that discourse performance could be further complicated if the episodic buffer fails to handle the processing demands of “contextual semantics”, whereby contextual information is integrated with conceptual semantic information.

For macrolevel discourse elements, Henderson et al. (2017) suggest examining the organization of episodic information, on the assumption that dysfunction in the episodic buffer could lead to difficulties in producing coherently linked episodes. Globally, such a deficit could manifest as an inability to establish a consistent relationship between episodes and the general topic of the discourse sample, reflecting breakdown in hierarchical discourse organization. Locally, PWA might experience problems linking new information to previously presented information, indicating disrupted sequential processing. Under such circumstances, one might expect, for example, failure to assess whether enough information has been introduced in the discourse to establish a semantic link between an antecedent and its referent, leading to ineffective reference resolution. In the face of sequential impairment, PWA might face an additional challenge affecting their adherence to the sequential demands of story grammar, where propositions are ordered in a particular way.

Exploration of the role of WM as related to language production in aphasia has thus clearly generated interest among researchers, but the extent to which WM affects discourse production in PWA deserves further investigation. The present study aimed to address this gap by using a cross-sectional design to examine data collected from a sample larger than is typically included in aphasia studies. The dataset comprised narrative samples of retellings of the children's story *Cinderella* and tests scores of WM. The *Cinderella* story can serve as a reliable discourse elicitation prompt because of its high level of familiarity among participants. The advantage in using monologic narratives lies not only in the fact such discourse samples are the most common data elicitation method in aphasia studies (Linnik et al., 2016). They also offer a window into the production of language in context that, even if somewhat artificial, neutralizes the complex effects of interactive processes that occur in natural conversation (e.g., Beeke, 2012).

In addition, monologic narratives allow for systematic coding and analysis of narrative content in terms of micro- (words and sentences) and macro- (groups of sentences) linguistic components, known to demonstrate multi-level discourse ability (Cherney, Coelho, & Shadden, 1998). The specific measures we chose for the current analysis (see Methods) were aimed to capture discourse patterns that might be sensitive to impairment in the episodic buffer, largely based on Henderson et al.'s (2017) work. For micro-linguistic components, we reasoned that a measure of lexical diversity might be indicative of potential issues with lexical access which, as noted earlier, could be adversely affected by an impaired episodic buffer. Similarly, because the episodic buffer is thought to handle discourse segments that exceed word level material, we included a measure of grammatical complexity, to highlight potential dysfunction of linguistic operations at the sentential level. For macro-linguistic components, we considered the nature of local and global links expressed between episodes, which, if disrupted, could be explained in terms of impaired sequential and hierarchical processing.

With respect to evaluation of WM, we used scores on verbal span tests made available through the AphasiaBank, including verbal span at the single word level, sentential level, and content level (see Methods). These tests differ from one another in terms of the type of linguistic information that they manipulate and the memory load they impose. From the most basic ability to remember increasingly longer word lists, through the more demanding ability recall words in increasingly longer sentential stimuli, to the most difficult requirement to hold in memory semantic content of sentences. As noted earlier, because the episodic buffer presumably binds information from multiple sources into unitary episodes, span tests that require integration of “complex” information might be more sensitive to the buffer's integrity. In the context of the current study, these include span tasks manipulating sentence length and sentence content, in which consideration of syntactic and semantic contextual information necessarily goes beyond the processing of lexical items. Integration of such information during performance may speak to distinct “chunking principles” underpinning the efficiency with which the episodic buffer is assumed to operate.

Span performance at the word-level, in contrast, might be more sensitive to central executive skills than to the episodic buffer. This idea is based on the long-standing consensus that a span greater than three items taxes executive skills (Baddeley & Hitch, 1974). The assumption is that longer spans increase the resource demands placed on WM capacity and so impede the efficiency with which the central executive system operates. Performance on this test may also be indicative of the potential contribution of impaired short-term memory (STM) to narrative production in aphasia (for related comments, see Majerus et al., 2010), as verbal span scores have been systematically associated with STM deficits in PWA (e.g., Minika, Rosenberg, Klinyak-Fliszar, & Martin, 2017). Note that an account that solely focuses on STM effects on discourse production in aphasia would capture mostly issues with the temporary *storage* of information, rather than potential problems with the *manipulation* of the stored information. The latter would require a closer look at PWA's performance on measures of WM (for related comments, see Martin, Kohen, Kalinyak-Fliszar, Soveri, & Laine, 2012).

Thus, to characterize the potential effects of memory problems on discourse production in PWA, we propose examining WM impairments, with a focus on possible deficits in Baddeley's episodic buffer. This goal is extremely challenging, especially in the absence of established measures that specifically assess the functionality of the episodic buffer (see also Henderson et al., 2017). Nonetheless, in a study with a sample size as large as ours, analysis of even standard WM measures such as those made available through the AphasiaBank could help identify potential trends that could then be verified in a later study. Because it was impossible to make *a priori* assumptions about the degree to which each of the AphasiaBank tests taps the episodic buffer, our premise was that each of our three measures could potentially reflect distinct, even if interlinked, effects on discourse production in aphasia.

To examine the potential effects of WM on discourse production in PWA, we asked three research questions: (1) Does WM predict production of micro-level narrative components in PWA? (2) Does WM predict production of macro-level narrative components in PWA? (3) Are effects of WM on narrative production predicted by aphasia type? For questions (1) and (2), we predicted a positive association between WM and narrative components, largely based on the findings reported in Henderson et al. (2017). That is, we expected poor performance on WM tasks to predict compromised narrative production patterns in PWA, affecting both micro- and macro-level discourse components. We predicted that those with decreased verbal WM would also have lower lexical diversity and reduced grammatical complexity, as well as lesser local and global coherence and disrupted story grammars. Because the WM measures we used differed in terms of memory load, we considered the possibility that not all measures would follow the same pattern.

For question (3), we did not expect to find differences in discourse performance, based on aphasia type. The reason is twofold. First, descriptions of WM impairments in PWA are usually based on case studies or small samples that rarely reference aphasia type (see Cahana-Amitay & Albert, 2015, for a review), so it is not at all clear that a specific aphasia classification would necessarily involve a specific kind of WM deficit. Second, the aphasia literature is also vague on group distinctions in the production of macro-level discourse components, making it impossible to predict whether such differences would, in fact, emerge.

2. Methods

2.1. Participant data

Participant data were obtained through an online database known as the AphasiaBank (MacWhinney, Fromm, Forbes, & Holland, 2011). It comprises a collection of demographic data and different language tests from over 100 PWA made publicly available for the use of researchers worldwide. Data about each participant are de-identified. Of interest to the present study, are basic demographic data, including aphasia classification based on Western Aphasia Battery (WAB; Kertesz, 2007), samples of monologic narratives (i.e., *Cinderella*), and three tests assessing aspects of WM (e.g., word span).

For purposes of data analysis, we selected from the database only participants for whom complete data sets of relevance to our study aims were reported: working memory scores and narrative samples containing at least one story-related thematic unit (see Section 2.3.1). Age-matched healthy controls were not included in the sample because WM data were not collected for them. Our search yielded 15 participants with Wernicke's aphasia, to whom we then randomly assigned age- and gender-matched 15 people with Broca's aphasia and 15 people with Anomic aphasia (6 Female per group). This random assignment was done before any linguistic coding or analysis was started. Our final sample was thus composed of 45 participants, including 15 people with Anomia, 15 people with Broca's aphasia, and 15 people with Wernicke's. Aphasia type was determined, on the basis of participants' WAB scores. All participants were at least six months post onset and spoke English as their primary language.

2.2. Narrative elicitation

Each AphasiaBank participant in the sample was asked to retell the children's story of *Cinderella* from a wordless 36 panel storybook (Grimes, 2005). The book uses vignettes to outline the general story line. Participants were allowed to review the book for as long as they wished, before retelling the story. Once they said they were ready, the book was removed and they were asked to retell the story in their own words.

2.3. Discourse measures

Evaluation of narrative production was done using micro- and macro-level discourse measures, as detailed below. Six different measures of discourse structure were used to assess narrative production: A) three measures of micro-linguistic components: lexical diversity, grammatical complexity, and narrative length; B) three measures of macro-linguistic components: local coherence, global coherence, and story organization.

2.3.1. Micro-linguistic components

2.3.1.1. Lexical diversity. Lexical Diversity is used as a fractional index (i.e., 0–1) of lexical variation in a given language sample and is conventionally measured by Type-Token Ratio (TTR) (Richards, 1987). A high ratio (i.e., closer to 1) is thought to be indicative of more lexically diverse words and a larger vocabulary (Graessar, McNamara, Louwerse, & Cai, 2004). A common critique of using a simple TTR is that sample lengths can greatly influence such indices. This in turn can make comparisons between multiple speakers and samplings difficult to interpret (e.g., Chipere, Malvern, Richards, & Duran, 2001). Instead, Covington and McFall (2010) devised an automated program for averaging TTR across a "moving window" of words throughout the sample. This measure, known as Moving-Average-Type-Token Ratio (i.e., MATTR), has been shown to be more reliable across text of varying lengths. In this study, we thus used the MATTR, with a moving average window set for 10 words.

2.3.1.2. Grammatical complexity. A common measure of grammatical complexity in discourse is the Thematic Unit (T-Units), with each unit defined as a main clause with any subordinated clauses (Hunt, 1965). *Cinderella* story examples of T-Units containing zero, one, and two subordinated clauses are provided below. Subordinated clauses are highlighted in bold:

- Subordination of 0 clauses: *Cinderella was an only child*
- Subordination of 1 clause: *Cinderella was an only child* ***whose father remarried***
- Subordination of 2 clauses: *Cinderella,* ***who was very kind-hearted****, was an only child* ***whose father remarried***

In this study, grammatical complexity of each narrative was defined as the total number of subordinated clauses in each narrative. The higher the number of subordinated clauses in a narrative, the more grammatically complex it is.

2.3.1.3. Narrative length. Narrative Length is defined as the number of main clauses and any subordinated clauses (Hunt, 1965). This gives at least a coarse measure of the amount of information included in a narrative retelling of *Cinderella*.

2.3.2. Macro-linguistic components

2.3.2.1. Local coherence. Following Van Leer and Turkstra (1999), local coherence was defined as the meaning of any given T-Unit to the meaning of the immediately preceding T-Unit. Each T-Unit was ranked on how the topic of the previous T-Unit is continued, elaborated, or coordinated. The rating was based on a 5-point scale with the following ranking, ranging from least to most relevant: 1 - The current T-Unit doesn't continue the topic of the previous T-Unit, because it drastically shifts topic, comments on the discourse, or is unintelligible. 2 - The current T-Unit contains many clauses, wherein one clause generally refers to the previous T-Unit but is vague or shift the topic. 3 - The topic of the previous T-Unit is generally inferred, but may be vague or have a shift in topic. 4 - The current T-Unit contains many clauses, where one continues the topic of the previous T-Unit. 5 - The topic of the previous T-Unit is continued. Story examples inspired by C*inderella* are provided below:

- Ranking of 1: *Cinderella wanted to go to the ball. That's all I remember about the story.*
- Ranking of 2: *Cinderella wanted to go to the ball. I think she would be happy traveling there.*
- Ranking of 3: *Cinderella wanted to go to the ball. It would be a busy time.*
- Ranking of 4: *Cinderella wanted to go to the ball. She paused to consider ways to make a dress.*
- Ranking of 5: *Cinderella wanted to go to the ball. So she made a dress of rags to attend.*

Higher rankings reflect better local coherence.

2.3.2.2. Global coherence. This measure refers to the relationship between the content of any given T-Unit to the established topic of the entire discourse. Since this was a narrative re-telling of *Cinderella*, each T-Unit was ranked on how well it relates the overall story of *Cinderella*. For global coherence to be maintained throughout the narrative, there should be as many T-Units as possible relating to this topic. Similar to local coherence, global coherence we used a 5-point rating scale to rank each T-Unit for relevance to the overall topic, ranging from least to most related (Van Leer & Turkstra, 1999): 1 - The T-Unit doesn't have a relation to the general topic or is a comment on the discourse. 2 - The T-Unit contains many clauses, wherein one clause possibly relates to the topic and the other(s) does not. 3 - The T-Unit provides content that is possibly related to the general topic, is an evaluative statement, or must be inferred. 4 - The T-Unit contains many clauses, where one clause relates directly to the topic and one relates indirectly. 5 - The T-Unit provides content directly related to the general topic. Story examples inspired by *Cinderella* are provided below:

- Ranking of 1: *I don't really like this story that much.*
- Ranking of 2: *Cinderella wanted to meet this important guy, but can't figure how to do it.*
- Ranking of 3: *Cinderella was really crazy for her desires.*
- Ranking of 4: *Cinderella wanted to meet the prince, because she was really star-struck.*
- Ranking of 5: *Cinderella wanted to meet the prince.*

Again, higher rankings reflect better global coherence.

2.3.2.3. Story organization. Story Grammar was used as the primary measure of narrative organization (Lê, Coelho, Mozeiko, & Grafman, 2011). Each narrative typically includes a series of episodes that express the different events that occurred. These episodes have an internal structure composed of three components: an initiating event, in which a character is motivated to achieve a goal, an attempt, in which there is an attempt to accomplish that goal, and a direct consequence, which marks the success/failure at attaining the goal. Story examples inspired by *Cinderella* are provided below:

- Initiating event: *Cinderella wanted to go to the palace to meet the prince.*
- Attempt: *She made a dress of rags to attend the ball.*
- Direct consequence: *The stepmother ripped her dress apart so Cinderella couldn't go.*

A complete episode consisted of all three of these components in that order; episodes that contain fewer than three components were considered incomplete. Stories that are considered better organized have more complete episodes than incomplete ones, and indicative of better discourse planning (Merritt & Liles, 1986). In our analyses, we only considered the number of complete episodes produced.

2.3.3. Reliability

The second author coded all language samples for the aforementioned discourse measures. Additionally, a Research Assistant, trained on discourse coding, coded 40% of the transcripts (n = 18; 6 Anomic, 6 Broca, & 6 Wernicke) for linguistic measures. Inter-rater reliability was at least 80% for all measures. Disagreements were resolved through discussion on a case-by-case basis. Three months later, the second author coded a subset of the samples (20%) to establish intra-rater reliability, which exceeded 90% for all measures.

2.4. Working memory (WM) elicitation

The tests of working memory (WM) included three verbal WM repetition tasks. The proctor said a given word or sentence and each participant was instructed to repeat that word or sentence back to the proctor. The tests could be perceived as placing greater cognitive demands on performance because the repetitions range from a single word, through full sentences, to sentences of different types.

2.4.1. Word span (WM1)

Each participant was asked to orally repeat a word list of increasing length to a maximum of 8 words (e.g., 1. Food; 2. Food, Sock; 3. Food, Sock, Night, etc.). Words were judged to be correct if they were an approximate oration of the word by one phoneme (e.g., *bock* = *sock*), and in any given order (e.g., *food, sock, night* = *sock, night, food*).

2.4.2. Sentence span (WM2)

Each participant was asked to repeat a sentence several times, with each oration increasing in length by adding a new noun, preposition, qualifier, or verb phrase. For example: 1. *The bus is coming*; 2. *The tour bus is coming*, 3. *The tour bus is coming into town*, etc. The number of total words correctly produced for each sentence was tallied for this measure (maximum of 65).

2.4.3. Unique sentence span (WM3)

Each participant was asked to repeat twelve sentences provided by the proctor. Stimuli were equally divided into sentences the content of which showed no interference (e.g., *The dog chased the cat up the tree*), informational interference (e.g., *Count to ten as fast as you can*), or semantic interference (e.g., *Books like to read children*). The task was to repeat the sentences verbatim even if the content was unusual or nonsensical. If a participant repeated the words verbatim, this was thought to be indicative of good working memory (i.e., not distracted by the content of the sentence). The number of total words correctly produced over all sentences was tallied for this measure (maximum of 88).

2.5. Analytical plan

Using a linear regression model, we identified the predictors of each level of micro- (i.e., lexical diversity and sentential complexity) and macro- (i.e., local coherence, global coherence, story grammar) levels of discourse production. We used regression predictors of PWA from the three aphasia groups (i.e., Wernicke vs. Anomia vs. Broca) and measures of WM (i.e., WM1, WM2, WM3). The groups were dummy coded to compare group means against the Group of people with Wernicke's aphasia (i.e., constant; B_0). The differences between the group of people with Wernicke's aphasia and the other groups were coded in two other predictors of groups for the people with Anomia (G1) and those with Broca's aphasia (G2). Additionally, because narrative length may allow for added production of discourse components, we also included a measure of length (i.e., number of T-Units; TU) in the model. Since each PWA group may be differentially affected by WM, we included interactions between each group and each measure of WM:

Table 1
Descriptive statistics by group.

Measure	Group	N	Mean	SD	Range
Age	Anomic	15	66.20	12.60	41.4–83.1
	Broca's	15	63.10	10.20	39.0–78.3
	Wernicke's	15	65.90	13.30	42.6–91.7
Aphasia Quotient	Anomic	15	82.75	7.42	68.5–93.4
	Broca's	15	57.96	7.25	45.5–72.1
	Wernicke's	15	64.61	8.64	53.0–93.4
Narrative Length (Thematic-Units)	Anomic	15	28.00	20.18	3–69
	Broca's	15	20.33	11.91	6–50
	Wernicke's	15	37.07	30.70	6–126
Lexical Diversity (Moving-Average Type-Token Ratio)	Anomic	15	0.982	0.011	.955–.994
	Broca's	15	0.977	0.008	.961–.983
	Wernicke's	15	0.984	0.009	.964–.993
Grammatical Complexity (subordination)	Anomic	15	7.53	8.07	0–21
	Broca's	15	1.87	3.57	0–11
	Wernicke's	15	9.27	11.23	0–38
Complete Episodes	Anomic	15	2.73	1.67	0–6
	Broca's	15	0.68	0.82	0–3
	Wernicke's	15	2.07	1.76	0–7
Local Coherence	Anomic	15	2.71	0.72	1.290–3.824
	Broca's	15	2.01	0.52	1.422–3.322
	Wernicke's	15	2.32	0.60	1.421–3.701
Global Coherence	Anomic	15	3.80	0.52	3.322–4.614
	Broca's	15	2.68	0.53	1.676–3.674
	Wernicke's	15	3.20	0.70	2.112–4.487

$$Y = \text{Constant} + \beta_1{*}G1 + \beta_2{*}G2 + \beta_3{*}WM1 + \beta_4{*}WM2 + \beta_5{*}WM3 + \beta_6{*}(WM1{*}G1) + \beta_7{*}(WM1{*}G2) + \beta_8{*}(WM2{*}G1) + \beta_9{*}(WM2{*}G2) + \beta_{10}{*}(WM3{*}G1) + \beta_{11}{*}(WM3{*}G2) + TU + \text{error}$$

Results were analyzed using the Enter method in linear regression modeling in SPSS software (IBM Corp, 2016).

3. Results

Descriptive findings are presented in Table 1. Results show that narratives produced by people with aphasia of Broca's type are shorter, less grammatically complex, with fewer complete episodes, and reduced coherence. Additionally, almost all coefficients were significantly correlated with each other, suggesting that micro- and macro-production are indeed related (Table 2).

3.1. Lexical diversity (MATTR)

The regression model for lexical diversity was significant (F(12,32) = 3.386, $p < .05$), with an $R^2 = 0.559$ (Table 3). The

Table 2
Correlations between working memory and narrative length regression coefficients.

	G1	G2	WM1	WM2	WM3	WM1*G1	WM2*G2	WM2*G1	WM2*G2	WM3*G1	WM3*G2	TU
G1	1											
G2	**−0.5**	1										
WM1	**−0.457**	**0.305**	1									
WM2	**−0.645**	**0.303**	**0.584**	1								
WM3	**−0.736**	**0.463**	**0.556**	**0.866**	1							
WM1*G1	**0.855**	**−0.401**	**−0.606**	**−0.537**	**−0.593**	1						
WM1*G2	**−0.586**	**0.86**	**0.683**	**0.482**	**0.566**	**−0.608**	1					
WM2*G1	**0.917**	**−0.466**	**−0.404**	**−0.623**	**−0.627**	**0.85**	**−0.545**	1				
WM2*G2	**−0.643**	**0.8**	**0.458**	**0.405**	**0.471**	**−0.535**	**0.826**	**−0.711**	1			
WM3*G1	**0.952**	**−0.442**	**−0.398**	**−0.567**	**−0.633**	**0.856**	**−0.531**	**0.956**	**−0.659**	1		
WM3*G2	**−0.664**	**0.786**	**0.416**	**0.347**	**0.496**	**−0.55**	**0.814**	**−0.687**	**0.959**	**−0.681**	1	
TU	0.015	0.254	0.141	−0.11	−0.121	−0.063	0.269	0.009	0.121	0.005	0.067	1

Note: G1 & G2: Group differences.
WM1: Working Memory Word Span Measure.
WM2: Working Memory Sentence Span Measure.
WM3: Working Memory Unique Sentence Span Measure.
TU: T-Units (Narrative Length).
Significant correlations are in bold.

Table 3
Lexical diversity regression coefficients.

	B	SE	t	p
(Constant)	**0.978**	**0.005**	**194.086**	**< .001**
G1	−0.01	0.01	−0.932	0.358
G2	−0.004	0.009	−0.406	0.688
WM1	−0.000129	0.002	−0.077	0.939
WM2	5.22E-05	0.00018	0.278	0.783
WM3	-.000141	0.000124	−1.141	0.262
WM1*G1	.000430	0.003	0.168	0.867
WM1*G2	0.005	0.005	0.901	0.374
WM2*G1	−7.10E-05	0.000385	−0.184	0.855
WM2*G2	−0.000154	0.001	−0.29	0.774
WM3*G1	0.000144	0.000224	0.641	0.526
WM3*G2	7.52E-05	0.000341	0.221	0.827
TU	**0.000247**	**0.000061**	**4.021**	**< .001**

G1 & G2: Group differences.
WM1: Working Memory Word Span Measure.
WM2: Working Memory Sentence Span Measure.
WM3: Working Memory Unique Sentence Span Measure.
TU: T-Units (Narrative Length).
Significant predictors are in bold.

analysis revealed that only the constant (Beta = 0.978, t(44) = 194.086, p < .001) and the T-Unit (Beta = 0.00247, t(44) = 4.021, p < .001) coefficients were significant predictors of MATTR scores. Based on the significance of the constant, confidence intervals were constructed to compare group means. Post hoc analysis revealed that only the Anomic group was significantly greater than the Broca's group (Table 4). This suggests that narrative length and group membership, to some degree, were indicative of higher lexical diversity, and not working memory.

3.2. Sentential complexity

A significant regression model for sentential subordination was found (F(12,32) = 8.975, p < .001), with an $R^2 = 0.771$ (Table 5). This analysis showed that the T-Units coefficient was the only significant predictor in this model (Beta = 0.310, t(44) = 7.988, p < .001). This result is aligned with the idea that the amount of language produced influences the productivity of micro-linguitic components as measured by sentential subordination. All other predictors were non-significant.

3.3. Narrative organization

There was a significant regression model for complete narrative episodes (F(12,32) = 6.860, p < .001), with an $R^2 = 0.720$ (Table 6). Length was a significant predictor of organization (Beta = 0.043, t(44) = 5.140, p < .001), and also a nearing significant interaction effects of the group difference between Wernicke's and Anomic groups (i.e., G1) for the measures of first two working memory measures (WM1*G1, Beta = 0.687, t(44) = 1.983, p = .056; & WM2*G1, Beta = −0.108, t(44) = −2.070, p = .047). However, post hoc analysis to compare interaction effects for groups (Anomic vs. Broca's vs. Wernicke's) with the first two WM measures (Table 7) revealed no significant comparisons.

3.4. Local coherence

The analysis for Local Coherence showed a significant model (F(12,32) = 2.229, p < .05), with an $R^2 = 0.455$ (Table 8). The analysis only found the constant (Beta = 2.061, t(44), p < .001) and a significant interaction effect between the first working memory measure for Anomic and Wernicke's (WM1*G1; Beta = 0.495, t(44) = 2.579, p < .05), as well as a marginal one for the

Table 4
Confidence intervals for lexical diversity between groups.

	Estimate	SE	T Stat	Critical Value	CI-Low	CI-High
Anomic vs. Wernicke's	−0.010	−10.000	2.488	0.025	−0.035	0.015
Broca's vs. Wernicke's	−0.004	−0.004	2.488	0.022	−0.026	0.018
Anomic vs. Wernicke's	**−0.096**	**−0.096**	**2.488**	**0.026**	**−0.122**	**−0.071**

Significant differences are in bold.

Table 5
Sentential complexity (subordination) regression coefficients.

	B	SE	t	p
(Constant)	−2.008	3.183	−0.631	0.533
G1	4.635	6.473	0.716	0.479
G2	3.328	5.675	0.586	0.562
WM1	−0.529	1.051	−0.503	0.618
WM2	−0.068	0.119	−0.571	0.572
WM3	0.046	0.078	0.589	0.56
WM1*G1	−0.232	1.614	−0.144	0.887
WM1*G2	0.147	3.189	0.046	0.963
WM2*G1	−0.198	0.243	−0.812	0.423
WM2*G2	−0.002	0.336	−0.005	0.996
WM3*G1	0.008	0.142	0.055	0.957
WM3*G2	−0.071	0.215	−0.328	0.745
TU	**0.31**	**0.039**	**7.988**	**< .001**

G1 & G2: Group differences.
WM1: Working Memory Word Span Measure.
WM2: Working Memory Sentence Span Measure.
WM3: Working Memory Unique Sentence Span Measure.
TU: T-Units (Narrative Length).
Significant predictors are in bold.

Table 6
Narrative organization (# of complete episodes) regression coefficients.

	B	SE	t	p
(Constant)	−0.046	0.683	−0.068	0.947
G1	0.611	1.389	0.44	0.663
G2	0.176	1.218	0.144	0.886
WM1	0.228	0.226	1.011	0.32
WM2	−0.033	0.025	−1.291	0.206
WM3	0.013	0.017	0.795	0.432
WM1*G1	**0.687**	**0.346**	**1.983**	**0.056**
WM1*G2	0.907	0.684	1.325	0.195
WM2*G1	**−0.108**	**0.052**	**−2.07**	**0.047**
WM2*G2	−0.079	0.072	−1.091	0.283
WM3*G1	−0.006	0.03	−0.203	0.84
WM3*G2	−0.007	0.046	−0.149	0.883
TU	**0.043**	**0.008**	**5.14**	**< .001**

G1 & G2: Group differences.
WM1: Working Memory Word Span Measure.
WM2: Working Memory Sentence Span Measure.
WM3: Working Memory Unique Sentence Span Measure.
TU: T-Units (Narrative Length).
Significant predictors are in bold.

Table 7
Confidence intervals for narrative organization for interaction effects between groups and word span and sentence span.

		Estimate	SE	T Stat	Critical Value	CI-Low	CI-High
WM1	Anomic vs. Wernicke's	0.687	0.346	2.770	0.958	−0.271	1.645
	Broca's vs. Wernicke's	0.907	0.684	2.770	1.984	−1.077	2.891
	Anomic vs. Wernicke's	−0.220	0.944	2.770	2.338	−2.558	2.118
WM2	Anomic vs. Wernicke's	−0.108	0.052	2.770	0.144	−0.252	0.036
	Broca's vs. Wernicke's	−0.079	0.072	2.770	0.199	−0.287	0.129
	Anomic vs. Wernicke's	−0.029	0.089	2.770	0.247	−0.276	0.218

WM1: Working Memory Word Span Measure.
WM2: Working Memory Sentence Span Measure.

second working memory measure for Broca's and Wernicke's (WM2*G2; Beta = −0.080, t(44) = −1.997, = 0.054). However, post hoc analysis comparing group (Anomic vs. Broca's vs. Wernicke's) and interaction effects of group and working memory (WM1) did not show any significant effects after controlling for multiple comparisons (Table 9).

Table 8
Local coherence regression coefficients.

	B	SE	t	p
(Constant)	**2.061**	**0.379**	**5.439**	**< .001**
G1	−1.039	0.77	−1.348	0.187
G2	0.138	0.676	0.204	0.84
WM1	0.074	0.125	0.592	0.558
WM2	−0.017	0.014	−1.222	0.23
WM3	0.01	0.009	1.029	0.311
WM1*G1	**0.495**	**0.192**	**2.579**	**0.015**
WM1*G2	0.341	0.38	0.897	0.376
WM2*G1	−0.045	0.029	−1.547	0.132
WM2*G2	**−0.08**	**0.04**	**−1.997**	**0.054**
WM3*G1	0.008	0.017	0.486	0.63
WM3*G2	0.028	0.026	1.082	0.287
TU	0.005	0.005	1.114	0.274

G1 & G2: Group differences.
WM1: Working Memory Word Span Measure.
WM2: Working Memory Sentence Span Measure.
WM3: Working Memory Unique Sentence Span Measure.
TU: T-Units (Narrative Length).
Significant predictors are in bold.

Table 9
Confidence intervals for local coherence between groups and interaction effect between group and word span and sentence span.

		Estimate	SE	T Stat	Critical Value	CI-Low	CI-High
Group	Anomic vs. Wernicke's	−1.039	0.77	2.89	2.23	−3.269	1.191
	Broca's vs. Wernicke's	0.138	0.676	2.89	1.95	−1.812	2.088
	Anomic vs. Wernicke's	−1.177	1.07	2.89	3.09	−4.267	1.913
WM1* Group	Anomic vs. Wernicke's	0.495	0.192	2.89	0.555	−0.06	1.05
	Broca's vs. Wernicke's	0.341	0.38	2.89	1.098	−0.757	1.439
	Anomic vs. Wernicke's	0.154	0.396	2.89	1.058	−0.904	1.212
WM2* Group	Anomic vs. Wernicke's	−0.045	0.029	2.89	0.084	−0.129	0.039
	Broca's vs. Wernicke's	−0.08	0.04	2.89	0.116	−0.196	0.036
	Anomic vs. Wernicke's	−0.125	0.077	2.89	0.222	−0.347	0.097

WM1: Working Memory Word Span Measure.
WM2: Working Memory Sentence Span Measure.
WM3: Working Memory Unique Sentence Span Measure.

3.5. Global coherence

The analysis for Global Coherence revealed a significant predictive model ($F(12,32) = 3.728$, $p < .01$), with an $R^2 = 0.583$ (Table 10). Additionally, the model had a significant constant predictor (Beta = 0.363, $t(44) = 6.545$, $p < .001$), as well as a marginally significant coefficients of the first working memory measure (WM1; Beta = 0.120, $t(44) = 2.013$, $p = .053$), the first working measure between Wernicke's and Anomic groups (WM1*G1; Beta = 0.184, $t(44) = 1.854$, $p = .073$), and the second working measure between Wernicke's and the Broca's groups (WM2*G2; Beta = 0.038, $t(44)$ -1.699, $p = .099$; Table 11). Post hoc analysis did not reveal any significant effects after controlling for multiple comparisons between groups and WM measures (Table 11).

4. Discussion

In this study, we examined the potential effects of WM on discourse production in PWA, by asking two research questions: (1) Does WM predict production of micro- and macro-level narrative components in PWA? (2) Are effects of WM on narrative production predicted by aphasia type? Our findings offer preliminary evidence to both: some measures of WM span do predict certain aspects of narrative production in PWA, and these effects may be sensitive to aphasia type. These WM effects are consistent with the general claim that the performance of open-ended discourse tasks is mediated, at least, to some extent, by cognitive abilities (e.g., Alexander, 2006). It is important to note that the three WM measures we used (WM1, WM2, WM3) were highly correlated with one another, limiting the extent to which we can attribute WM effects on discourse in aphasia to any individual WM measure. However, as we point in the Introduction, given our larger than usual sample size, the trends we report here can offer a foundation for future studies of WM effects on discourse aphasia using more refined measures.

The specific WM effects that we observed involved two of three WM measures included in this study: word span (WM1) and sentence span (WM2). Poor performance on WM1 can be taken to reflect an executive dysfunction, as noted in Baddeley and Hitch

Table 10
Global coherence regression coefficients.

	B	SE	t	p
(Constant)	**2.379**	**0.363**	**6.545**	**< .001**
G1	0.32	0.739	0.433	0.668
G2	0.898	0.648	1.386	0.175
WM1	**0.241**	**0.12**	**2.013**	**0.053**
WM2	0.005	0.014	0.364	0.718
WM3	−0.002	0.009	−0.225	0.824
WM1*G1	**0.342**	**0.184**	**1.854**	**0.073**
WM1*G2	−0.042	0.364	−0.116	0.909
WM2*G1	−0.018	0.028	−0.655	0.517
WM2*G2	**−0.065**	**0.038**	**−1.699**	**0.099**
WM3*G1	−0.019	0.016	−1.163	0.253
WM3*G2	0.035	0.025	1.44	0.16
TU	−0.001	0.004	−0.199	0.844

G1 & G2: Group differences.
WM1: Working Memory Word Span Measure.
WM2: Working Memory Sentence Span Measure.
WM3: Working Memory Unique Sentence Span Measure.
TU: T-Units (Narrative Length).
Significant predictors are in bold.

Table 11
Confidence intervals for global coherence between groups.

	Estimate	SE	T Stat	Critical Value	CI-Low	CI-High
Anomic vs. Wernicke's	0.320	0.739	2.488	2.490	−1.520	2.160
Broca's vs. Wernicke's	0.898	0.648	2.488	2.490	−0.712	2.508
Anomic vs. Wernicke's	−0.578	0.910	2.488	2.490	−2.843	1.687

(1974) (see Introduction). Thus, we suggest that the challenges PWA face when producing narratives involve, to some degree, a deficit in Baddeley's central executive system. The implication of this finding for the episodic buffer, as conceived by Henderson et al. (2017), would be an impairment in the chunking of story propositions for temporary storage, which at least partly relies on executive functions that govern the overall goal of communicating a story effectively. These findings lend support to an independent body of literature suggesting that the complex array of utterances produced in discourse are interlinked with executive skills that allow for intentionality, planning, goal attainment necessary for accomplishing successful functional communication (Alexander, 2006; Ellis et al., 2016).

Span performance at the sentence level—WM2—likely provides more direct evidence for the potential role of the episodic buffer in narrative production in aphasia. We suggest that WM2 taps the extent to which a person can leverage syntactic structure as a gateway to the creation of story clusters within the episodic buffer. The sentential scope facilitates systematic access to information relevant to the story at hand, offering a first-pass structural frame into which narrative content is packaged. Because the stimuli in WM2 are limited to simple sentences, we propose a more basic chunking unit than a sentence: the *clause*. This idea is consistent with long-standing claims about the function of clauses in children's narrative discourse (e.g., Berman, 2008), which children learn to manipulate over time to convey increasingly more complex discourse information (e.g., Verhoeven et al., 2002). It is also in line with Sherratt's (2007) multi-level approach to discourse analysis in people with aphasia, in which the clause is viewed as an important building block of discourse. Thus, damage to the episodic buffer in PWA can potentially lead to difficulties in harnessing clause structure, which is a prerequisite for telling a well-formed story. The observation that poor WM2 is detrimental macro-level narratives components should come as no surprise, as local and global coherence represent interclausal links.

The absence of WM effects associated with our third WM measure—unique sentence span (WM3)—is likely attributable to the fact that it distinctly taps sentence content, implicating a somewhat different cognitive mechanism than those underpinning the other WM measures. This pattern suggests that the episodic buffer among PWA retains its ability to integrate semantic concepts into propositions to create a narrative. We argue that because WM3 taps the manipulation of semantic content within a sentence, the absence of its effects in our sample suggests that PWA are, in fact, capable of manipulating such semantic information. That is, their episodic buffer retains its ability to integrate semantic concepts into propositions and the deficits observed in their narratives likely stem from an impairment of a different "chunking principle." If PWA retain their ability to engage in such multimodal semantic processing, they may also be able to integrate semantic information from LTM with phonological information, sparing lexical access at the micro-linguisitc level. Such a scenario might explain why we failed to find effects of reduced memory span on micro-level discourse components in the current sample.

However, the absence of WM effects on micro-level discourse components that we report here remains unexpected, especially given that production of macro-level discourse components among PWA has been shown to be tightly related to that of micro-level discourse components (e.g., Andreetta et al., 2012; Wright & Capilouto, 2012). For example, Wright and Capilouto (2012) found that lexical diversity in PWA reliably predicted their coherence scores, explaining slightly more than half of the variance. One possibility is that the familiarity of the well-known children's story *Cinderella* masked the range of semantic difficulties that PWA may experience in less-constrained discourse production contexts. We expected that the AphasiaBank's specific administration protocol comprised 36 panels, a much longer narrative elicitation probe than the typical 1–5 panels used in picture description tasks, participants would offer participants ample opportunity to produce non-automatized semantic content. It is possible, however, that the high number of pictures in the book that participants were expected to reference in their retelling failed to impose a sufficiently WM demand on semantic processing, as a result of the overall accessibility of story content to the participants.

Although surprising, our results are consistent with other non-narrative discourse studies of functional communication among PWA, in which macro-rather than micro-level discourse components have been found to be affected by domain-general executive functions often evaluated with tests of working memory (Ramsberger, 1994; 2005). Roughly speaking, these studies indicate that sparing of executive functions aids the success of a conversational exchange even if a PWA suffers severe word-finding deficits (Irwin, Wertz, & Avent, 2002; Ramsberger & Rende, 2002). Other findings show that damage to executive systems limits the ability of PWA to use alternative modes of communication such as gesture (Purdy & Koch, 2006; Purdy, Duffy, & Coelho, 1994), increases linguistic perseveration, reduces shifts in conversational strategy (Frankel, Penn, & Ormond-Brown, 2007), and diminishes conversational independence and quality (Fridriksson, Nettles, Davis, Morrow, & Montgomery, 2006). Interestingly, Henderson et al. (2017) have proposed that the ability to shift from one conversational strategy to another to make up for lexical retrieval deficits, is linked to impairment in the episodic buffer. Their assumption is that PWA have difficulty integrating information from different modes of communication into a single, unified representation.

The WM effects on macro-level discourse components that we observed in this study were apparent in all measures of narrative production, although those affecting global coherence were more robust than those impacting local coherence. These findings suggest that impairment in the episodic buffer of PWA adversely affects, to a limited extent, their ability to engage in hierarchical processing, in which the relationship between the overall discourse topic and ongoing information is assessed. The primary challenge they face involves the ways in which they combine information sequentially in their narratives. Such problems could disrupt their ability to link propositions dictated by story grammar organization and interfere with the production of cohesive and coherent discourse. This observation is aligned with Wright and Capilouto's (2012) claim that macro-level discourse performance is dependent on cognitive factors, such as the order of presentation of elements in the story, and with independent aphasia studies describing deficits in sequential processing among participants (Bond, Ulatowska, Macaluso-Haynes, & May, 1983; Chapman & Ulatowska, 1989; Gleason et al., 1980).

Related results have also been reported by Manning and Franklin (2016), who demonstrated that PWA, but not age-matched controls, had problems with temporal sequencing in their narrative production. The deficits they reported were not correlated with impaired micro-level components, comparable to our findings, but they did not find a link between narrative performance and scores on cognitive screening, which according to them, were within normal limits for all participants. It is difficult to ascertain whether the participants' presumed spared cognition included scores on tests of WM, as there is no mention in the study of which cognitive tests were included in the evaluations. The authors maintained, though, that the temporal problems PWA exhibited likely reflect a broader extra-linguistic deficit, because in some participants, they found a double dissociation between sequencing and naming performance. This finding, while interesting, is difficult to interpret, as it leaves open the question of what cognitive mechanism might be implicated in the discourse deficit.

In terms of group differences, we found that aphasia classification may be weakly related to WM effects on discourse performance. Our aim here was not to offer an exhaustive analysis of the association of WM and discourse in aphasia, but rather to leverage the larger than usual sample of participants available through the AphasiaBank, to allow for some group comparisons. Because of the limited literature on WM and discourse in aphasia, the types of aphasia we focused on are the more commonly known ones. We observed that people with anomia outperform those with more severe aphasia types. These effects disappeared in our posthoc analysis, leaving open the question of whether aphasia type is, in fact, sensitive to WM-discourse interdependency.

In summary, this study demonstrates that WM impairment in PWA adversely affects their ability to produce macro-linguistic narrative components. Individuals with reduced word and sentential spans have trouble creating narrative-appropriate local and global coherence and well-organized story-grammar, possibly reflecting an episodic buffer disruption that impedes multimodal integration of information that expresses sequential and hierarchical information. The degree to which these patterns are sensitive to clinical classification of aphasia remains underdetermined. Because the patterns we observed derive from a larger than usual sample, our study provides a good starting point for the design of larger-scale research that would examine the neural underpinnings of the effects of WM on discourse production in aphasia.

Competing interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

We wish to thank AphasiaBank, Michela Sguera for her reliability coding, and Dr. Hariharan Swaminathan for his statistical consulting.

References

Alexander, M. P. (2006). Impairments of procedures for implementing complex language are due to disruption of frontal attention processes. *Journal of the International Neuropsychological Society, 12*(2), 236–247.

Andreetta, S., Cantagallo, A., & Marini, A. (2012). Narrative discourse in anomic aphasia. *Neuropsychologia, 50*(8), 1787–1793.

Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology, 14*(9), 875–892.

Armstrong, E., Ferguson, A., & Simmons-Mackie, N. (2013). Discourse and functional approaches to aphasia. In I. Papthanasiou, P. Coppens, & C. Potagas (Eds.). *Aphasia and related neurogenic communication disorders* (pp. 217–232). Boston, MA: Jones and Bartlett Publishers.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829–839.

Baddeley, A. D. (2007). Working memory: Multiple models, multiple mechanisms. In H. L. Roediger, Y. Dudai, & S. M. Fitzpatrick (Eds.). *Science of memory: Concepts* (pp. 151–153). Oxford: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8*, 47–89.

Baldo, J. V., & Dronkers, N. F. (2006). The role of inferior parietal and inferior frontal cortex in working memory. *Neuropsychology, 20*(5), 529–538.

Beeke, S. (2012). Aphasia: The pragmatics of everyday conversation. In H. J. Schmid (Ed.). *Cognitive pragmatics* (pp. 345–371). Berlin, Germany: Mouton de Gryter.

Beeson, P. M., Bayles, K. A., Rubens, A. B., & Kaszniak, A. W. (1993). Memory impairment and executive control in individuals with stroke-induced aphasia. *Brain and Language, 45*(2), 253–275.

Berman, R. (2008). The Psycholinguistics of developing text construction. *Journal of Child Language, 35*, 735–771.

Bloom, R. L., Borod, J. C., Santschi-Haywood, C., Pick, L. H., & Obler, L. K. (1996). Left and right hemispheric contributions to discourse coherence and cohesion. *International Journal of Neuroscience, 88*(1–2), 125–140.

Bond, S. L., Ulatowska, H. K., Macaluso-Haynes, S., & May, E. B. (1983). Discourse production in aphasia: Relationship to severity of impairment. In R. H. Brookshire (Ed.). *Proceedings of the clinical aphasiology conference* (pp. 202–210). Minneapolis, MN: BRK Publishers.

Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics and Phonetics, 30*(7), 489–518.

Byng, S., & Black, M. (1989). Some aspects of sentence production in aphasia. *Aphasiology, 3*(3), 241–263.

Cahana-Amitay, D., & Albert, M. L. (2015). *Redefining recovery from aphasia.* New York: Oxford University Press.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences, 22*(1), 77–94.

Carpenter, P. A., & Just, M. A. (1989). The role of working memory in language comprehension. In K. Rayner (Ed.). *Complex information processing: The impact of Herbert A. Simon* (pp. 31–68). Hillsdale, NJ: Erlbaum.

Carpenter, P. A., Just, M. A., & Reichle, E. D. (2000). Working memory and executive function: Evidence from neuroimaging. *Current Opinion in Neurobiology, 10*(2), 195–199.

Caspari, I., Parkinson, S. R., LaPointe, L. L., & Katz, R. C. (1998). Working memory and aphasia. *Brain and Cognition, 37*(2), 205–223.

Chapman, S. B., & Ulatowska, H. K. (1989). Discourse in aphasia: Integration deficits in processing reference. *Brain and Language, 36*(4), 651–668.

Cherney, L. R., Coelho, C. A., & Shadden, B. B. (1998). *Analyzing discourse in communicatively impaired adults.* Gaithersburg, MY: Aspen Publishers.

Chipere, N., Malvern, D., Richards, B., & Duran, P. (2001). Using a corpus of school children's writing to investigate the development of vocabulary diversity. In P. Rayson, A. Wilson, T. McEnery, A. Hardie, & S. Khoja (Eds.). *Proceedings of the corpus linguistics 2001 conference* (pp. 126–133). Lancaster, UK: University Centre for Computer Corpus Research on Language Technical Papers.

Christensen, S. C., & Wright, H. H. (2010). Verbal and non-verbal working memory in aphasia: What three n-back tasks reveal. *Aphasiology, 24*(6–8), 752–762.

Code, C. (2010). Aphasia. In J. S. Damico, N. Müller, & M. J. Ball (Eds.). *The handbook of language and speech disorders* (pp. 317–336). West Sussex, UK: Wiley-Blackwell.

Connor, L. T., MacKay, A. J., & White, D. A. (2000). Working memory: A foundation for executive abilities and higher-order cognitive skills. In G. Ramsberger (Vol. Ed.), *Executive functions: Description, disorders, assessment, management. Seminars in Speech and Language: Vol. 21*, (pp. 109–120).

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type–token ratio (MATTR). *Journal of Quantitative Linguistics, 17*(2), 94–100.

Davidson, B., Worrall, L., & Hickson, L. (2003). Identifying the communication activities of older people with aphasia: Evidence from naturalistic observation. *Aphasiology, 17*(3), 243–264.

Doyle, P. J., McNeil, M. R., Park, G., Goda, A., Rubenstein, E., Spencer, K., et al. (2000). Linguistic validation of four parallel forms of a story retelling procedure. *Aphasiology, 14*, 537–549.

Ellis, C., Henderson, A., Wright, H. H., & Rogalski, Y. (2016). Global coherence during discourse production in adults: A review of the literature. *International Journal of Language & Communication Disorders, 51*(4), 359–367.

Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology, 25*(11), 1414–1430.

Frankel, T., Penn, C., & Ormond-Brown, D. (2007). Executive dysfunction as an explanatory basis for conversation symptoms of aphasia: A pilot study. *Aphasiology, 21*(6–8), 814–828.

Fridriksson, J., Nettles, C., Davis, M., Morrow, L., & Montgomery, A. (2006). Functional communication and executive function in aphasia. *Clinical Linguistics and Phonetics, 20*, 401–410.

Friedmann, N., & Gvion, A. (2003). Sentence comprehension and working memory limitation in aphasia: A dissociation between semantic-syntactic and phonological reactivation. *Brain and Language, 86*(1), 23–39.

Gleason, J. B., Goodglass, H., Obler, L., Green, E., Hyde, M. R., & Weintraub, S. (1980). Narrative strategies of aphasic and normal-speaking subjects. *Journal of Speech & Hearing Research, 23*(2), 370–382.

Glosser, G., & Deser, T. (1991). Patterns of discourse production among neurological patients with fluent language disorders. *Brain and Language, 40*(1), 67–88.

Graessar, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments & Computers, 36*(2), 193–202.

Grimes, N. (2005). *Walt Disney's Cinderella.* New York: Random House.

Gvion, A., & Friedmann, N. (2012). Does phonological working memory impairment affect sentence comprehension? A study of conduction aphasia. *Aphasiology, 26*(3–4), 494–535.

Halliday, M. A., & Christian, M. I. M. (2004). *An introduction to functional grammar.* London: Edward Arnold.

Henderson, A., Kim, H., Kintz, S., Frisco, N., & Wright, H. H. (2017). Working memory in Aphasia: Considering discourse processing and treatment implications. *Seminars in Speech and Language, 38*(01), 40–51.

Hunt, K. (1965). *Differences in grammatical structures written at three grade levels (NCTE; research report No. 3)*Urbana, IL: National Council of Teachers of English.

IBM Corp. Released (2016). *IBM SPSS statistics for windows, version 24.0.* New York: IBM Corp.

Irwin, W. H., Wertz, R. T., & Avent, J. R. (2002). Relationships among language impairment, functional communication, and pragmatic performance in aphasia. *Aphasiology, 16*(8), 823–835.

Kertesz, A. (2007). *Western aphasia battery (Revised).* San Antonio: PsychCorp.

Kong, A. P. H. (2016). *Analysis of neurogenic disordered discourse production: From theory to practice.* New York, NY: Routledge Psychology Press.

Laures-Gore, J. S., Marshall, R. M., & Verner, E. (2011). Digit span differences in aphasia and right brain damage. *Aphasiology, 25*(1), 43–56.

Lê, K., Coelho, C., Mozeiko, J., & Grafman, J. (2011). Measuring goodness of story narratives. *Journal of Speech, Language, and Hearing Research, 54*(1), 118–126.

Lemme, M. L., Hedberg, N. L., & Bottenberg, D. E. (1984). Cohesion in narratives of aphasic adults. In R. H. Brookshire (Ed.). *Clinical Aphasiology: Conference proceedings* (pp. 215–222). Minneapolis, MN: BRK Publishers.

Linnik, A., Bastiaanse, R., & Höhle, B. (2016). Discourse production in aphasia: A current review of theoretical and methodological challenges. *Aphasiology, 30*(7), 765–800.

Lock, S., & Armstrong, L. (1997). Cohesion analysis of the expository discourse of normal, fluent aphasic and demented adults: A role in differential diagnosis? *Clinical Linguistics and Phonetics, 11*(4), 299–317.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*(11), 1286–1307.

Majerus, S., D'argembeau, A., Perez, T. M., Belayachi, S., Van der Linden, M., Collette, F., ... Maquet, P. (2010). The commonality of neural networks for verbal and visual short-term memory. *Journal of Cognitive Neuroscience, 22*, 2570–2593.

Manning, M., & Franklin, S. (2016). Cognitive grammar and aphasic discourse. *Clinical Linguistics and Phonetics, 30*(6), 417–432.

Martin, N., Kohen, F., Kalinyak-Fliszar, M., Soveri, A., & Laine, M. (2012). Effects of working memory load on processing of sounds and meanings of words in aphasia. *Aphasiology, 26*(3–4), 462–493.

Merritt, D. D., & Liles, B. Z. (1986). Story grammar ability in children with and without language disorder. *Journal of Speech, Language, and Hearing Research, 30*(4), 539–552.

Minika, I., Rosenberg, S., Klinyak-Fliszar, M., & Martin, N. (2017). Short-term memory and aphasia: From theory to treatment. *Seminars in Speech and Language, 38*(1), 17–18.

Murray, L. L. (2012). Attention and other cognitive deficits in aphasia: Presence and relation to language and communication measures. *American Journal of Speech-Language Pathology, 21*(2), S51–S64.

Nicholas, M., Obler, L. K., Albert, M. L., & Helm-Estabrooks, N. (1985). Empty speech in Alzheimer's disease and fluent aphasia. *Journal of Speech & Hearing Research, 28*(3), 405–410.

Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech, Language, and Hearing Research, 42*, 626–648.

Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2017). Reviewing the quality of discourse information measures in aphasia. *International Journal of Language & Communication Disorders*. http://dx.doi.org/10.1111/1460-6984.12318.

Purdy, M., Duffy, R., & Coelho, C. (1994). An investigation of the communicative use of trained symbols in aphasic adults following multimodality training. In P. Lemme (Ed.). *Clinical Aphasiology: Conference proceedings* (pp. 345–356). Austin, TX: Pro-Ed.

Purdy, M., & Koch, A. (2006). Prediction of strategy usage by adults with aphasia. *Aphasiology, 20*(2–4), 337–348.

Ramsberger, G. (1994). Functional perspective for assessment and rehabilitation of persons with severe aphasia. *Seminars in Speech and Language, 15*(1), 1–16.

Ramsberger, G. (2005). Achieving conversational success in aphasia by focusing on non-linguistic cognitive skills: A potentially promising new approach. *Aphasiology, 19*, 1066–1073.

Ramsberger, G., & Rende, B. (2002). Measuring transactional success in the conversation of people with aphasia. *Aphasiology, 16*(3), 337–353.

Richards, B. (1987). Type/token ratios: What do they really tell us? *Journal of Child Language, 14*(2), 201–209.

Rönnberg, J., Larsson, C., Fogelsjöö, A., Nilsson, L. G., Lindberg, M., & Ängquist, K. A. (1996). Memory dysfunction in mild aphasics. *Scandinavian Journal of Psychology, 37*(1), 46–61.

Sherratt, S. (2007). Multi-level discourse analysis: A feasible approach. *Aphasiology, 21*(3–4), 375–393.

Sung, J. E., McNeil, M. R., Pratt, S. R., Dickey, M. W., Hula, W. D., & Szuminsky, N. J. (2009). Verbal working memory and its relationship to sentence-level reading and listening comprehension in persons with aphasia. *Aphasiology, 23*(7–8), 1040–1052.

Ulatowska, H. K., North, A. J., & Macaluso-Haynes, S. (1981). Production of narrative and procedural discourse in aphasia. *Brain and Language, 13*(2), 345–371.

Van Leer, E., & Turkstra, L. (1999). The effect of elicitation task on discourse coherence and cohesion in adolescents with brain injury. *Journal of Communication Disorders, 32*(5), 327–349.

Verhoeven, L., Aparici, M., Cahana-Amitay, D., Hell, J., Kriz, S., & Viguié, A. (2002). Clause packaging: A cross-linguistic developmental analysis. *Written Language & Literacy, 5*(2), 135–162.

Ween, J. E., Verfaellie, M., & Alexander, M. P. (1996). Verbal memory function in mild aphasia. *Neurology, 47*(3), 795–801.

Whitworth, A. (2010). Using narrative as a bridge: Linking language processing models with real-life communication. *Seminars in Speech and Language, 31*(1), 64–76.

Whitworth, A., Claessen, M., Leitão, S., & Webster, J. (2015). Beyond narrative: Is there an implicit structure to the way in which adults organise their discourse? *Clinical Linguistics and Phonetics, 29*(6), 455–481.

Worrall, L., Sherratt, S., Rogers, P., Howe, T., Hersh, D., & Ferguson, A. (2011). What people with aphasia want: Their goals according to the ICF. *Aphasiology, 25*(3), 309–322.

Wright, H. H., & Capilouto, G. J. (2012). Considering a multi-level approach to understanding maintenance of global coherence in adults with aphasia. *Aphasiology, 26*(5), 656–672.

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*(6–8), 802–813.