AJSLP

Research Article

Macrostructural Analyses of Cinderella Narratives in a Large Nonclinical Sample

Kathryn J. Greenslade,[a] Jade E. B. Stuart,[a] Jessica D. Richardson,[b] Sarah Grace Dalton,[c] and Amy E. Ramage[a,d]

Purpose: Macrostructural narrative analyses are important clinical measures, revealing age-related declines and disorder-related impairments in the accuracy, completeness, logical sequencing, and organization of content. The current study aims to provide preliminary data on typical aging and psychometric evidence supporting multilevel Main Concept, Sequencing, and Story Grammar (MSSG) analyses that capture these aspects of narratives.
Method: Transcripts of *Cinderella* narratives for 92 healthy control participants stratified across four age brackets from the online database AphasiaBank were coded by Richardson and Dalton (2016) for main concept (MC) analysis. In the current study, MSSG analyses were completed for (a) logical sequencing, independently and in combination with MC accuracy and completeness (MC + sequencing), and (b) story grammar organization (i.e., inclusion of episodic components and complexity of episodes). Interrater agreement (99%–100%) revealed highly reliable scoring.
Results: Descriptive statistics for the typically aging sample are presented for sequencing, MC + sequencing, total episodic components, and episodic complexity. Scores for participants over 60 years of age were lower (poorer) than scores for those 20–59 years of age, supporting the construct validity of score use for identifying age-related declines in performance.
Conclusions: This study's novel MSSG analyses of narrative production efficiently assess the logical sequencing and story grammar organization of content in healthy controls. Preliminary reliability and validity evidence support the use of all scores to measure age-related changes in narrative macrostructure. Data from this typically aging sample provide a foundation for future research and clinical assessment aimed at quantifying narrative deficits in adults with communication disorders.
Supplemental Material: https://doi.org/10.23641/asha.12683495

Narrative discourse, or storytelling, is critical to everyday communication, allowing us to entertain, impart important life lessons, and revisit events with others. Successful narration recounts a chain of causally linked events, either real or imagined, in a logically sequenced and organized manner (Coelho, 2002; Mozeiko et al., 2011; Paul, 1995). Because narration is inherently complex, breakdowns in both microstructure (e.g., lexical selection, grammatical structure) and macrostructure (e.g., overarching "gist" and organization) can occur. Thus, narratives can reveal age-related declines in lexical selection, grammatical complexity, efficiency of language production, and inclusion and organization of essential story events (Cannizzaro & Coelho, 2013; Marini et al., 2005; Wright et al., 2005). Clinically, narrative analyses can reveal inaccurate, incomplete, or insufficient content and poor organization in individuals with neurological insult or disease impacting language and/or cognition, including aphasia (e.g., Dalton & Richardson, 2019; Fromm et al., 2017; Richardson & Hudspeth, 2014) and cognitive-communication disorders (CCDs; Elbourn et al., 2019; Galettoet al., 2013; Lê, Coelho, Mozeiko, & Grafman, 2011; Marini et al., 2017; Stout et al., 2000), such as those resulting from right hemisphere damage, traumatic brain injury (TBI), or dementia (Christman Buckingham & Sneed, 2017). Two common approaches to narrative analysis examine the accuracy and completeness of content and/or adherence to story grammar.

[a]Department of Communication Sciences & Disorders, University of New Hampshire, Durham
[b]Department of Speech & Hearing Sciences, The University of New Mexico, Albuquerque
[c]Department of Speech Pathology and Audiology, Marquette University, Milwaukee, WI
[d]Interdisciplinary Program in Neuroscience and Behavior, University of New Hampshire, Durham
Correspondence to: Kathryn J. Greenslade: kathryn.greenslade@unh.edu

Editor-in-Chief: Julie M. Barkmeier-Kraemer
Editor: Kristie A. Spencer
Received October 16, 2019
Revision received March 15, 2020
Accepted May 20, 2020
https://doi.org/10.1044/2020_AJSLP-19-00151

Disclosure: *Kathryn J. Greenslade, Jessica D. Richardson, and Amy E. Ramage are contributors to and users of the password-protected online database AphasiaBank, which is supported by the National Institutes of Health. Control data from AphasiaBank were utilized for the current research. All other authors have declared that no competing interests existed at the time of publication.*

Yet, in adults, these types of analyses often fail to measure logical sequencing and organizational complexity, areas of challenge in aging, aphasia, and CCD populations. Furthermore, multilevel analyses combining all these measures have not been applied to a single task. To effectively capture deficits, new analyses applied to a complex storytelling task are needed. Such measures would permit clinicians to match analyses to perceived deficits, compare productions of neurologically healthy individuals and those with aphasia and/or CCDs, and identify profiles of strengths and weaknesses. Such information could optimize planning and delivery of treatments and improve diagnostic and therapeutic outcomes. As a critical first step, the current study introduces multilevel analyses: Main Concept, Sequencing, and Story Grammar (MSSG). MSSG is designed to efficiently capture an adult's ability to accurately and completely convey important content in a logically sequenced and organizationally complex manner when telling a story, such as *Cinderella*. In other words, it assesses their ability to effectively convey the story's "message." Data from a large neurologically healthy sample, along with preliminary reliability and construct validity evidence, are presented.

Narrative Analyses Assessing Accuracy, Completeness, and Story Grammar Organization

Two primary approaches to analyzing narrative macrostructure include (a) assessing the accuracy and/or completeness of content (Lê, Coelho, Mozeiko, Krueger, & Grafman, 2011; Nicholas & Brookshire, 1995; Richardson & Dalton, 2016) and (b) examining how participant content aligns with story grammar organization (Lê, Coelho, Mozeiko, Krueger, & Grafman, 2011; Lindsey et al., 2019).

Accuracy and Completeness of Content: Main Concept Analysis

Main concept (MC) analysis quantifies the presence, accuracy, and completeness of key story content by comparing a participant's utterances with a closed list of MCs compiled from narratives of healthy controls (Nicholas & Brookshire, 1993, 1995; Stark, 2010). MCs are defined as utterances that contain a subject, one main verb, and, if appropriate, an object (e.g., "Cinderella's father remarried a lady"); subordinate clauses may be included if only one main verb is present (Nicholas & Brookshire, 1995). Essential elements of each MC are identified. Corresponding participant utterances are evaluated to determine whether any essential element is conveyed inaccurately (e.g., "mother" or "woman" instead of "stepmother") or incompletely (e.g., "Cinderella's father remarried"—omits "a lady" or "a woman"). Each MC receives a code for the corresponding participant utterance: (a) accurate and complete, (b) accurate but incomplete, (c) inaccurate but complete, (d) inaccurate and incomplete, or (e) absent (Nicholas & Brookshire, 1995; Richardson & Dalton, 2016). MC analysis identifies discourse-level deficits in microstructure (e.g., accuracy/precision of lexical choices) and macrostructure (e.g., inclusion/completeness of key concepts; Nicholas & Brookshire, 1993).

Richardson and Dalton (2016) developed an MC checklist for the *Cinderella* narrative task from the standardized AphasiaBank and TBIBank discourse protocols (www.Talkbank.org). *Cinderella* was selected as it is a well-known fairytale that elicits lengthy, complex linguistic output. Based on 92 control transcripts from AphasiaBank, 34 MCs, each with two to four essential elements, were identified. A checklist of these MCs included a generalized production (e.g., "[1]Cinderella [2]was a [3]servant," where numbering denotes essential elements) and a nonexhaustive list of alternative productions (e.g., "[1]She [2]had to be [3]a maid"). Consistent with prior research showing age-related declines in narrative production (Capilouto et al., 2005; Marini et al., 2005; Wright et al., 2005), MC scores revealed better performance in younger (< 60 years) than older (≥ 60 years) age groups, providing construct validity evidence (expected age differences) and highlighting the need for age-stratified control data to evaluate client performance (Richardson & Dalton, 2016). Construct validity evidence also supported the use of MC scores to identify reduced accuracy and completeness in persons with aphasia and those with primary progressive aphasia (Dalton et al., 2020; Dalton & Richardson, 2019; Fromm et al., 2017; Kong, 2009; Nicholas & Brookshire, 1993, 1995), as well as in individuals with TBI at 3, 6, 9, and 12 months postinjury (Elbourn et al., 2019). Finally, based on *Cinderella* narratives collected at 3, 6, 9, and 12 months post-TBI, MC score trajectories improved over the first year post-TBI, consistent with spontaneous recovery in narrative production (Elbourn et al., 2019).

Since MC analysis measures the accuracy and completeness of content, productions that include content not specified in the MC list or that occur out of sequence are not penalized. Thus, on its own, MC analysis does not capture deficits in the organization or logical sequencing of narrative content, challenges that are observed in adults with aphasia and/or CCDs (Coelho, 2002; Hameister & Nickels, 2018; Jorgensen & Togher, 2009; Mozeiko et al., 2011). To examine these challenges, additional analyses are required (Hameister & Nickels, 2018).

Story Grammar Analyses

Story grammar refers to an organizational framework used in Western narratives that facilitates comprehension by conveying logical relationships between characters/agents and events (Cannizzaro & Coelho, 2002; Coelho, 2002; Mozeiko et al., 2011; Richardson & Hudspeth, 2014; Stein & Glenn, 1979). Story grammar categories include a setting introducing characters, location, and timing; one or more episodes conveying the story's middle scenes; and a conclusion/coda drawing story events to a close (Roth & Spekman, 1986; Stein & Glenn, 1979). Arguably, "episodes" are the most critical category of the story grammar framework (Mozeiko et al., 2011; Stein & Glenn, 1979). A complete episode minimally includes three components: (a) an "initiating event" that prompts the character to engage in a goal-directed behavior, (b) an "attempt" or action intended to attain the goal, and (c) a "direct consequence" or outcome of the attempt relative to the goal (Coelho, 2002; Lê, Coelho, Mozeiko, Krueger, & Grafman, 2011; Stein & Glenn, 1979). Although not required, episodes can also include plans, or intended character actions; internal responses, or character feelings, thoughts, or goals; and reactions, or character's feelings about successful or failed goal outcomes (Roth & Spekman, 1986). Initiating events, attempts, and direct consequences may extend over more than one utterance, and a single utterance may include more than one component. Following Lê, Coelho, Mozeiko, and Grafman (2011), partial episodes contain two of the three required episodic components. In this study, the term "complex episodes" was used to refer to episodes having two or all three of these required components, encapsulating partial and complete episodes and clearly differentiating between complete content (MCs) and complete episodes (story grammar). As partial or complete episodes, complex episodes typically convey sufficient information for the listener to follow the logical sequence of events and/or infer missing information, whereas including only one episodic component can lead to listener confusion. Use of story grammar components to organize narrative productions is a known challenge for people with aphasia (e.g., Richardson & Hudspeth, 2014) and TBI (e.g., Lê, Coelho, Mozeiko, & Grafman, 2011).

Stark (2010) completed story grammar analyses based on the Cinderella narratives of 10 healthy controls. Forty-one propositions or units akin to MCs were identified and chunked into a story grammar framework, consisting of a setting (four propositions), five episodes (four to eight propositions each), and a conclusion/coda (three propositions). The number of propositions within the setting, each episode, and conclusion were coded as being omitted, produced explicitly, or produced implicitly. Propositions within episodes were not divided further into initiating events, attempts, direct consequences, or other episodic components. Thus, episodic complexity was not addressed because coding did not identify whether episodes included at least two of three required episodic components (Stark, 2010).

Similar to Stark (2010), most story grammar research in adults does not focus on episodic complexity, despite analyses for children often addressing narrative robustness in this way (e.g., Gillam et al., 2016; Heilmann et al., 2010). Importantly, omitting more than two episodic components may lead to listener confusion and signal macrolinguistic processing deficits (Lê, Coelho, Mozeiko, & Grafman, 2011). Thus, measuring episodic complexity in adults, as we do in children, could reveal the reduction in complex organization often observed in narratives of individuals with aphasia and/or CCDs.

Story Goodness Index: An Integration of Content Analysis and Story Grammar

The Story Goodness Index, based on participants' retell of a wordless picture book, examines both story grammar and story completeness (Lê, Coelho, Mozeiko, Krueger, & Grafman, 2011). The Story Goodness Index's story grammar scores reflect the total number of T-units within complex episodes divided by the total number of T-units produced (note that a T-unit refers to an independent clause with all modifiers and attached clauses, which excludes incomplete sentences). Story completeness scores reflect the number of included critical content units divided by the total number of critical content units that were produced by at least 80% of a control sample (n = 46 for Lê, Coelho, Mozeiko, Krueger, & Grafman, 2011; n = 47 for Lindsey et al., 2019). After scoring story grammar and story completeness independently, scores are interpreted together on a two-coordinate grid. Quadrants of the grid are created by horizontal lines (story grammar) and vertical lines (story completeness) at 1 and 2 *SD*s below the mean for each measure based on the control samples. Clockwise from the top left, quadrants represent (a) good organization and reduced content, (b) good organization and good content, (c) poor organization and good content, and (d) poor organization and reduced content (Lê, Coelho, Mozeiko, Krueger, & Grafman, 2011).

The Story Goodness Index's story grammar score has shown macrostructural deficits in individuals with TBI compared with neurologically healthy adults (Lê, Coelho, Mozeiko, Krueger, & Grafman, 2011; Lindsey et al., 2019). However, this discrepancy in scores was not observed between adolescents with typical development and those with autism spectrum disorder (ASD) and typical intellectual functioning (IQ > 80; Canfield et al., 2016). Of note, this story grammar score does not capture a story's robustness nor its adherence to a story grammar framework. For example, a story composed of two T-units within one complex episode (e.g., 2 T-units in a complex episode / 2 T-units total) would receive the same story grammar score as a retell consisting of four T-units within two complex episodes (e.g., 4 T-units in complex episodes / 4 T-units total), despite the latter being a more robust narrative. Rather than measuring how sparse or impoverished narratives are, the Story Goodness Index's story grammar score unequally measures the impact of setting and concluding statements that occur outside episodes as well as vague, irrelevant, and/or off-topic utterances. For example, the story grammar score for "The Bear and the Fly" would be lowered by setting statements, such as "It's a family of bears," or vague/hard to interpret utterances, such as "He was gonna through ⟨any⟩ anything to get rid of the fly" (examples quoted directly from Lindsey et al., 2019, p. 340; for complete scoring examples, please see the appendixes of Lindsey et al., 2019; Lê, Coelho, Mozeiko, Krueger, & Grafman, 2011). Lê, Coelho, Mozeiko, Krueger, and Grafman (2011) reportedly used this story grammar score because it was more sensitive than measures of narrative robustness, such as the number of complex episodes.

Results regarding the story completeness score's sensitivity have been mixed, with lower scores in individuals with severe penetrating head injuries (Lê, Coelho, Mozeiko, Krueger, & Grafman, 2011; Lê et al., 2012), but not in adults with closed head injuries (Lindsey et al., 2019). Furthermore, although adolescents with ASD and typical intellectual functioning showed poorer story composites overall than those with typical development, no statistically significant reductions were noted in story completeness (Canfield et al., 2016). Lindsey et al. (2019) and Canfield et al. (2016) attributed the lack of group differences to the use of short, simplistic stories compared to the original study's book. Yet, all three stories used with the Story Goodness Index have a relatively small number of critical content units (i.e., five to seven in total), potentially limiting the narratives' complexity. Furthermore, the Story Goodness Index's story completeness measure codes critical content as present or absent rather than assessing their completeness and accuracy, potentially failing to capture the quality of speaker contributions. Also, consistent with the lack of significant group differences on story completeness, Lindsey et al. (2019) showed no significant associations between group and discourse performance at 1 *SD* below the mean when comparing Story Goodness Index quadrants. Thus, evidence supporting the utility of the Story Goodness Index's story completeness and overarching measure is mixed.

Beyond this, researchers and clinicians interested in using this tool face several obstacles. First, the control data are limited, with small samples, restricted age ranges, and few published distribution properties. Second, the Story Goodness Index requires coding each T-unit separately for the two measures, which are then plotted against one another; therefore, its clinical efficiency is suboptimal for capturing the multitude of deficits seen in adults with aphasia and/or CCDs. Furthermore, the Story Goodness Index does not account for (a) the accuracy and completeness of content, (b) the logical sequencing of content, and (c) the robustness of complex episodes. Most importantly, the Story Goodness Index's stories are relatively simple with a limited number of episodes, so they sometimes fail to capture group differences.

Thus, sensitive and efficient measures based on a complex story would be beneficial for clinical use to match perceived areas of deficit and compare productions of individuals with aphasia and/or CCDs with neurologically healthy individuals in order to optimize planning and delivery of treatment. These measures must also identify and differentiate strengths and weaknesses, which may vary across disorders and individuals, as well as document changes over time. As a first step in developing such measures, this study sought to develop clinically efficient narrative analyses measuring logical sequencing and production of complex episodes, which are coded secondary to MC analysis for *Cinderella* narratives. For this combined set of MSSG analyses, our aim was to report data from adults without impaired language or cognition, with stratification across four age brackets (20–39, 40–59, 60–79, and 80+ years), using the same sample as that of Richardson and Dalton (2016). Two secondary aims were to (a) provide construct validity evidence for using MSSG analyses to identify age-related declines in performance and (b) determine whether individual participants demonstrated consistent or discrepant performance in their use of story grammar components versus their accuracy, completeness, and logical sequencing of content. Consistent with the study of Richardson and Dalton, we hypothesized that younger participants (< 60 years) would earn higher scores than older participants (60+ years), revealing expected age-related declines in narrative production. Following patterns of control data from Lê, Coelho, Mozeiko, and Grafman (2011) and Lindsey et al. (2019), we predicted that, while most neurologically healthy adults would demonstrate performance consistency across MSSG's measures, some would perform better on story grammar use than on the accuracy, completeness, and logical sequencing of content, and others would demonstrate the opposite profile.

Method

Participants

The current study used the same sample as that of Richardson and Dalton (2016), who reported data for MC analysis based on 92 transcripts of typically aging adults from AphasiaBank. AphasiaBank is an National Institutes of Health–funded online database that permits researchers to access transcripts and media files from healthy controls and participants with various communication disorders (Forbes et al., 2012; MacWhinney et al., 2011). Transcripts included in current analyses were contributed by the Capilouto, Richardson, and Wright laboratories, per each participant's consent (https://aphasia.talkbank.org/). Participants were typically aging, had no history of neurological conditions, and reported English as their primary language. All participants aged 40 years and older were required to pass the Mini-Mental State Examination (MMSE).

Table 1 summarizes the sample's demographic characteristics. There were slightly more women (n = 55) than men (n = 37), with a mean education of 15.6 years (± 2.5) and a mean MMSE score of 29.207 (± 1.097). The sample was primarily Caucasian (88 Caucasian, two African American, two Hispanic/Latino) and ranged in age from 21 to 89.5 years. The sample was stratified by age, with 22–24 participants in each of four brackets (20–39, 40–59, 60–79, and 80+ years), adjusting for an assignment error in the original stratification, in which each bracket had 23 participants.

Procedure

Transcripts from *Cinderella* narratives were elicited according to the AphasiaBank protocol (MacWhinney et al., 2011). First, a text-covered picture book depicting Disney's version of *Cinderella* (Grimes, 2005) was presented. The examiner stated, "I'm going to ask you to tell a story. Have you ever heard the story of *Cinderella*?" All participants reported knowing this story. Next, the examiner provided instructions: "Do you remember much about it? These pictures might remind you of how it goes. Take a look at the pictures and then I'll put the book away, and ask you to tell me the story in your own words." After the participant reviewed the book, it was removed, and they recounted the story. If necessary, the examiner stated: "Now tell me as much of the story of *Cinderella* as you can. You can use any details you know about the story, as well as the pictures you just looked at." Although the prompts "What

Table 1. Demographic information for the 92 transcripts selected by Richardson and Dalton (2016) as the typically aging sample from the AphasiaBank database, with the error adjustment made.

Group	*n*	Age in years	Sex (F:M)	Education in years	Race/ethnicity	MMSE
All	92	58.3 (± 21.6)	55:37	15.6 (± 2.5)[a]	88 Caucasian 2 African American 2 Hispanic/Latino	29.207 (± .776) for *n* = 82 (10 missing)
20–39 years	22	29.9 (± 5.96)	13:9	15.6 (± 1.9) for *n* = 21 (1 not reported)	19 Caucasian 2 African American 1 Hispanic/Latino	29.462 (± .776) for *n* = 13 (9 missing)
40–59 years	23	48.4 (± 6.3)	15:8	15.7 (± 2.5)[a]	22 Caucasian 1 Hispanic/Latino	29.773 (± .429) for *n* = 22 (1 missing[b])
60–79 years	24	71.29 (± 4.7)	14:10	15.67 (± 2.3)	24 Caucasian	29.375 (± 1.013)
80+ years	23	83.9 (± 2.9)	13:10	15.3 (± 2.8)[a]	23 Caucasian	28.348 (± 1.301)

Note. F = female; M = male; MMSE = Mini-Mental State Examination.
[a]As reported by Richardson and Dalton (2016). [b]MMSE adjusted score of 60, indicating a passing score.

happened next?" and "Go on" were allowed, no participants required prompting once they began telling the story.

MSSG Measures

MSSG analyses were conducted using a single coding sheet for each participant, which consolidated all measures in one place. The coding sheet along with general instructions and materials for conducting reliability analyses are available in Supplemental Material S1.

MC Analysis

MC codes for all transcripts were obtained from the study by Richardson and Dalton (2016). Transcripts were checked to ensure that all MCs were identified and coding rules were consistently applied. A limited number of codes were changed in this checking process. Reliability was conducted for 20% of transcripts, yielding a point-to-point agreement of 96.540% and a Cohen's kappa of .965. As previously described, each of *Cinderella*'s 34 MCs was assigned a code, with associated scores as follows: 3 points for accurate/complete, 2 points for accurate/incomplete or inaccurate/complete, 1 point for inaccurate/incomplete, and 0 points for absent. See Richardson and Dalton's (2016) Appendix 4 for examples of utterances matching each score. A total MC score (maximum = 102) was calculated by adding the points assigned to each MC.

Sequencing, "MC + Sequencing," Total Episodic Components, and Episodic Complexity

Each MC was assigned a sequencing score based on whether it was presented in a logical sequence. Utterances not matched to an MC were not evaluated for logical sequencing. To be logically sequenced, each MC had to come after the MC(s) before it (e.g., MC 2 must follow MC 1). An exception to this rule was when the sequence of MCs was judged to be interchangeable. For example, MCs 3–5 (Cinderella's stepfamily was mean to her, she was their servant, she was forced to do the chores) could be presented in any sequence among themselves. Other exceptions were made based on the content of MCs. For instance, MC 13 ("Cinderella was upset") could follow MC 10, 11, or 12 (she can't go to the ball, her dress was torn by her stepsisters, or they went to the ball without her), given that this reaction could logically follow any of those events. Similarly, depending on how it was presented, MC 18 ("She had to be home by midnight") could come before or after MC 17 or MC 21, functioning either as a direct consequence (e.g., a consequence of the fairy godmother's spell was that she had to be home by midnight) or as an initiating event (e.g., she was reminded she had to go home as the clock struck midnight). Appendix A summarizes scoring rules for sequencing and provides examples of how to apply these rules.

Any MC stated in a logical sequence was assigned 3 points. An MC stated out of sequence but signaled as being in the wrong sequence by the speaker (e.g., "I forgot to say") was assigned 2 points. An MC stated out of sequence but not signaled as being in the wrong sequence was assigned 1 point. An absent MC was assigned a score of 0. Of note, except for scores of 0 indicating an absent MC, sequencing and MC scores were independent of each other. For example, an MC could be accurate and complete (score = 3) but out of sequence without being signaled (score = 1). If all 34 MCs were logically sequenced, the maximum sequencing score was 102 (34 MCs multiplied by 3). A combined MC + sequencing score was calculated by adding the MC score and sequencing score for each MC and totaling these values across MCs (maximum total = 204; "34 × 3" for MC scores plus "34 × 3" for sequencing scores).

Building on Stark (2010), while further specifying episodic components, a story grammar code was assigned to each MC, based on the category/component that MC was most likely to serve. As with sequencing, utterances that did not match an MC were not evaluated for story grammar. Story grammar codes included setting, initiating event, attempt, direct consequence, mental state (either an internal response or reaction), and conclusion (Roth & Spekman, 1986; Stein & Glenn, 1979). Four MCs were determined to potentially correspond with two story grammar components, depending on how the participant constructed the MC (e.g., "Cinderella had to be home by midnight" could be a "direct consequence" or an "initiating event"). For these MCs, the coder assigned the most appropriate code. Furthermore, a fifth MC (17, "Cinderella went to the ball in the coach") could be coded as a direct consequence, an initiating event, or both, if the speaker included Cinderella's departure from her home and her arrival at the ball (e.g., "So off to the ball she goes in her carriage. Cinderella gets to the ball"). Appendix B summarizes rules for story grammar coding with examples of how to assign codes for these five MCs.

Five episodes, along with setting and concluding statements, were established based on the story grammar codes. Two scores examined how complex participants' episodes were. First, for each episode, coders determined whether a speaker included at least one instance of the three required episodic components: initiating event, attempt, or direct consequence. This yielded a total episodic components score with a maximum score of 15 (i.e., 1 point per episodic component, or 3 possible points per episode × 5 episodes). This score is similar to ratings of initiating events and direct consequences in the Monitoring Indicators of Scholarly Language (Gillam et al., 2016), while allowing higher scores based on the number of episodes in which each component was included. Second, each episode was assigned a complexity score of 1 if it included at least two of the three required episodic components, and a score of 0 if it had one or zero components. The maximum episodic complexity score was 5 (1 point per episode).

Reliability

To ensure coding reliability, two graduate students completed a 2-week training for assigning sequencing, total episodic components, and episodic complexity scores. Training was completed when the pair independently scored and reached at least 80% reliability for point-to-point agreement and .7 for Cohen's kappa for four consecutive practice samples across all scores. Training transcripts were blinded and recoded after coding was completed for the remainder of the sample. Point-to-point agreement and Cohen's kappa were calculated for 20% of the study sample to ensure continued fidelity. Cohen's kappa was 1.00 for total episodic components and episodic complexity and .99 for sequencing, where coefficients of .80 or higher are regarded as "very good" (McHugh, 2012). Point-to-point agreement was 100% for total episodic components and episodic complexity and 99% for sequencing, where values above 80% are considered "very good" (Kazdin, 1982). To resolve discrepant coding, consensus was conducted. Also, one independently assigned sequencing score was changed following clarification of scoring manual rules. Of note, these high levels of reliability resulted from coding based on agreed upon matches of participant utterances to MCs. Reliability might be lower if coders had to assign scores for sequencing, total episodic components, and episodic complexity without first agreeing upon utterance matching.

Analysis

Statistical Package for the Social Sciences (SPSS v25) was used to generate descriptive statistics, skew and kurtosis, and Shapiro–Wilk's normality tests for sequencing, MC + sequencing, total episodic components, and episodic complexity scores for the entire sample and for the four age brackets (20–39, 40–59, 60–79, and 80+ years). Variables were considered to be normally distributed if skew and kurtosis z scores were less than a conservative value of $\pm$ 2.58 and if Shapiro–Wilk's test p values were greater than .05 (Laerd Statistics, n.d.).

To compare sequencing and story grammar performance of older versus younger participants, Welch's t tests were conducted for normally distributed variables, and Mann–Whitney U tests were conducted for nonnormally distributed variables. Scores were compared for the two younger (20–59 years) and two older (60–80+ years) age brackets, following Richardson and Dalton (2016). Alpha was set at .05 for all significance testing. Two-way analyses of variance to assess the main effect of sex and interaction of sex and age group were not conducted due to heterogeneity of variance and (in two cases) nonnormal distributions.

To determine whether neurologically healthy controls demonstrated consistent versus discrepant performance across the macrostructure measures, z-score cutoffs were generated at 1 and 2 *SD*s below the mean for MC + sequencing and total episodic components. Comparisons were made across z scores for individual participants to identify those with consistently good story grammar component use and accuracy, completeness, and logical sequencing of content; good story grammar component use in the presence of poor content; good content in the presence of poor story grammar component use; and poor story grammar component use and content. Results are descriptively summarized below.

Results

Descriptive Statistics

Due to the unequal sex ratio across age brackets, descriptive statistics assessed potential sex-based score differences. For 20–39, 40–59, and 80+ years age brackets, performance across variables did not differ statistically for men versus women ($p > .140$ for all measures). For 60- to 79-year-olds, women performed better than men, as assessed by t tests for sequencing, $t(22) = 3.083$, $p = .005$, and MC + sequencing, $t(22) = 3.224$, $p = .042$, and as assessed by Mann–Whitney U tests for total episodic components, $U = 30.500$, $z = -2.360$, $p = .019$, and episodic complexity, $U = 29.000$, $z = -2.544$, $p = .016$.

To determine the potential impact of mental status on narrative performance, Spearman rank correlations were run between MMSE unadjusted raw scores and each macrostructure variable, across the full sample and in those participants over 60 years of age. Nonparametric correlations were run due to the limited range of MMSE scores (26–30). Across the full sample, correlations were significant for all variables (ps < .015), except total episodic components (p = .112). However, scatter plots with fit lines for older and younger subgroups revealed no linear relationships. Following this finding, no statistically significant correlations were found between MMSE scores and any of the macrostructure variables in the older age group (ps > .180).

Sequencing, MC + Sequencing, Total Episodic Components, and Episodic Complexity

Table 2 summarizes descriptive statistics for sequencing, MC + sequencing, total episodic components, and episodic complexity scores, including means, standard deviations, medians, ranges, skew, and kurtosis values for the full sample and each age bracket. Overall sequencing scores (max = 102) ranged from 9 to 92, with a mean of 61.098 (*SD* = 18.880), while MC + sequencing scores (max = 204) ranged from 17 to 183, with a mean of 121.109 (*SD* = 37.509). For each age bracket, z scores for skewness and kurtosis fell within acceptable ranges (≤ ± 2.58) for both measures; however, despite being mesokurtic, negative skew was detected for both measures in the full sample. Shapiro–Wilk's test indicated nonnormal distributions for the full sample, in the presence of normally distributed scores for all age brackets for both measures.

Total episodic components score (max = 15) for the full sample had a mean of 11.457 (*SD* = 2.771) and a median of 12. In each age bracket, means and medians were proximal in value. For the full sample, skew and kurtosis z scores for total episodic components both indicated a nonnormal distribution, which was confirmed by Shapiro–Wilk's test ($p < .001$). There was negative skew ($z = -3.139$) in the 60- to 79-year-old age bracket, while Shapiro–Wilk's test indicated nonnormal distributions in the age brackets of 20–39 ($p = .019$), 40–59 ($p = .008$), and 60–79 ($p = .002$) years. Based on both z scores and Shapiro–Wilk's tests, the age bracket of 80+ years demonstrated normally distributed total episodic components scores.

For the episodic complexity score (max = 5), the full sample median and mean were both 4.00, with an *SD* of 1.167. Means and medians for each age bracket were proximal. Skew and kurtosis z scores for episodic complexity indicated nonnormal distributions for the full sample, which was confirmed by Shapiro–Wilk's test ($p < .001$). In the 40- to 59-year-old age bracket, negative skew was detected based on the z score (−2.676), while Shapiro–Wilk's test indicated nonnormal distributions in the age brackets of 20–39, 40–59, and 60–79 years (ps < .001). Thus, only the eldest age bracket (80+ years) demonstrated normally distributed episodic complexity scores based on z scores and Shapiro–Wilk's tests.

Table 2. Descriptive statistics for sequencing, "main concept + sequencing," total episodic components, and episodic complexity scores for the full sample and each age stratification.

MSSG measure	All (*N* = 92)	20–39 years (*n* = 22)	40–59 years (*n* = 23)	60–79 years (*n* = 24)	80+ years (*n* = 23)
Sequencing (max 102)					
M	61.098	68.591	70.826	58.625	46.783
SD	± 18.880	± 11.438	± 13.753	± 18.273	± 20.865
Mdn	64.000	69.500	73.000	60.000	45.000
Range	9–92	42–92	34–89	16–91	9–85
Skew (*SE*)	−0.833 (0.251)[a]	−0.405 (0.491)	−1.043 (0.481)	−0.773 (0.472)	−0.061 (0.481)
Kurtosis (*SE*)	0.436 (0.498)	0.540 (0.953)	1.214 (0.935)	0.676 (0.918)	−0.132 (0.935)
Main concept + sequencing (max 204)					
M	121.109	136.136	141.043	115.583	92.565
SD	± 37.509	± 22.592	± 27.466	± 35.944	± 41.240
Mdn	128.000	138.000	145.000	118.500	88.000
Range	17–183	82–183	67–177	32–179	17–169
Skew (*SE*)	−0.837 (0.251)[a]	0.422 (0.491)	−1.098 (0.481)	−0.754 (0.472)	−0.098 (0.481)
Kurtosis (*SE*)	0.454 (0.498)	0.827 (0.953)	1.329 (0.935)	0.716 (0.918)	−0.133 (0.935)
Total episodic components score (max 15)					
M	11.457	12.727	12.783	11.042	9.348
SD	± 2.771	± 1.518	± 1.808	± 2.440	±3.433
Mdn	12.000	13.000	13.000	12.000	10.000
Range	2–15	9–15	9–15	4–14	2–15
Skew (*SE*)	−1.355 (0.251)[a]	−0.929 (0.491)	−0.909 (0.481)	−1.482 (0.472)[a]	−0.598 (0.481)
Kurtosis (*SE*)	1.975 (0.498)[a]	−0.352 (0.953)	0.002 (0.935)	2.080 (0.918)	0.103 (0.935)
Episodic complexity score (max 5)					
M	4.000	4.455	4.565	3.833	3.174
SD	± 1.167	± .671	± .662	± 1.239	± 1.370
Mdn	4.000	5.000	5.000	4.000	3.000
Range	0–5	3–5	2–5	1–5	0–5
Skew (*SE*)	−1.314 (0.251)[a]	−0.860 (0.491)	−1.288 (0.481)[a]	−1.151 (0.472)	−0.574 (0.481)
Kurtosis (*SE*)	1.400 (0.498)[a]	−0.242 (0.953)	0.625 (0.935)	−0.567 (0.918)	−0.082 (0.935)

Note. MSSG = Main Concept, Sequencing, and Story Grammar.
[a]Significant skew or kurtosis.

Age-Related Differences in Macrostructural Narrative Performance Across Variables

Sequencing, MC + sequencing, total episodic components, and episodic complexity were compared for younger (20–59 years) versus older (60+ years) participants. Based on normal distributions and unequal variances across groups, Welch's t tests were conducted for sequencing and MC + sequencing scores. Two outliers were identified for each score. Tests with and without outliers yielded similar results, so reported results include outliers. Between-groups differences are reported as mean ± standard error. Significantly higher sequencing scores were found for younger (M = 69.733, SD = 12.581) as compared to older participants (M = 52.830, SD = 20.272), with a mean difference of 16.904 ± 3.536, $t(77.365) = 4.827$, $p < .001$. The younger group also performed better on MC + sequencing (younger: M = 138.644, SD = 25.039; older: M = 104.319, SD = 39.933), for a statistically significant mean difference of 34.325 ± 6.918, $t(77.817) = 4.962$, $p < .001$.

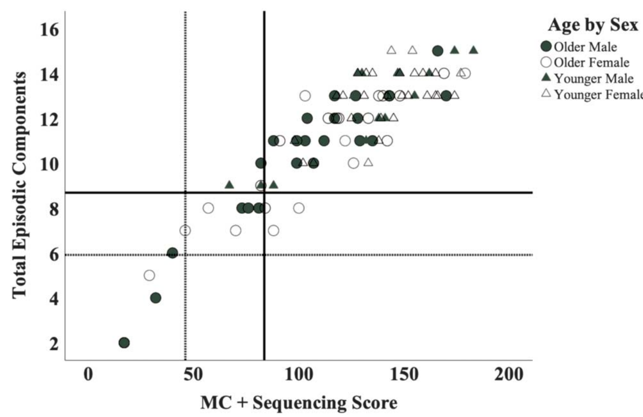
Due to nonnormal distributions, Mann–Whitney U tests were conducted to assess age-related differences in total episodic components and episodic complexity. As assessed by visual inspection, distributions of total episodic components and episodic complexity scores were dissimilar for the younger and older age groups. Scores in the younger group were significantly higher than in the older group for total episodic components (younger mean rank = 59.533, older mean rank = 34.021), $U = 471.000$, $z = -4.635$, $p < .001$, and episodic complexity (younger mean rank = 57.289, older mean rank = 36.170), $U = 572.000$, $z = -4.028$, $p < .001$.

Consistent Versus Discrepant Performance Across Macrostructural Analyses

To determine whether participants who had difficulty telling accurate, complete, and logically sequenced story content also struggled with story grammar organization, participants' MC + sequencing scores (x-axis) were plotted against their total episodic components scores (y-axis; see Figure 1). Similar to the Story Goodness Index, vertical and horizontal lines represented z scores that were 1 SD (solid lines: MC + sequencing = 82, $z = -1.043$; 8 < total episodic components < 9, [$-.886 < z < -1.247$]) and 2 SDs (dotted lines: MC + sequencing = 46, $z = -2.002$; 5 < total episodic components < 6, [$-1.969 < z < -2.330$]) below the mean.

Visual inspection of Figure 1 revealed an expected linear relationship between MC + sequencing and total episodic components, indicating that speakers who produced more accurate, complete, and logically sequenced story content typically generated more episodic components. The bottom left quadrant of the plot identified 11 narratives produced by speakers in the older group, representing narratives with poor content (≥ 1 SD below on MC + sequencing) and few episodic components (≥ 1SD below on total episodic components). Of the 11, eight were men, and 10 were over 70 years of age. Thus, this quadrant illustrates

Figure 1. Scatter plot comparing total episodic components to MC + sequencing for younger and older participants by sex. Solid lines indicate 1 *SD* below the mean; dotted lines indicate 2 *SD*s below the mean. Lines divide the graph into four quadrants, with the poorest performance in the bottom left quadrant. MC = main concept.



an overall decline in macrostructural narrative quality in some older participants, particularly men. The lower right quadrant identified three older (> 70 years) women who used fewer episodic components than expected, despite average accuracy, completeness, and logical sequencing. Finally, in the upper left quadrant, two older and two younger participants were identified who had below average accuracy, completeness, and logical sequencing, in the presence of average episodic component use. The older participants in this quadrant included one man and one woman, both over 85 years of age; the younger were 19- and 41-year-old men.

Discussion

This study furthered our understanding of how neurologically healthy controls structure their ideas when telling a complex narrative and extended Richardson and Dalton's (2016) work to develop clinically efficient and effective macrostructural narrative analyses: MSSG. Based on a large sample of AphasiaBank control transcripts, this study generated age-stratified data for four analyses related to logical sequencing and story grammar organization for *Cinderella* narratives: sequencing, MC + sequencing, total episodic components, and episodic complexity. The entire sample and each age bracket were described demographically, allowing clinicians to determine whether MSSG measures would be applicable for their use. Furthermore, score differences between older and younger participants provided construct validity evidence supporting the use of scores to detect age-related declines in narrative production. Finally, as expected, plotting episodic component use against the accuracy, completeness, and logical sequencing of content revealed consistently good or poor performance for most participants (n = 85). Yet, even among neurologically healthy controls, the combined analyses revealed discrepant performance in some, suggesting strong potential to pinpoint subtle communication strengths and weaknesses in individuals with aphasia and/or CCDs. These findings are an important step toward streamlining clinically useful, multilevel analyses of complex narrative production through MSSG.

MSSG Data From a Typically Aging Sample

This study's typically aging data for four MSSG analyses were consistent with expectations. Most speakers produced logically sequenced narratives, consisting of multiple episodes that included at least two of three required episodic components. In the full sample, negative skew led to nonnormal distributions across all variables. Negative skew was especially pronounced for total episodic components and episodic complexity, with skew detected within age brackets and in the full sample. Thus, as expected, most neurologically healthy controls produced complex episodes throughout their *Cinderella* narratives, with particularly high performance noted in younger age brackets. Within age brackets, sequencing and MC + sequencing scores demonstrated normal distributions. Thus, generating typically aging data that clustered participants closely in age was critical for providing appropriate comparisons. Clinicians using these analyses are advised to use age-stratified data to compare client scores to expected performance and bear in mind the limitations of current data described below (e.g., relatively small samples in each age bracket).

Age-Related Differences in MSSG Scores

As hypothesized, scores for each variable were significantly lower (poorer) for older as compared to younger participants. The two older age groups (60+ years) produced fewer MCs that were complete, accurate, and logically sequenced (MC + sequencing score), indicating poorer narrative content with age. Older participants also produced fewer episodic components, leading to less complex episodes (i.e., fewer episodes that had at least two of three episodic components). Qualitatively, these MSSG scores translated to less thorough narratives and/or content that was less relevant to the classic fairytale. Within the older and younger age brackets, these declines in narrative production were not related to participants' mental status, as measured by the MMSE. Thus, results were consistent with Richardson and Dalton's (2016) MC analysis, as well as previous research indicative of age-related declines in narrative performance (Capilouto et al., 2005; Marini et al., 2005; Wright et al., 2005). Of note, main effects for sex were also observed in the 60- to 79-year-old age bracket. Due to heterogeneity of variance and (in some cases) nonnormal distributions, interactions of age and sex could not be assessed with two-way analyses of variance; thus, it is possible that such interactions exist.

Expected age-related declines provide construct validity evidence for using MSSG scores to determine whether an individual is producing less accurate, complete, and/or logically sequenced content or less complex story grammar due to aging. MSSG analyses may also prove useful in monitoring these declines. Combining these analyses with assessment of other cognitive domains (e.g., executive functions, memory) could help illuminate the mechanisms underlying age-related declines in narrative productions (Cannizzaro & Coelho, 2013). Finally, evidence of age-related declines highlights the importance of generating and using age-stratified norms to best identify narrative production challenges in adults with aphasia and/or CCDs.

Consistency and Discrepancies Across Macrostructural Analyses

Overall, sequencing, total episodic components, and episodic complexity scores were highly consistent with MC analysis scores. The consistency in scores is logical because current variables were coded only for MCs. Furthermore, neurologically healthy controls would be expected to produce narratives with good content and use of story grammar components. As adults age, we expect to see declines in both content and use of story grammar components—a pattern identified in 11 current participants whose scores were poor across all measures. This subset of participants was predominantly men (eight of 11) and over 70 years of age (10 of 11). These participants' stories tended to be quite short, averaging about 16 utterances, six episodic components, and less than two complex episodes per story. Qualitatively, many of these participants indicated difficulty remembering how story events took place (e.g., "somehow she wound up at the ball," "I don't know how the word got around [about the ball]"). Thus, MSSG analyses detected the poor narrative performance observed qualitatively.

Importantly, even in the present control sample, seven of 92 participants demonstrated discrepant scores. Four demonstrated average use of episodic components in the presence of poorer-than-expected accuracy, completeness, and logical sequencing of content, and three exhibited average content that used fewer episodic components than expected. For example, one woman who was over 80 years of age obtained an MC + sequencing score of 100 out of 204 ($z = -0.563$), indicating accurate, complete, and logically sequenced content. In contrast, she produced only eight of 15 required episodic components (total episodic complexity: $z = -1.247$) and only two episodes with at least two required components ($z = -1.713$), indicating lower-than-expected episodic complexity. Such reductions in episodic complexity could lead to confusion, forcing the listener to make assumptions about missing events that prompted character action, attempts to attain their goal(s), and consequences of those attempts. Further research is needed to determine whether adults with aphasia and/or CCDs will demonstrate similar discrepancies and whether these combined measures will efficiently and effectively capture macrostructural deficits.

Establishment of MSSG Analyses, Limitations, and Future Directions

In designing this study's novel MSSG analyses, we sought to create clinically useful, effective, and efficient measures; thus, recommendations from prior research were considered. First and foremost, Lindsey et al. (2019) postulated that a complex narrative was needed to reveal deficits in populations with CCDs. Thus, *Cinderella* narratives were used, given that multiple, complex episodes are required to tell the story. This high level of complexity was sufficient to detect age-related declines in narrative production; future research is needed to determine if *Cinderella* narratives are sufficiently complex to identify deficits in those with aphasia and/or CCDs.

In addition, Lê, Coelho, Mozeiko, Krueger, and Grafman (2011) argued that the Story Goodness Index's effectiveness in identifying narrative deficits in individuals with TBI could be attributed to the combined power of their story grammar and story completeness measures. Canfield et al. (2016) similarly reported that the combination of quantitative measures revealed deficits in adolescents with ASD that were not found based on story grammar or story completeness alone. Our combined MSSG analyses followed suit and added a measure of logical sequencing. To assess story completeness, MSSG used MC analysis, which has revealed age-related declines (Richardson & Dalton, 2016), differences between healthy controls and adults with communication disorders (Dalton & Richardson, 2019; Elbourn et al., 2019), and improvements due to spontaneous recovery following a TBI (Elbourn et al., 2019). For sequencing and story grammar measures, only utterances that matched an MC were evaluated, promoting clinical efficiency and coding reliability by eliminating the need to code each transcript utterance. This allowed efficient analysis of logical sequencing, a macrostructure analysis that has the potential to reveal narrative disorganization in individuals with aphasia and/or CCDs. Furthermore, only five of Cinderella's 34 MCs required active coding to identify which story grammar component an utterance served. From there, a 1 or 0 was assigned to indicate whether each episode included an initiating event, attempt, and direct consequence, efficiently yielding measures of story grammar complexity (total episodic components and episodic complexity). While neither of these measures identified deficits in producing specific episodic components (e.g., a participant who consistently omits attempts), a microanalysis of the data from which these scores were generated could provide this valuable information for researchers and clinicians. (Note, the coding sheet in Supplemental Material S1 color codes the three required episodic components to facilitate such microanalyses.) Our goal in designing MSSG measures was for experienced clinicians to be able to analyze narratives with minimal to no transcription. To determine if this is the case, future research should examine whether these measures could be reliably scored in real time or based on review of videos without transcription.

Beyond efficiency, MSSG measures were effective in detecting age-related declines in sequencing and story grammar complexity. Future research should explore their clinical utility in identifying challenges of individuals with aphasia and/or CCDs and document additional psychometric evidence, including test–retest reliability and construct validity for measuring changes over time. Researchers should aim to utilize large, diverse, and well-stratified samples, bearing in mind that communication deficits may present differently across distinct etiologies. Furthermore, although our sequencing and story grammar analyses were more efficient than traditional analyses that code each transcript utterance, future research should compare the two methods to determine whether they capture similar challenges or whether omitting utterances that do not match an MC fails to adequately represent a speaker's true narrative abilities. MSSG's story grammar measures could also be compared with a measure of the proportion of utterances within complex episodes to the total number of transcript utterances, following the Story Goodness Index's story grammar measure. Such a comparison would determine if these measures detect similar challenges in individuals with CCDs.

While current results provide strong preliminary evidence supporting MSSG analyses, a few limitations should be acknowledged. Although our typically aging sample was roughly twice that of prior research on the Story Goodness Index (Lê, Coelho, Mozeiko, Krueger, & Grafman, 2011; Lindsey et al., 2019; Mozeiko et al., 2011), the number of participants in each age bracket was relatively small. The sample also lacked diversity in terms of race, ethnicity, and educational attainment. Furthermore, the sample included more women than men, and women received higher scores on MCs in the full sample (Richardson & Dalton, 2016) and on sequencing and story grammar in the 60- to 79-year-old age bracket. Future research should increase the sample's size and diversity and equalize the sex ratio, which could provide normative data that are representative of the broader adult population. Additionally, narrower age stratifications and analyses of sex and education effects would ensure that such normative data adequately reflect the population.

Finally, it should be noted that the *Cinderella* narrative presents its own challenges. Even in neurologically healthy controls, few individuals approached a maximum MC + sequencing score. Since *Cinderella* is a familiar story, speakers might have assumed shared listener knowledge and produced shorter, vaguer narratives as a result. Alternately, speakers may have (a) failed to remember the story adequately, (b) lacked the working memory abilities to recall the story, or (c) told a version of the story that did not align with the MCs. To address these issues, an unfamiliar story, similar to but more complex than those used in the Story Goodness Index, could be used, and speakers could view images associated with the story while telling it. Furthermore, future research should examine how MSSG scores relate to participants' executive functions and memory abilities. For example, because the MMSE has poor sensitivity to subtle declines in cognitive function secondary to aging (cf. Pinto et al., 2019), future research should use a memory screening that is more sensitive to subtle changes (e.g., Montréal Cognitive Assessment) or a more thorough cognitive assessment. Last, it is worth noting that retelling a fictional story or fairy tale may differ from other discourse contexts, such as personal narratives or expository tasks.

Task Relevance

On a final note, we recognize that some may question the relevance of telling a fairytale to functional, everyday communication. Successfully telling the story of *Cinderella* is complex, requiring intact micro- and macrolinguistic production, while also placing demands on executive functioning (e.g., working memory to keep track of previously shared content, inhibition to prevent sharing of irrelevant information). Speakers who fail to include accurate, complete, and/or logically sequenced content in an organized manner when telling *Cinderella* may also have difficulty relaying details in day-to-day interactions. Thus, while omitting that "Cinderella must return home by midnight" or inaccurately stating "her stepmother created a dress for her" may seem trivial, deficits in generating accurate, complete, and well-organized information in real-world contexts could lead to loss of employment, missed medical appointments, or miscommunication with family members. In short, content is important, and the expression of vague, incorrect, or poorly organized information can lead to confusion, frustration, and/or communicative breakdowns and misunderstandings. Using the familiar narrative of *Cinderella* could allow researchers and clinicians to quantify the accuracy, completeness, and/or organization of content in an efficient, standardized way, which might not be possible with personal stories. That said, we acknowledge that performance on a fictional narrative may differ in important ways from performance on personal narratives or expository/procedural discourse. For example, across adulthood, lexical density was the highest for personal narratives, then fictional narratives, and least for procedural discourse; furthermore, older adults demonstrated significantly greater lexical density on personal narratives and procedural discourse than younger adults (Fergadiotis et al., 2011). Alternate analyses, such as high-point analysis for personal narratives, should be explored to examine macrostructure in other discourse forms.

Conclusion

This study introduced novel, multilevel narrative analyses: MSSG. MSSG's four novel analyses were coded secondary to MC analysis for *Cinderella* narratives. The aim was to capture the overarching accuracy, completeness, and logical sequencing of narrative content and story grammar organization, focusing on the complexity of episodes. Based on transcripts from a large sample of neurologically healthy controls, we provided preliminary data for a typically aging sample for MSSG's four new analyses: sequencing, MC + sequencing, total episodic components, and episodic complexity. Together with Richardson and Dalton (2016), distributional characteristics are available for all five MSSG analyses for *Cinderella* narratives. Appendixes A and B provide guidelines for using current analyses, which should be used in conjunction with Richardson and Dalton's Appendixes 1–3 and Nicholas and Brookshire's (1995) Appendix A. Use of age-stratified data is recommended, given age-related score differences.

The current findings provide a strong foundation for efficient and effective measurement of narratives. While further research is needed to confirm the reliability, validity, and clinical utility of sequencing and story grammar measures, MSSG analyses have the potential to enable clinicians to match analyses to perceived areas of deficit, compare narrative productions of neurologically healthy adults and those with aphasia and/or CCDs, identify profiles of strengths and weaknesses, and improve diagnostic and therapeutic outcomes for adults with communication disorders.

Author Contributions

Kathryn Joy Greenslade: Conceptualization (Equal), Data curation (Supporting), Formal analysis (Lead), Investigation (Equal), Method–section head (Equal), Project administration (Supporting), Resources (Equal), Supervision (Lead), Validation (Equal), Visualization (Supporting), Writing–original draft (Equal), Writing–review and editing (Lead). **Jade E. B. Stuart:** Conceptualization (Equal), Data curation (Lead), Investigation (Equal), Methodology (Equal), Project administration (Lead), Writing - Original Draft (Equal), Writing - Review & Editing (Equal). **Jessica D. Richardson:** Conceptualization (Supporting), Investigation (Supporting), Resources (Equal), Writing - Review & Editing (Equal). **Sarah Grace Dalton:** Investigation (Supporting), Resources (Equal), Writing - Review & Editing (Equal). **Amy E. Ramage:** Conceptualization (Supporting), Formal analysis (Supporting), Visualization (Lead), Writing - Review & Editing (Equal).

Acknowledgments

J. D. Richardson's research was supported by an Institutional Development Award from the National Institute of General Medical Sciences (P20GM109089). We would like to thank Michelle LaCroix for her assistance in coding transcripts and her attention to detail and Melanie Pond for her assistance with data entry. We also thank the creators of AphasiaBank (NIDCD R01-DC008524) for developing a repository for researchers to share their data and the Capilouto, Richardson, and Wright laboratories and their participants for sharing the *Cinderella* transcripts used in this research.

References

Canfield, A. R., Eigsti, I.-M., de Marchena, A., & Fein, D. (2016). Story goodness in adolescents with autism spectrum disorder (ASD) and in optimal outcomes from ASD. *Journal of Speech, Language, and Hearing Research, 59*(3), 533–545. https://doi.org/10.1044/2015_JSLHR-L-15-0022

Cannizzaro, M. S., & Coelho, C. A. (2002). Treatment of story grammar following traumatic brain injury: A pilot study. *Brain Injury, 16*(12), 1065–1073. https://doi.org/10.1080/02699050210155230

Cannizzaro, M. S., & Coelho, C. A. (2013). Analysis of narrative discourse structure as an ecologically relevant measure of executive function in adults. *Journal of Psycholinguistic Research, 42*(6), 527–549. https://doi.org/10.1007/s10936-012-9231-5

Capilouto, G., Wright, H. H., & Wagovich, S. A. (2005). CIU and main event analyses of the structured discourse of older and younger adults. *Journal of Communication Disorders, 38*(6), 431–444. https://doi.org/10.1016/j.jcomdis.2005.03.005

Christman Buckingham, S. S., & Sneed, K. E. (2017). Cognitive-communication disorder. In J. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of clinical neuropsychology* (pp. 1–8). Springer. https://doi.org/10.1007/978-3-319-56782-2_872-3

Coelho, C. A. (2002). Story narratives of adults with closed head injury and non-brain-injured adults: Influence of socioeconomic status, elicitation task, and executive functioning. *Journal of Speech, Language, and Hearing Research, 45*(6), 1232–1248. https://doi.org/10.1044/1092-4388(2002/099)

Dalton, S. G. H., Hubbard, H. I., & Richardson, J. D. (2020). Moving toward non-transcription based discourse analysis in stable and progressive aphasia. *Seminars in Speech and Language, 41*(1), 32–44. https://doi.org/10.1055/s-0039-3400990

Dalton, S. G. H., & Richardson, J. D. (2019). A large-scale comparison of main concept production between persons with aphasia and persons without brain injury. *American Journal of Speech-Language Pathology, 28*(1S), 293–320. https://doi.org/10.1044/2018_AJSLP-17-0166

Elbourn, E., Kenny, B., Power, E., Honan, C., McDonald, S., Tate, R., Holland, A., MacWhinney, B., & Togher, L. (2019). Discourse recovery after severe traumatic brain injury: Exploring the first year. *Brain Injury, 33*(2), 143–159. https://doi.org/10.1080/02699052.2018.1539246

Fergadiotis, G., Wright, H. H., & Capilouto, G. J. (2011). Productive vocabulary across discourse types. *Aphasiology, 25*(10), 1261–1278. https://doi.org/10.1080/02687038.2011.606974

Forbes, M. M., Fromm, D., & MacWhinney, B. (2012). AphasiaBank: A resource for clinicians. *Seminars in Speech and Language, 30*(3), 217–222. https://doi.org/10.1055/s-0032-1320041

Fromm, D., Forbes, M., Holland, A., Dalton, S. G., Richardson, J., & MacWhinney, B. (2017). Discourse characteristics in aphasia beyond the Western Aphasia Battery cutoff. *American Journal of Speech-Language Pathology, 26*(3), 762–768. https://doi.org/10.1044/2016_AJSLP-16-0071

Galetto, V., Andreetta, S., Zettin, M., & Marini, A. (2013). Patterns of impairment of narrative language in mild traumatic brain injury. *Journal of Neurolinguistics, 26*(6), 649–661. https://doi.org/10.1016/j.jneuroling.2013.05.004

Gillam, S. L., Gillam, R. B., Fargo, J. D., Olszewski, A., & Segura, H. (2016). Monitoring Indicators of Scholarly Language: A progress-monitoring instrument for measuring narrative discourse skills. *Communication Disorders Quarterly, 38*(2), 96–106. https://doi.org/10.1177/1525740116651442

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

Hameister, I., & Nickels, L. (2018). The cat in the tree—Using picture descriptions to inform our understanding of conceptualisation in aphasia. *Language, Cognition and Neuroscience, 33*(10), 1296–1314. https://doi.org/10.1080/23273798.2018.1497801

Heilmann, J., Miller, J., Nockerts, A., & Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-age children. *American Journal of Speech-Language Pathology, 19*(2), 154–166. https://doi.org/10.1044/1058-0360(2009/08-0024)

Jorgensen, M., & Togher, L. (2009). Narrative after traumatic brain injury: A comparison of monologic and jointly-produced discourse. *Brain Injury, 23*(9), 727–740. https://doi.org/10.1080/02699050903133954

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings*. Oxford University Press.

Kong, A. (2009). The use of main concept analysis to measure discourse production in Cantonese-speaking persons with aphasia: A preliminary report. *Journal of Communication Disorders, 42*(6), 442–464. https://doi.org/10.1016/j.jcomdis.2009.06.002

Laerd Statistics. (n.d.). *Testing for normality in SPSS statistics*. Retrieved March 11, 2020, from https://statistics.laerd.com/premium/spss/tfn/testing-for-normality-in-spss.php

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. https://doi.org/10.1044/1092-4388(2010/09-0022)

Lê, K., Coelho, C., Mozeiko, J., Krueger, F., & Grafman, J. (2011). Measuring goodness of story narratives: Implications for traumatic brain injury. *Aphasiology, 25*(6–7), 748–760. https://doi.org/10.1080/02687038.2010.539696

Lê, K., Coelho, C., Mozeiko, J., Krueger, F., & Grafman, J. (2012). Predicting story goodness performance from cognitive measures following traumatic brain injury. *American Journal of Speech-Language Pathology, 21*(2), S115. https://doi.org/10.1044/1058-0360(2012/11-0114)

Lindsey, A., Hurley, E., Mozeiko, J., & Coelho, C. (2018). Follow-up on the Story Goodness Index for characterizing discourse deficits following traumatic brain injury. *American Journal of Speech-Language Pathology, 28*(1S), 330–340. https://doi.org/10.1044/2018_AJSLP-17-0151

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*(11), 1286–1307. https://doi.org/10.1080/02687038.2011.589893

Marini, A., Boewe, A., Caltagirone, C., & Carlomagno, S. (2005). Age-related differences in the production of textual descriptions. *Journal of Psycholinguistic Research, 34*(5), 439–463. https://doi.org/10.1007/s10936-005-6203-z

Marini, A., Zettin, M., Bencich, E., Bosco, F. M., & Galetto, V. (2017). Severity effects on discourse production after TBI. *Journal of Neurolinguistics, 44,* 91–106. https://doi.org/10.1016/j.jneuroling.2017.03.005

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282. https://doi.org/10.11613/BM.2012.031

Mozeiko, J., Lê, K., Coelho, C., Krueger, F., & Grafman, J. (2011). The relationship of story grammar and executive function following TBI. *Aphasiology, 25*(6–7), 826–835. https://doi.org/10.1080/02687038.2010.543983

Nicholas, L. E., & Brookshire, R. H. (1993). A system for scoring main concepts in the discourse of non–brain-damaged and aphasic speakers. *Clinical Aphasiology Conference, 21,* 87–99.

Nicholas, L. E., & Brookshire, R. H. (1995). Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. *Journal of Speech and Hearing Research, 38*(1), 145–156. https://doi.org/10.1044/jshr.3801.145

Paul, R. (1995). *Language disorders from infancy through adolescence: Assessment and intervention*. Mosby.

Pinto, T. C. C., Machado, L., Bulgacov, T. M., Rodrigues-Júnior, A. L., Costa, M. L. G., Ximenes, R. C. C., & Sougey, E. B. (2019). Is the Montréal Cognitive Assessment (MoCA) screening superior to the Mini-Mental State Examination (MMSE) in the detection of mild cognitive impairment (MCI) and Alzheimer's disease (AD) in the elderly? *International Psychogeriatrics, 31*(4), 491–504. https://doi.org/10.1017/S1041610218001370

Richardson, J. D., & Dalton, S. G. (2016). Main concepts for three different discourse tasks in a large non-clinical sample. *Aphasiology, 30*(1), 45–73. https://doi.org/10.1080/02687038.2015.1057891

Richardson, J. D., & Hudspeth, S. G. (2014). *Story grammar analysis in persons with mild aphasia*. The Aphasiology Archive. http://aphasiology.pitt.edu/id/eprint/2593

Roth, F. P., & Spekman, N. J. (1986). Narrative discourse: Spontaneously generated stories of learning-disabled and normally achieving students. *Journal of Speech and Hearing Disorders, 51*(1), 8–23. https://doi.org/10.1044/jshd.5101.08

Stark, J. A. (2010). Content analysis of the fairy tale Cinderella—A longitudinal single-case study of narrative production: "From rags to riches." *Aphasiology, 24*(6–8), 709–724. https://doi.org/10.1080/02687030903524729

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing* (pp. 53–120). Ablex.

Stout, C., Yorkston, K., & Pimentel, J. (2000). Discourse production following mild, moderate, and severe traumatic brain injury: A comparison of two tasks. *Journal of Medical Speech-Language Pathology, 8*(1), 15–25.

Wright, H. H., Capilouto, G. J., Wagovich, S. A., Cranfill, T. B., & Davis, J. E. (2005). Development and reliability of a quantitative measure of adults' narratives. *Aphasiology, 19*(3–5), 263–273. https://doi.org/10.1080/02687030444000732

Appendix A

Sequencing Scoring Rules and Examples

Main concepts (MCs) must be presented in order unless otherwise specified. An MC is not considered out of order if previous or subsequent MCs are omitted. If a lower numbered MC is included "after" a higher numbered MC or MCs, the lower numbered MC is considered out of order. If multiple utterances can be applied to one MC, code sequencing was based on the utterance that occurs first.

In the following example, all MCs are presented in order:

MC 6:	Line 1233	the king in the land had decided that it was time for his son to get married
MC 7:	Line 1242	to do that he wanted to have a great ball
MC 8:	Line 1256	Cinderella's mother had gotten the invitation

In the following example, MC 27 follows a higher numbered MC and is therefore out of order:

MC 26:	Line 988	and the coach takes her home
MC 28:	Line 1013	and she hears the duke come to the door
MC 27:	Line 1026	so he's going to all the houses in the kingdom to see which maiden can fit into the glass slipper

For each MC that is present, assign one of the following order scores:

3 points: MC stated in the correct order
2 points: MC not stated in the correct order but marked as being out of order by the speaker (e.g., "I forgot to say"; "and in the process of")
1 point: MC not stated in the correct order and not marked as being out of order by the speaker
0 points: MC absent

The following are exceptions to sequencing scores based on the story grammar function that the MC serves:

1. MCs 3–5 are all related to setting and may be presented in any order amongst themselves.
2. MCs 6–8 are all initiating events and may be presented in any order amongst themselves.
3 MC 9 is a mental state and may be present after MC 7 or 8, which are both initiating events.
4. The order of MCs 10 and 11 is interchangeable, unless MC 10 is coded as an attempt, in which case it must precede MC 11.
5. MC 13 is a mental state that may follow MC 10, 11, or 12.
6. MCs 15 and 16 are both attempts and their order is interchangeable.
7. If MC 18 is coded as a direct consequence for Episode 2, then it must come before, after, or in place of MC 17. If MC 18 is coded as an initiating event for Episode 4, then it must come before, after, or in place of MC 21.
8. The order of MCs 19 and 20 is interchangeable.
9. MCs 24–26 are direct consequences may be presented in any order amongst themselves.

Appendix B

Story Grammar Coding Examples

Story grammar codes are preassigned for 31 main concepts (MCs). Coders make decisions for 5 MCs based on how the speaker presents the content. See Richardson and Dalton (2016) for details of each MC.

MCs with preassigned story grammar codes are listed by episode and story grammar element:

Setting	Episode 1	Episode 2	Episode 3	Episode 4	Episode 5	Conclusion
MC 1–5	*IE:* MC 6–8 *A:* MC 11 *DC*: MC 12 *MS:* MC 9, 13	*IE:* MC 14 *A:* MC 15 *DC:* MC 16	*See below*	*IE:* MC 21 *A:* MC 22 *DC:* MC 23–26	*IE:* MC 27 *A:* MC 28, 29, 31 *DC:* MC 30, 32	MC 33, 34

Note. *IE* = initiating event; *A* = attempt; *DC* = direct consequence; *MS* = mental state.

Rules for coding the story grammar of MCs without preassigned story grammar codes are listed below, in order of occurrence:

MC 10: *Episode 1 Attempt* or *Episode 1 Direct Consequence*
Code as an attempt if a condition is included:
Cinderella cannot go to the ball unless she does her chores.
Code as a direct consequence if a reason is included:
Cinderella cannot go to the ball because her dress is torn.
If neither a condition nor reason is included, code was based on sequencing:
The stepsisters tear her dress and she can't go (Direct Consequence)

MC 17: *Episode 2 Direct Consequence* and/or *Episode 3 Initiating Event*
Code as a direct consequence if the wording indicates that Cinderella left/goes to the ball:
Cinderella went to the ball.
Code as an initiating event if the wording indicates that Cinderella arrived/got to the ball:
Cinderella reached the ball.
Code as both a direct consequence and initiating event if wording indicates that Cinderella left/goes to the ball and arrived/got to the ball.
So off to the ball she goes in her carriage and her beautiful dress and her glass slippers. Cinderella gets to the ball.

MC 18: *Episode 2 Direct Consequence* or *Episode 4 Initiating Event*
Code as a direct consequence if this MC is presented at the end of Episode 2:
Everyone went off to the ball…told her that at the stroke of midnight she had to be home.
Code as an initiating event if this MC is presented at the beginning of Episode 4:
Of course, the prince falls in love with her. But then Cinderella finally remembers that the fairy godmother warned her that all the stuff turns back to what it used to be at midnight.

MC 19: *Episode 3 Attempt* or *Direct Consequence*
Code as an attempt if it precedes MC 20, or if MC 20 is not included by the speaker.
He danced with her and he fell in love with her.
Code as a direct consequence if it follows MC 20.
The prince falls in love with her instantly. As they're taking one of their last dances, she hears the gong from the clock tower.

MC 20: *Episode 3 Attempt* or *Direct Consequence*
Code as an attempt if it precedes MC 19.
The prince falls in love with her instantly. As they're taking one of their last dances, she hears the gong from the clock tower.
Code as a direct consequence if it follows MC 19, or if MC 19 is not included by the speaker.
But he dances with her the rest of the night, and he falls in love with her.