Summer 7-14-2021

Main Concept, Sequencing, and Story Grammar (MSSG) Analyses of the Cinderella Story in Latent Aphasia

Janet B. Adams

University of New Mexico - Main Campus

Janet Adams

Candidate

Speech and Hearing Sciences

Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Jessica Richardson, Ph.D., CCC-SLP, Chairperson

Rick Arenas, Ph.D.

Sarah Grace Hudspeth Dalton, Ph.D., CCC-SLP

Main Concept, Sequencing, and Story Grammar (MSSG) Analyses of the Cinderella Story in Latent Aphasia

BY

Janet Adams

B.A., SPEECH AND HEARING SCIENCES,

UNIVERSITY OF NEW MEXICO, 2018

THESIS

Submitted in Partial Fulfillment of the

Requirement for the Degree of

Masters of Science

Speech-Language Pathology

The University of New Mexico

Albuquerque, New Mexico

July 2021

Main Concept, Sequencing, and Story Grammar (MSSG) Analyses of the Cinderella Story in Latent Aphasia

By Janet Adams

B.A., Speech and Hearing Sciences, University of New Mexico, NM 2018

M.S., Speech-Language Pathology, The University of New Mexico 2021

ABSTRACT

Commonly used standardized tests, like the Western Aphasia Battery-Revised (WAB-R), are not sensitive to higher level discourse deficits, leading to certain individuals not meeting diagnostic criteria for aphasia. Consequently, individuals with aphasia are excluded from receiving potentially beneficial services and/or from being included in research. In a large sample of persons with stroke-induced aphasia, this study analyzed discourse samples using Main Concept, Sequencing, and Story Grammar (MSSG) Analyses to examine macrostructural discourse characteristics of persons with latent aphasia (PWLAs) compared to persons with no underlying brain injury (PNBIs) and persons classified as having anomic aphasia (PWAAs) by the WAB-R. A secondary analysis was conducted comparing PWLAs and a subgroup of PWAAs who scored in the upper range on the WAB-R (PWAAs-top). Results revealed PWLAs demonstrate macrostructural discourse impairments differing from PNBIs, PWAAs, and PWAAs-top and should be considered a unique aphasic group deserving of clinical attention and services.

Table of Contents

List of Figures

List of Tables

Main Concept, Sequencing, and Story Grammar (MSSG) Analyses of the Cinderella Story in Latent Aphasia

CHAPTER 1

INTRODUCTION

Latent Aphasia

Previous research highlights a population of individuals with a previous diagnosis of aphasia who describe the negative impact of aphasia in their lives, but who do not meet diagnostic criteria for aphasia on commonly used standardized tests (e.g., the Western Aphasia Battery-Revised [WAB-R]; Kertesz, 2007). These individuals receive labels such as "clinically undetectable" aphasia (Olsen et al., 1986), latent aphasia (Boller & Vignolo, 1966; DeDe & Salis, 2020; Pichot, 1955; Silkes et al., 2020), not aphasic by WAB (NABW; Dalton & Richardson, 2015; Fromm et al., 2017), and very mild aphasia (Cavanaugh & Haley, 2020). Due to the lack of sensitivity to mild deficits of the WAB-R and similar assessments, such individuals have erroneously been described in the literature as "recovered" and have even been included as "controls" during normative reference development (Kertesz, 2007). However, closer examination of these individuals reveals everyday language difficulties, activity and participation limitations, and negative psychosocial impacts (e.g., Armes et al., 2020). Current mainstream assessment measures lack the sensitivity to detect their mild language deficits. Consequently, certain individuals who should qualify for clinical services or meet research inclusion criteria are excluded from receiving potentially beneficial services. Unfortunately, treatment programs that address these higher-level deficits are also scarce. In order to provide opportunities and services for this population, a detailed characterization of these higher-level deficits for this population is needed. For the purposes of this study, we will refer to this group as persons with latent aphasia (PWLAs).

Latent Aphasia and Discourse Assessment

Previous research has documented deficits in other measures such as discourse, which is now listed as a primary outcome measure for aphasia treatment research (Brady et al., 2016), and revealed that PWLAs significantly differ from both closest neighbors – persons not brain injured (PNBIs) and persons with anomic aphasia (PWAAs). Studies analyzing Core Lexicon (CoreLex), a measure of typical vocabulary needed to produce meaningful discourse during picture description and storytelling tasks, revealed significant differences between healthy controls (PNBIs), persons with diagnosed aphasia (PWAs), and persons with latent aphasia (PWLAs) (Dalton & Richardson, 2015; Dalton et al., 2020). Other studies revealed significant differences in discourse performance between PWLAs and PNBIs, suggesting they perform closest to high-scoring PWAAs (Dalton et al., 2020; Dalton & Richardson, 2015, 2019; Fromm et al., 2017) for Main Concept production (i.e., a measure of essential content and how accurately and completely an individual conveys the content) and derived efficiency measures (i.e., accurate and complete concepts per minute) during a variety of discourse tasks. PWLAs performance also differed from PNBIs and PWAAs for other microstructural (i.e., measures informativeness below the sentence level) and temporal discourse measures (e.g., words per minute [WPM], moving average type-token ratio [MATTR], mean length of utterance [MLU], etc.), so that PWLAs do not fit neatly into either group. Higher-level language measures are still needed to characterize and/or detect mild aphasic deficits in order to justify the clinical need for this population and highlight treatment targets. While the focus of this study is on spoken language abilities and deficits, it should be noted that very mild (i.e., “sub-clinical”) auditory comprehension deficits in persons with latent aphasia have been detected via in-depth assessment using the Token Test (Boller & Vignolo, 1966).

Hybrid micro/macrostructural discourse measures, such as Main Concept Analysis (MCA), characterize deficits of different aphasia subtypes and capture differences between PNBIs and PWAs as well as between aphasia subtypes. As reviewed briefly above, MCA studies have helped to focus attention on the needs of PWLAs. There is a current lack of information regarding the added usefulness of other discourse measures for sensitive detection of high-level deficits, and further exploration into these measures is warranted. Story grammar measures (i.e., measures narrative organization and how successful relationships between characters, agents, and events are conveyed) may provide useful information, as more recent studies have revealed deficits in grammar and syntax of persons with mild aphasia (Coelho et al., 1994; Armstrong, 2000; Beeke et al., 2003; Andreeta et al., 2012) that will be reviewed below. Further, story grammar and other measures of discourse macrostructure, such as reference chains, topic coherence, and Predicate Argument Structure (PAS), have been recently endorsed as informative and psychometrically strong measures (Pritchard et al., 2018). Recently, Greenslade et al. (2020) developed a multilevel measure that combines analysis of both informativeness and story grammar called Main Concepts, Sequencing, and Story Grammar Analyses, or MSSG. The following sections will discuss the development of MSSG, its components, and how they are combined to analyze micro/macro structural discourse performance.

Main Concept Checklists and Analysis

Main concept analysis (MCA) was first introduced by Nicholas and Brookshire (1993) and provides information about the presence of essential content and how accurately and completely an individual conveys the content in story retelling tasks. Nicholas and Brookshire (1995) developed this checklist-based analysis from utterances produced by 80% or more of 20 non-brain-damaged adult controls (10 males, 10 females; mean age 64.2 [SD = 7]; mean years of education 12.8 [SD = 2.2]) during structured discourse tasks. MCA checklists have since been created to analyze different types of discourse: picture sequence descriptions (i.e., Broken Window and Refused Umbrella), picture description (i.e., Cat in the Tree), story retell (i.e., Cinderella), and procedural discourse (i.e., Peanut Butter and Jelly Sandwich) (Kong, 2009; Richardson & Dalton, 2016, 2019). MCA is conducted by prompting an individual to complete the discourse task, orthographically transcribing their responses, comparing their utterances to the existing checklists, and scoring them on the bases of presence and accuracy of their produced concept relative to the target concept on the checklist. Each utterance that corresponds to a target main concept (MC) receives one of the following codes: accurate-complete (AC), accurate-incomplete (AI), inaccurate-complete (IC), Inaccurate-incomplete (II) and absent (AB) (Nicholas & Brookshire, 1995; Richardson & Dalton, 2016). Once all MCs are coded, the codes are assigned a numeric value (i.e., 3 points for AC, 2 points for AI and IC, 1 point for II, and 0 points for AB) and then totaled to provide the MC composite score, which represents a combination of accuracy and completeness abilities. Of note, Kong (2009) first developed a composite score formula which is nearly identical to the one above, but the IC and II codes were combined into a single coding category - IN (inaccurate) - and both received 1 point. Normative research utilizing and developing MC checklists has revealed significant differences in the number of total accurate and complete concepts (AC) and MC composite scores between age groups (i.e., 59 and younger v. 60 and older). This checklist development and normative research supported the use of MCA as a potentially useful measurement for analyzing components for relaying discourse narrative in PWAs (Richardson & Dalton, 2016).

Main Concepts in Aphasia

Recently, much work has been accomplished with MCA in PWAs. For example, Dalton and Richardson (2019) conducted a large-scale study comparing MC production between 238 PWAs and 145 PNBIs. They found PWAs exhibited lower informativeness scores (i.e., MC composite scores), made fewer MC attempts overall, and more frequently produced statements that were inaccurate and/or incomplete when compared to PNBIs for all main concept measures (i.e., MC composite scores, MC attempts, and each MC code) and all discourse tasks (i.e., Broken Window, Refused Umbrella, Cat Rescue, Cinderella, and PB&J sandwich). The data showed varying patterns of differences by aphasia subgroup and discourse task when examined against PNBIs. The results demonstrated the benefit of utilizing MCA in research as well as clinical practice, as the approach was sensitive to differences in discourse as a result of aphasia. Additional evidence supporting MCA's sensitivity to group differences (i.e., PNBIs, PWAs, and aphasia subtypes) was provided by Kong and colleagues (Kong et al., 2016). Perhaps most relevant for this present study is the work by Fromm et al. (2017) who investigated discourse characteristics (i.e., words per minute [WPM], percent word errors, lexical diversity [moving average type-token ratio; MATTR], MC production, number of utterances, mean length of utterance [MLU], and proposition density) in PWLAs (i.e., individuals who scored at or above the WAB-R aphasia quotient of 93.8, referred to as not aphasic by WAB [NABW] in the original study). They found significant differences between PNBIs and the PWLAs on all discourse measures except proposition density and MLU, where both groups were found to be statistically similar. Further comparison of the PWLAs and PWAAs found significant differences in MLU, but both groups were statistically similar when looking at total utterances. The authors performed a secondary analysis between the PWLA group and higher scoring PWAAs with the same AQ range (6.2 points) as PWLAs, instead of the 30-point range for the entire group of PWAAs. Comparison between the PWLAs and the highest scoring PWAAs (i.e., WAB-R AQ range 87.5 to 93.7), designated the PWAAs-top, revealed significant differences between the PWLAs and the PWAAs-top on all discourse measures except total utterances and MC production. Their findings demonstrated that PWLAs should not be categorized as "recovered" because their discourse characteristics were significantly different from PNBIs and that their discourse impairments warranted further investigation.

MCA was recently endorsed as a psychometrically strong measure because it consistently demonstrates high levels of acceptability, intra- and interrater reliability, content validity, convergent validity, and discriminant validity (Pritchard et al., 2017). In addition, a recent study conducted by Cruice et al. (2020) surveyed current SLP practices and views of discourse analysis to identify barriers and facilitators for clinical feasibility. For SLPs who reportedly did not use discourse analysis, the following barriers were described: (1) time constraints, (2) patient severity, (3) lack of expertise, training, confidence, resources and equipment. Overall, clinicians were highly engaged in completing aspects of discourse analysis but improvements needed to be made to increase its use in clinical practice (e.g., more accessible discourse measures, reduced training requirements, etc.). Taken together, MCA is a promising tool for efficient and informative discourse analysis that, given its sensitivity to mild aphasic deficits and reliance on orthographic transcription to date, is likely to be useful in clinical settings (Dalton et al., 2020).

Limitations of Main Concepts. While MCA has demonstrated good sensitivity to distinguishing aphasia subtypes, characterizing deficits, and is psychometrically strong, it has some shortcomings when looking at the whole picture (i.e., discourse organization, how successful relationships between characters, agents, and events are conveyed, etc.) of narrative discourse deficits. MCA normative development to date is potentially limited by gender bias and a lack of diversity in race, ethnicity, and age (Richardson & Dalton, 2016). Previous researchers utilizing MCA have revealed common scoring errors related to pronoun referents (i.e., pronouns that replace nouns that have been previously stated) and the application of the completeness rule (i.e., statements containing some of the essential information). As scores can interpret utterances differently from other scorers, an utterance may be assigned a code of accurate/incomplete (AI) and other scores may assign the same utterance a code of inaccurate/complete (IC), which can cause fidelity issues. Finally, and most importantly for this project, MCA focuses solely on the essential content provided but does not take into account how the content is organized and sequenced, or how much of the framework of the story is provided.

Story Grammar Analyses

Story grammar analyses are well-known and commonly used narrative analyses that identify whether or not speakers are able to produce the structured "cognitive skeleton" of a story. Story grammar measures aim to capture how narratives are organized and whether speakers are successful in conveying logical relationships between characters, agents, and events within and between episodes of the story narrative. Story grammar analyses commonly examine episodes, which are composed minimally of three components: 1) an initiating event (i.e., event that prompts the character to engage in a goal directed behavior), 2) an attempt (i.e., an action intended to attain a goal), and 3) a direct consequence (i.e., the outcome of the attempt/action in relation to the goal). An episode can also include internal responses (i.e., a character's thoughts, feeling and reaction) and plans (i.e., intended character actions), but those components are not required for an episode to be scored as a complete/complex episode. Episodes that include two or more of the required components are labeled as a complex episode, as they are thought to provide the listener with sufficient information to understand or infer the sequencing and adequacy of the narrative's events (Greenslade et al., 2020; Richardson et al., 2021).

Story Grammar in Aphasia (and other Communication Disorders)

Story grammar research has focused more on individuals with cognitive and/or communication deficits following TBI. Lê et al. (2011) examined narrative in cognitive-communication disorders to validate a new measure – the Story Goodness Index (SGI), a two-dimensional classification system of discourse organization and completeness. The SGI allows for visual evaluation of narrative organization and informativeness (i.e., statements considered accurate/complete). To obtain the SGI, utterances (i.e., T-unit) are first coded as one of the essential story grammar components that makes up an episode (i.e., initiating event, attempt, or direct consequence) and are then totaled. This is the story grammar score. Next, the story's completeness is analyzed utilizing a checklist similar to the MC checklists described above. The number of correctly produced utterances (similar to an accurate and complete (AC) code for MCA) provides the story completeness score. The SGI is then visualized by plotting story grammar and completeness scores onto an x,y graph for interpretation, where: Quadrant 1: low content-high organization; Quadrant 2: high content-high organization; Quadrant 3: low content-low organization; and Quadrant 4: high content-low organization (Lê et al., 2011). Using the Story Goodness Index (SGI), Lê et al. showed that PNBIs were more likely to fall into Quadrant 2 of the index, meaning that their discourse had high content and high organization, whereas individuals with TBI were more likely to fall into Quadrant 1, meaning that their discourse was organized but contained incomplete content. This demonstrated the potential of SGI to examine discourse in TBI and other clinical populations for sensitive detection of macrolinguistic deficits and potential determination of treatment targets (Lê et al., 2011).

The study conducted by Stout et al. (2000) also analyzed how TBI severity influences discourse performance for the following story grammar variables during a picture description and story retelling task: number of T-units, number of mazes, and total number of content units. They found that PWTBIs performed significantly different from PNBIs (e.g., spoke more slowly, produced longer mazes, had shorter retellings, and conveyed less content) but displayed varying patterns of narrative deficits, possibly due to the differing cognitive demands of the task. This was supported by a study conducted by Galetto et al. (2013) who examined narrative discourse impairment in PWTBIs (mild). Their results showed higher incidence of errors in global coherence and lower levels of lexical informativeness. They suggested that TBI-induced cognitive impairment negatively affected narrative organization and macrolinguistic skills (Galetto et al., 2013). Power et al. (2020), investigated a picture description narrative (i.e., Cat Rescue) in participants with severe TBI from acute recovery to 6 months post TBI. Included in their analysis are story grammar variables of episodic complexity and proportion of utterances within an episode structure (i.e., the number of utterances within episode structure divided by the total number of utterances in the story). Their findings show that PWTBIs produced less complete episodes in a story retelling than PNBIs. They also found no significant changes between 3 and 6 months of recovery to narrative production. However, individual improvement over time was observed by the authors (Power et al., 2020).

Overall, little is known about story grammar or its relationship to other discourse measures in PWAs. As reviewed in Richardson et al. (2021), there are a few noteworthy studies of story grammar in PWAs that utilize a wide variety of analytic approaches. Ulatowska et al. (1981) examined performance of narrative superstructure elements (i.e., story grammar components: setting, complicating event, and resolution as well as optional components: codas, abstracts, evaluation) in 10 PNBIs and 10 PWAs (i.e., 7 mild, 3 mild-moderate) during structured narrative tasks (i.e., telling stories, producing summaries, giving morals to the stories, and producing procedures). The results revealed no significant differences from PNBIs for story grammar variables, leading the authors to suggest that PWAs with mild deficits had intact sequencing and produced the minimum criteria for complete episodes. Ulatowska et al. (1983) also examined a group of 15 PNBIs and 15

PWAs with moderate deficits during the following tasks: telling, summarizing, and giving morals to stories. They found that PWAs with moderate impairments produced fewer required and optional story grammar components and demonstrated impaired sequencing abilities when compared to PNBIs. In a case study of an individual with mild fluent aphasia throughout his recovery up until 1 year post stroke conducted by Coelho et al. (1994), they analyzed changes in narrative discourse abilities, specifically story grammar (i.e., number of complete episodes). The study revealed impaired story grammar abilities (e.g., 0 complete episodes produced at baseline and 2 complete episodes produced at 12 months) that only slightly improved over the 12-month recovery period, whereas other measures of language abilities (i.e., Porch Index Communicative Ability [PICA]) showed improvement. The authors suggest that multilevel analysis, including story grammar, is an appropriate and sensitive measure for assessing change in communication abilities over recovery, and importantly, may identify deficits not detected via other more commonly used instruments. Another study investigating story grammar outcomes of 2 PWAs (i.e., 1 fluent and 1 nonfluent) following completion of the discourse treatment NARNIA (i.e., Novel Approach to Real-life communication: Narrative Intervention in Aphasia), showed significant improvements in story grammar (Whitworth, 2010).

Limitations of Story Grammar. While the results from previous studies utilizing story grammar indicate its sensitivity for detecting narrative deficits in clinical populations, there are several limitations to using this measure. First, there is a lack of consistent usage of story grammar measures, preventing comparison across studies. Second, there is a lack of control data (i.e., sample size, age ranges, etc.) for story grammar measures. Finally, although story grammar measures provide information about a story's completeness, it does not take into account how accurately and/or completely one conveys the story, which may limit their discriminatory power (Greenslade et al., 2020).

Main Concept, Sequencing, and Story Grammar (MSSG) Analyses

Given the limitations described above for MCA and Story Grammar analyses as separate analyses, these approaches were combined into an analytic approach titled Main Concept, Sequencing, and Story Grammar (MSSG) analyses (Greenslade et al., 2020). For MSSG, MCA coding and composite scoring is completed utilizing the Cinderella MC checklist created by Richardson and Dalton (2016). Each MC produced then receives a story grammar code (i.e., initiating event, attempt, direct consequence, etc.) based on its purpose in the narrative. The resulting story grammar component code is then given a score as follows: 2 points if two components are present, 1 point if one component is present, and 0 points if the component is absent. After the MCs and SG components are scored and coded, sequencing rules are applied to each present MC as follows: 3 points for MCs stated in correct order, 2 points for MC stated in incorrect order and is marked by the speaker as being out of sequence, 1 point for MC not stated in the correct order and are not marked by the speaker as being out of sequence, and 0 points for MCs that are absent. Finally, the coder will score each episodic component as present (1) or absent (0) to calculate the total episodic components score and the episodic complexity score. The five episodes of the Cinderella narrative are assigned a complexity score of (1) if it includes at least two of the three required components and a score of (0) if it includes one or zero components, resulting in a maximum episodic complexity score of 5.

Greenslade et al. (2020) investigated the usefulness of MSSG for measuring age-related changes in 92 PNBIs (i.e., age 20-39: 22, age 40-59: 23, age 60-79: 24, and age 80+: 23). The data revealed differences in MSSG variables (i.e., sequencing, MC + sequencing, total episodic component, and episodic complexity) when comparing participants 60 and older to scores from participants aged 20-59. These results support the use of all MSSG variables for identifying age-related changes and disorder-related impairments in narrative macrostructure. In a follow-up paper, Richardson et al. (2021) applied MSSG to 110 PNBI and 370 PWA (i.e., 122 anomic, 85 Broca's, 67 conduction, 54 NABW, 12 transcortical motor, and 24 Wernicke's) to characterize differences between PWAs and PNBIs as well as differences between aphasia subtypes. The authors also completed a secondary analysis focusing on demographic variables (i.e., age, gender, and education) and how they influence MSSG outcomes. They found that there were significant differences between PWAs and PNBIs for all MSSG variables (i.e., MC, sequencing, MC + sequencing, essential story grammar components, total episodic components, and episodic complexity). Differences in discourse performance were also noted between each aphasia subtype and those classified as having latent aphasia (i.e., not aphasic by WAB [NABW]). As a result, the authors called for further investigation of differences between PNBIs, PWLAs, and PWAAs. Their secondary analysis on demographic variables revealed age and gender as significant predictors of PNBI's discourse performance for all MSSG variables, except total episodic components, where age was the only significant predator. For PWAs, education and gender were the significant predictors of discourse performance for all MSSG variables.

Purpose of the Present Study

The purpose of this study is to further examine macrostructural discourse characteristics of persons with mild anomic and latent aphasia utilizing the AphasiaBank database along with our own participant database. This study will characterize discourse performance on all MSSG variables (i.e., main concepts, story sequencing, and story grammar) for the Cinderella storytelling task utilizing the procedures outlined in Greenslade et al. (2020) and Richardson and Dalton (2016). Previous work by Richardson and colleagues (2021) provided descriptive statistical and comparative information for MSSG variables (but not for MC codes) for PNBIs and PWAs of varying subtypes. We will provide descriptive statistical information and compare the discourse performance (MSSG variables and MC codes) of three specific groups: Group 1 - persons with no brain injury (PNBIs); Group 2 - persons with stroke-induced latent aphasia who were diagnosed as non-aphasic by the WAB-R, but still report and demonstrate impaired language (PWLAs); and Group 3 - persons with stroke-induced clinically-diagnosed anomic aphasia (PWAAs). As in Fromm et al. (2017), we will conduct an analysis between PWLAs and a subgroup of PWAAs who scored in the upper range on the WAB-R (PWAAs-top). In a larger sample than has been reported upon previously, we will provide an in-depth characterization and analysis of all discourse variables. This research aims to effectively capture narrative discourse deficits of PWAs who are most often at or beyond the margins of service delivery in order to improve diagnostics, provide justification for services, and inform planning and delivery of treatment.

CHAPTER 2

METHODS

Participants and Transcripts

Transcripts for 112 persons with no brain injury (PNBIs), 60 persons with latent aphasia (PWLAs), and 139 persons with anomic aphasia (PWAAs) have been retrieved from from AphasiaBank and our lab database, with databases including the following: Adler Aphasia Center (Adler), Aphasia Center of California (Elman), Aphasia Center of Tucson (Tucson), Aphasia Center of West Texas (ACWT), Aphasia House (Whiteside), Aphasia Lab of the University of South Carolina (Fridriksson), Boston University (BU), Carnegie Mellon University (CMU), East Carolina University (Wright), Emerson College (Kempler), InteRACT: Intensive Residential Aphasia Communication Therapy (Wozniak), Montclair State University (MSU), Northwestern University (Thompson), private practice (Garrett), Stroke Comeback Center (Williamson), Snyder Center for Aphasia Life Enhancement (SCALE), Stroke Aphasia Recovery Program (STAR), Texas Christian University (TCU), Triangle Aphasia Project (TAP), University of Kansas (Kansas), University of Kentucky (Capilouto), University of Massachusetts - Amherst (Kurland), University of New Hampshire (UNH), University of New Mexico (Richardson), and University of North Carolina - Chapel Hill (Haley, Jacks).

For PNBIs, the average age was 58.3 years (*SD* = 20.8) with an average education of 15.8 years (*SD* = 2.5). For PWLAs, the average WAB-R Aphasia Quotient (AQ) score was 96.7 (*SD* = 1.8), the average age was 60.3 years (*SD* = 13.2), and the average years of education was 15.9 (*SD* = 2.7). For PWAAs, the average WAB-R Aphasia Quotient (AQ) score was 85.3 (*SD* = 6.5), the average age was 62.1 years (*SD* = 12.5), and the average years of education was 15.8 (*SD* = 2.9). For the smaller group, PWAAs-top (where the WAB-R AQ scores range is between 87.5 and 93.7 as in Fromm et al. [2017]), the average WAB-R Aphasia Quotient (AQ) score was 90.7 (*SD* = 1.8), the average age was 61.8 years (*SD* = 11.4), and the average years of education was 15.7 (*SD* = 2.7).

Discourse samples were collected following the AphasiaBank protocol and procedures (https://aphasia.talkbank.org/protocol/). Each participant reviewed a wordless version of the Cinderella picture book by Grimes (Grimes, 2005) before attempting the story retell. Once the participant indicated they were finished reviewing the book, the clinician provided the following instructions: *"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."* Transcripts were then orthographically transcribed and converted in CHAT to be scored and coded for the variables described below. For transcripts that had not yet been submitted to the AphasiaBank database, CHAT transcripts or orthographic transcripts were assigned line numbers and then coded and scored.

Table 1. Demographic information

	PNBI (N = 112)	**PWLA (N = 60)**	**PWAA (N = 139)**	**PWAA-top (N = 63)**
Age (years)	58.3 (± 20.8) 20 – 89.5	60.3 (± 13.2) 25 – 88	62.1 (± 12.5) 18 – 85.7	61.8 (± 11.4) 34.4 – 83.2
WAB-R Aphasia Quotient (AQ)	N/A	96.7 (± 1.8) 93.8 – 100	85.3 (± 6.5) 63.4 – 93.7	90.7 (± 1.8) 87.8 – 93.7
Gender	65 Female 47 Male 2 Unspecified	30 Female 20 Male	57 Female 82 Male	24 Female 39 Male
Education (years)	15.8 (± 2.5) 11 – 23	15.9 (± 2.7) 12 – 21	15.8 (± 2.9) 11 – 23	15.7 (± 2.7) 12 – 22
Race/Ethnicity	94 Caucasian 3 African American 3 Hispanic/Latino 12 Unspecified	49 Caucasian 2 African American 8 Hispanic/Latino 1 Other	122 Caucasian 12 African American 2 Hispanic/Latino 1 American Indian 1 Mixed 1 Unspecified	57 Caucasian 5 African American 1 Unspecified

PNBI: persons non-brain injured, PWLA: persons with latent aphasia, PWAA: persons with anomic aphasia, PWAA-top: persons with anomic aphasia whose WAB-R AQ is between 87.5 – 93.7

Discourse Analysis

Main Concepts

Transcripts were scored for main concepts (MCs) utilizing the checklist-based analysis produced by Nicholas and Brookshire (1995) and Richardson and Dalton (2016). Each MC consists of a main verb and its constituent arguments, and can include prepositional phrases and/or subordinate clauses (e.g., MC18: She knew 1) ***she*** 2) **had to be** 3) ***home by midnight*** because everything will turn back at midnight). MCs are coded based on presence and accuracy. An utterance that corresponds to a main concept (Cinderella – 34 MCs) can receive one of the following five codes: accurate/complete (AC), accurate/incomplete (AI), inaccurate/complete (IC), inaccurate/incomplete (II), or absent (AB) (Nicholas & Brookshire, 1995). Each checklist includes alternative productions to aid in scoring (e.g., alternatives for "Cinderella was upset" include "Cinderella was sad", "Cinderella was crying", etc.). These codes are then converted to numeric scores using the formula presented in Richardson and Dalton (2016): (3 x AC) + (2 x AI) + (2 x IC) + (1 x II). The scores for each MC were calculated within stories to provide a story MC composite score, which provides information about the individual's overall performance. The maximum composite score for the *Cinderella* story retell task was 102. We also reported on each MC code, and the maximum for any MC code is 34 as that is the total number of MCs in *Cinderella*. In previous work, we have reported MC attempts by adding the number of statements produced by the participant that were coded as AC, AI, IC, and II (Dalton & Richardson 2019). As this is essentially equivalent to essential story grammar components (see explanation below), we only report here on essential story grammar components.

Story Sequencing

After the completion of MC scoring, line numbers were added to the orthographic transcripts to determine whether MCs were produced in a logical sequence using Appendix A of Greenslade et al. (2020). For each MC to be considered logically sequenced, it had to come after the MC before it (e.g., MC 1 followed by MC 2 followed by MC 3, etc.), except when the sequence was judged to be interchangeable and MCs could be presented in any sequence among themselves (e.g., MCs 3 – 5; "Stepmother/stepsisters were mean to Cinderella", "Cinderella was a servant", and "Cinderella has to do the housework"). Sequencing rules are applied to each present MC as follows: 3 points for MCs stated in correct order, 2 points for MCs stated in incorrect order and is marked by the speaker as being out of sequence (e.g., “I forgot to say”), 1 point for MCs not stated in the correct order and are not marked by the speaker as being out of sequence, and 0 points for MCs that are absent. The maximum sequencing score for *Cinderella* was 102 (34 MCs x 3 points).

Main Concepts and Sequencing

Once the MCs were scored and sequenced, a combined score was calculated by adding the MC and sequencing scores for each concept (MC + sequencing). Each utterance corresponding to a main concept could receive a maximum score of 6 for this measure (AC MC [3 points] + correctly sequenced [3 points]). These scores were then combined across the 34 main concepts for a maximum total of 204 (34 accurate and complete MCs x 3 points, plus 34 recognizable MC attempts in correct order x 3 points).

Essential Story Grammar Components

Each of the 34 MCs received a story grammar code (i.e., setting, initiating event, attempt, direct consequence, mental state, and conclusion). Of the 34 MCs, twenty-nine were pre-assigned one of the six story grammar component codes. The remaining 5 MCs were assigned coding options based on the order of their production, neighboring MCs, and/or specific formulation to determine its story grammar component label. MC 17 (i.e., “Cinderella went/goes/arrives to the ball) provides an example of this occurrence, as the story grammar component assigned to it was determined based on whether the verb indicated Cinderella was on the way to the ball (direct consequence) or was at the ball (initiating event). After each MC received a story grammar component it receives a numerical value; absent MCs (0 points), present MCs (1 point). MC 17 is the only exception to this rule, as it could receive 2 points if the speaker indicated both that Cinderella was going to the ball and that she arrived at the ball. These scores were combined for a maximum total score of 35. Essential story grammar components were not reported in the original MSSG study by Greenslade et al (2020) but was reported by Richardson et al. (2021). We report it here in this study because it provides information about the amount of recognizable attempts at essential concepts and story grammar components. This variable can also be compared to previously reported MC attempts (Dalton & Richardson, 2015; Dalton & Richardson, 2019).

Total Episode Components

Each MC was pre-assigned an episode number (1-5), or episode options, as in the case of MC 17 “Cinderella went to the ball” and MC 18 “She had to be home by midnight” where the order and formulation determined the story grammar component and corresponding episode. An episode is made up of an initiating event, attempt, and direct consequence and can receive a total episodic components score of 3 if all components are present, 2 if only two components are present, 1 if only one component is present, and 0 if all components are absent. If each required episodic story grammar component is produced for all episodes, the maximum score is 15 (5 episodes x by 3 episodic story grammar components).

Episodic Complexity

Episodes that contained at least two of the required episodic story grammar components (i.e., initiating event, attempt, and direct consequence) were considered complex episodes. The five episodes of the Cinderella narrative were assigned a complexity score of

(1) if it included at least two of the three required components and a score of (0) if it included one or zero of the required components. The complexity scores of the five episodes were combined and could result in a maximum episodic complexity score of five if all episodes were complex.

MSSG Classification

Greenslade et al. (2020) introduced a modified version of the Story Goodness Index (SGI) (Coelho et al., 2013; Lê et al., 2011) and plotted the participants' MC+sequencing dimension (x-axis) against their total episodic components dimension (y-axis) for the MSSG Classification (see Figure 1). This allowed for a visual representation of the relationship between the participants' ability to tell accurate, complete, and logically sequenced stories (i.e., MC+sequencing) and their ability to maintain overall episodic structure (i.e., total episodic components) (Greenslade et al., 2020). Each quadrant was defined by cut off points at 1 *SD* (i.e., solid lines) and 2 *SD* (i.e., dotted lines) below the control mean for MC+sequencing and total episodic components (See Table 2 for *M* and *SD* values for each variable). For MC+sequencing, 1 *SD* below the mean was 87.4, and 2 *SD* was 51.8. For total episodic components, 1 *SD* below the mean was 9, and 2 *SD* was 6.4. For MSSG classification, of greatest interest for this study are Quadrant 2, which represents high sequenced content-high episodic structure, and Quadrant 3, which represents low sequenced content-poor episodic structure.

Assessment and Scoring Fidelity

Transcription and Coding

Transcripts collected for this study were elicited and transcribed by multiple AphasiaBank contributors, leading to potential variability in the data collection. Previous studies have examined this concern and found “excellent adherence” to the protocol mentioned above as well as transcript reliability (Richardson & Dalton, 2016). We have rechecked 20% of the videos to ensure proper administration according to the AphasiaBank procedures. For the rest of the locally collected transcripts awaiting submission to AphasiaBank, the transcripts were orthographically transcribed by a research assistant and reviewed by another research assistant for accuracy.

Over the years of developing the MC checklist (Richardson & Dalton, 2016) and the training manual, researchers have uncovered commonly encountered scoring errors. Pronoun referents are a commonly visited rule from the Nicholas and Brookshire (1995) scoring manual to determine the accuracy and completeness, as scorers can interpret utterances differently from other scorers. For example, some raters will score an utterance as accurate/incomplete (AI) and others will score the same utterance as inaccurate/complete (IC). Since AI and IC scores receive the same numerical value (i.e., 2 points), this will not affect the MC composite score, but it does affect point-to-point agreement for coding and is important to consider for fidelity of the results. Another inconsistent scoring error discovered involves the application of the completeness rule (e.g., “Statements containing some of the essential information”), which allows a non-MC utterance to be used if it contains an important story component (e.g., character, pivotal event, etc.). Due to these findings, 20% of the participant transcripts have been rescored for intra-rater reliability.

Scoring and training of the MSSG training dataset was completed by Richardson, Dalton, and Greenslade (Greenslade et al., 2020; Richardson et al., 2021). Graduate students completed training for main concepts, sequencing, and/or story grammar scoring. The training lasted approximately 2-4 weeks depending upon the variable they were assigned and their accuracy/reliability of scoring and coding. Training was completed once the raters reached a minimum of 80% reliability for point-to-point agreement on the training files (i.e., 5 PNBIs and 5 PWAs) provided by the authors of Richardson et al. (2021). Meetings were held to discuss scoring differences to support the learning of the trainee.

Intra-Rater Reliability

Point-to-point agreement intra-rater reliability was calculated for 20% of the clinical study sample (19 PWLAs and 43 PWAAs [includes 19 PWAAs-top]), as these are the transcripts more likely to suffer from decreased reliability. For PWLAs, intra-rater reliability was as follows: 99.8% for MC composite, 99.2% for sequencing, 99.6% for MC+sequencing, 99.7% for essential story grammar components, 100% for total episodic components, and 100% for episodic complexity. For PWAAs, intra-rater reliability was as follows: 99.9% for MC composite, 99.4% for sequencing, 99.6% for MC+sequencing, 99.8% for essential story grammar components, 99.2% for total episodic components, and 100% for episodic complexity. For our secondary analysis, point-to-point agreement intra-rater reliability for PWAAs-top is as follows: 100% for MC composite, 98.9% for sequencing, 99.5% for MC+sequencing, 99.9% for essential story grammar components, 99.3% for total episodic components, and 100% for episodic complexity. Overall, excellent intra-rater reliability was observed in this study, and combining this with the inter-rater reliability information reported in Richardson et al. (2021) (i.e., all MSSG variables above 95.3%), we are confident in the integrity of the data entered into analyses.

Data Analysis

Data Screening

Normality was evaluated using skew and kurtosis, where skew > +/-2 and kurtosis > +/-4 indicated a non-normal distribution (Fabrigar et al., 1999; West et al., 1995). For all groups (and subgroups), all MSSG variables had acceptable skew and kurtosis. For all groups (and subgroups), all MC codes had acceptable skew and kurtosis, except inaccurate/complete (IC) for PNBI and inaccurate/incomplete (II) for PNBI and PWLA. Homogeneity of variance for MSSG variables was visually evaluated using histograms and revealed that PNBIs had negatively skewed distributions and PWLAs and PWAAs MSSG variables generally had positively skewed distributions (exceptions: total episodic components and episodic complexity). For the secondary analysis, PWAAs-top had negatively skewed distributions. Homogeneity of variance for MC codes was also visually evaluated using histograms and revealed that PNBIs generally had positively skewed distributions (exception: Accurate/Complete [AC]) and PWLAs and PWAAs generally had positively skewed distributions (exception: Absent [AB]). For the secondary analysis, PWAAs-top had positively skewed distributions. Our screening revealed that the homogeneity of variance assumption was not met, so appropriate alternative statistical tests were conducted and are detailed below.

Significance Testing

In order to maximize clinical utility of these measures, we report descriptive statistics (mean, median, SD, range, skew, and kurtosis) for all PNBIs, PWLAs, and PWAAs (including the top subgroup). Omnibus tests and planned pairwise group comparisons were conducted to identify differences between the groups under study. Since assumptions of normality were met but assumptions for homogeneity of variance were violated, Welch's ANOVAs were used to evaluate between-group differences. Games-Howell tests were used for pairwise follow-up comparisons. Results for the MSSG variables are first presented, followed by results for each MC code.

CHAPTER 3

RESULTS

MSSG Variables

MC Composite

PWLAs produced lower MC composite scores (M=43.1, SD=16.5) than PNBIs (M=60.8, SD=17.9) but received higher MC composite scores than PWAAs (M=28.8, SD=17.6) and PWAAs-top (M=33.5, SD=16.6). For the primary analysis (PNBI v. PWLA v. PWAA), omnibus ($F(2,160.96)=100.037$, $p<.001$) and all pairwise tests ($p<.001$) were significant. For the secondary analysis (PNBI v. PWLA v. PWAA-top), omnibus ($F(2,133.64)=55.228$, $p<.001$) and all pairwise tests were significant (PNBI v. PWLA and PNBI v. PWAA-top, $p<.001$; PWLA v. PWAA-top, $p<.01$).

Sequencing

PWLAs produced lower sequencing scores (M=46.8, SD=17.4) than PNBIs (M=62.2, SD=17.9) but received higher sequencing scores than PWAAs (M=34.2, SD=19.4) and PWAAs-top (M=38.8, SD=18.4). For the primary analysis, omnibus ($F(2,162.13)=70.286$, $p<.001$) and all pairwise tests ($p<.001$) were significant. For the secondary analysis, omnibus ($F(2,130.16)=37.113$, $p<.001$) and all pairwise tests were significant (PNBI v. PWLA and PNBI v. PWAA-top, $p<.001$; PWLA v. PWAA-top, $p<.05$).

MC Composite + Sequencing

PWLAs produced lower MC composite + sequencing scores (*M*=89.9, *SD*=33.8) than PNBIs (*M*=123, *SD*=35.6) but received higher MC composite + sequencing scores than PWAAs (*M*=63.2, *SD*=37) and PWAAs-top (*M*=72.3, *SD*=35). For the primary analysis, omnibus ($F(2,161.83)=84.111$, $p<.001$) and all pairwise tests ($p<.001$) were significant. For the secondary analysis, omnibus ($F(2,131.83)=45.415$, $p<.001$) and all pairwise tests were significant (PNBI v. PWLA and PNBI v. PWAA-top, $p<.001$; PWLA v. PWAA-top, $p<.05$).

Essential Story Grammar

PWLAs produced lower essential story grammar scores (*M*=16.1, *SD*=5.7) than PNBIs (*M*=21.3, *SD*=6.1) but received higher essential story grammar scores than PWAAs (*M*=11.9, *SD*=6.7) and PWAAs-top (*M*=13.4, *SD*=6.3). For the primary analysis, omnibus ($F(2,164.76)=66.914$, $p<.001$) and all pairwise tests ($p<.001$) were significant. For the secondary analysis, omnibus ($F(2,131.15)=36.387$, $p<.001$) and all pairwise comparisons were tests (PNBI v. PWLA and PNBI v. PWAA-top, $p<.001$; PWLA v. PWAA-top, $p<.05$).

Total Episodic Components

PWLAs produced fewer total episodic components (*M*=9.2, *SD*=2.9) than PNBIs (*M*=11.6, *SD*=2.6) but produced more total episodic components than PWAAs (*M*=6.9, *SD*=3.9) and PWAAs-top (*M*=7.9, *SD*=3.6). For the primary analysis, omnibus

(F(2,163.05)=65.704, p<.001) and all pairwise tests (p<.001) were significant. For the secondary analysis, omnibus (F(2,131.15)=36.387, p<.001), PNBI v. PWLA, and PNBI v. PWAA-top pairwise tests (p<.001) were significant. The PWLA v. PWAA-top pairwise test (p<0.071) did not reach statistical significance.

Episodic Complexity

PWLAs produced lower episodic complexity scores (M=3.2, SD=1.3) compared to PNBIs (M=4.1, SD=1.1) but higher episodic complexity scores than PWAAs (M=2.3, SD=1.6) and PWAAs-top (M=2.7, SD=1.5). For the primary analysis, omnibus (F(2,158.14)=54.741, p<.001) and all pairwise comparisons (p<.001) were significant. For the secondary analysis, omnibus (F(2,117.38)=24.833, p<.001), PNBI v. PWLA, and PNBI v. PWAA-top pairwise tests (p<.001) were significant. The PWLA v. PWAA-top pairwise test (p<0.084) did not reach statistical significance.

Table 2. Descriptive statistics for each MSSG variable for Cinderella

Statistic	**Participant groups**			
	PNBI (N = 112)	**PWLA (N = 60)**	**PWAA (N = 139)**	**PWAA-top (N = 63)**
MC Composite				
Mean	60.8	43.1	28.8	33.5
SD	17.9	16.5	17.6	16.6
Median	64	41	28	33
Range	8 – 91	7 – 86	2 – 75	2 – 66
Skew	-0.720	0.216	0.375	-0.106

Table 2. [cont.]				
Kurtosis	0.380	-0.024	-0.681	-0.736
Sequencing				
Mean	62.2	46.8	34.2	38.8
SD	17.9	17.4	19.4	18.4
Median	65	45	34	39
Range	9 – 92	9 – 90	3 – 81	3 – 76
Skew	-0.797	0.112	0.199	-0.222
Kurtosis	0.446	-0.209	-0.924	-0.721
MC Composite + Sequencing				
Mean	123	89.9	63.2	72.3
SD	35.6	33.8	37.0	35.0
Median	128	87	64	75
Range	17 - 183	16 – 176	5 – 156	5 – 142
Skew	-0.770	0.151	0.281	-0.163
Kurtosis	0.430	-0.123	-0.824	-0.726
Essential Story Grammar Components				
Mean	21.3	16.1	11.9	13.4
SD	6.1	5.7	6.7	6.3
Median	22	16	12	14
Range	3 – 33	3 – 30	1 – 30	1 – 25
Skew	-0.679	0.072	0.238	-0.228
Kurtosis	0.398	-0.182	-0.790	-0.764
Total Episodic Components				
Mean	11.6	9.2	6.9	7.9
SD	2.6	2.9	3.9	3.6
Median	12	10	8	9
Range	2 – 15	1 – 14	0 – 14	0 – 13

Table 2. [cont.]				
Skew	-1.287	-0.756	-0.264	-0.583
Kurtosis	1.842	0.306	-1.214	-0.874
Episodic Complexity				
Mean	4.1	3.2	2.3	2.7
SD	1.1	1.3	1.6	1.5
Median	4	3.5	2	3
Range	0 – 5	0 – 5	0 – 5	0 – 5
Skew	-1.459	-0.663	-0.077	-0.327
Kurtosis	2.204	-0.249	-1.257	-1.014

Table 3. Welch analysis of variance (ANOVA) results for between group comparisons for MSSG variables – primary analysis

		Post hoc tests – Games-Howell		
		Participant Group Comparisons		
MSSG variable	**F Value**	**PNBI vs. PWLA**	**PNBI vs. PWAA**	**PWLA vs. PWAA**
MC Composite	$F(2,160.96)=100.037$***	***	***	***
Sequencing	$F(2,162.13)=70.286$***	***	***	***
MC Composite + Sequencing	$F(2,161.83)=84.111$***	***	***	***
Essential Story Grammar Components	$F(2,164.76)=66.914$***	***	***	***
Total Episodic Components	$F(2,163.05)=65.704$***	***	***	***
Episodic Complexity	$F(2,158.14)=54.741$***	***	***	***

Note: *$p < .05$; **$p < .01$; ***$p < .001$

Table 4. Welch analysis of variance (ANOVA) results for between group comparisons for MSSG variables – secondary analysis

		Post hoc tests – Games-Howell		
		Participant Group Comparisons		
MSSG variable	**F Value**	**PNBI vs. PWLA**	**PNBI vs. PWAA-top**	**PWLA vs. PWAA-top**
MC Composite	$F(2,133.64)=55.228$***	***	***	**
Sequencing	$F(2,130.16)=37.113$***	***	***	*
MC Composite + Sequencing	$F(2,131.83)=45.415$***	***	***	*
Essential Story Grammar Components	$F(2,131.15)=36.387$***	***	***	*
Total Episodic Components	$F(2,119.34)=31.913$***	***	***	0.071
Episodic Complexity	$F(2,117.38)=24.833$***	***	***	0.084

Note: *$p < .05$; **$p < .01$; ***$p < .001$

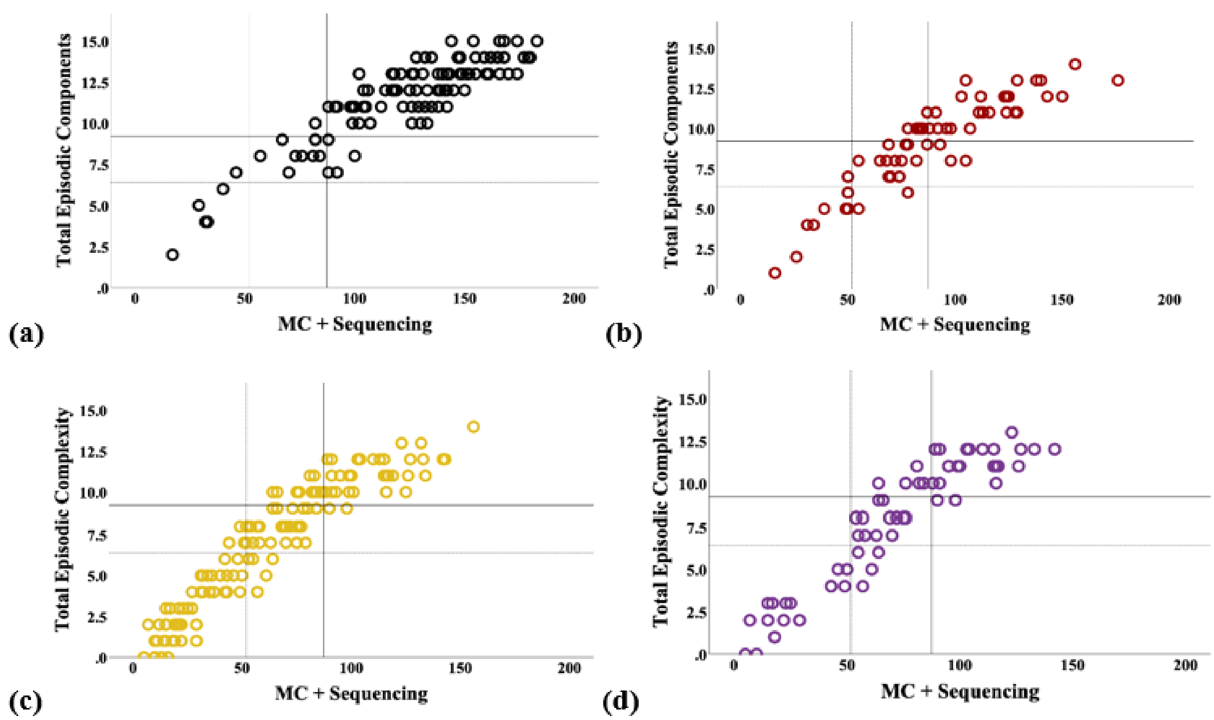
MSSG Classification

Each quadrant was defined by cut off points at 1*SD* (i.e., solid lines) and 2 *SD* (i.e., dotted lines) below the control mean for MC+sequencing and total episodic components (See Table 2 for *M* and *SD* values for each variable). For MC+sequencing, 1 *SD* below the mean was 87.4, and 2 *SD* was 51.8. For total episodic components, 1 *SD* below the mean was 9, and 2 *SD* was 6.4. Using 1-*SD* values, the distribution of scores for Quadrant 2 (high scores) was as follows: *PNBI*: 83%; *PWLAs*: 41.6%; *PWAAs-top*: 38%; and *PWAAs*: 27.3%. Using

2-*SD* values, the distribution of scores for Quadrant 2 (high scores) was as follows: *PNBI*: 94.6%; *PWLAs*: 83.3%; *PWAAs-top*: 68.3%; and *PWAAs*: 55.4%. Using 1-*SD* values, the distribution of scores for Quadrant 3 (low sequenced content-poor episodic structure) was as follows: *PNBI*: 10.7%; *PWLAs*: 33.3%; *PWAAs-top*: 49.2%; and *PWAAs*: 58.3%. Using 2-*SD* values, the distribution of scores for Quadrant 3 (low scores) was as follows: *PNBI*: 4.5%; *PWLAs*: 13.3%; *PWAAs-top*: 25.4%; and *PWAAs*: 38.1%.

Figure 1. MSSG Classification

Figure 1. MSSG Classification



Note: MSSG Classification for Cinderella narratives, plotted as a function of content and sequencing (MC+Sequencing; *x* axis) and overall episodic structure (total episodic components; *y* axis). Quadrants were defined by cutoff points at 1SD (solid line) and 2SD (dotted lines) below the mean for both variables. Quadrant 2 contains the highest scores for both dimensions, Quadrant 3 contains the lowest scores for both dimensions. Quadrant 1 contains low sequenced content and high overall episodic structure. Quadrant 4 contains high sequenced content and low overall episodic structure. (**a**) controls, or persons not brain injured (PNBIs) (**black**); (**b**) persons with latent aphasia (PWLAs) (**red**); (**c**) persons with anomic aphasia (PWAAs) (**yellow**); (d) persons with aphasia-top (PWAAs-top) (**purple**).

MC Codes

Accurate Complete

PWLAs produced fewer accurate/complete MC codes (*M*=11.2, *SD*=5.3) than PNBIs (*M*=18.6, *SD*=6.1) but produced more accurate/complete MC codes than PWAAs (*M*=5.9, *SD*=5.1) and PWAAs-top (*M*=7.5, *SD*=4.9). For the primary analysis, omnibus ($F(2,156.38)=155.757$, $p<.001$) and all pairwise tests ($p<.001$) were significant. For the secondary analysis, omnibus ($F(2,137.37)=90.105$, $p<.001$) and all pairwise tests ($p<.001$) were significant.

Accurate Incomplete

PWLAs produced more accurate/incomplete MC codes (*M*=3.2, *SD*=2.0) than PNBIs (*M*=1.3, *SD*=1.4) but produced fewer accurate/incomplete MC codes than PWAAs (*M*=3.8, *SD*=2.8) and PWAAs-top (*M*=3.4, *SD*=2.6). For the primary analysis, omnibus ($F(2,151.55)=52.962$, $p<.001$), PNBI v. PWLA, and PNBI v. PWAA pairwise tests ($p<.001$) were significant. The PWLA v. PWAA pairwise comparison ($p<0.203$) did not reach statistical significance. For the secondary analysis, omnibus ($F(2,107.97)=33.519$, $p<.001$), PNBI v. PWLA, and PNBI v. PWAA pairwise tests ($p<.001$) were significant. The PWLA v. PWAA-top pairwise comparison ($p<0.898$) did not reach statistical significance.

Inaccurate Complete

PWLAs produced similar amounts of inaccurate/complete MC codes ($M=1.3$, $SD=1.3$) as PNBIs ($M=1.1$, $SD=1.3$), PWAAs ($M=1.6$, $SD=1.6$), and PWAAs-top ($M=1.8$, $SD=1.6$). For the primary analysis, omnibus ($F(2,164.74)=3.533$, $p<.05$) and PNBI v. PWAA pairwise tests ($p<.05$) were significant. PNBI v. PWLA ($p<0.561$) and PWLA v. PWAA ($p<0.417$) pairwise comparisons did not reach statistical significance. For the secondary analysis, omnibus ($F(2,125.7)=5.233$, $\mathrm{p}<.01$) and PNBI v. PWAA-top pairwise tests($p<.01$) were significant. The PNBI v. PWLA ($p<0.561$) and PWLA v. PWAA-top ($p<0.094$) pairwise comparisons did not reach statistical significance.

Inaccurate Incomplete

PWLAs produced similar amounts of inaccurate/incomplete MC codes ($M=0.2$, $SD=0.5$) as PNBIs ($M=0.1$, $SD=0.3$) but produced fewer inaccurate/incomplete MC codes than PWAAs ($M=0.5$, $SD=0.8$) and PWAAs-top ($M=0.5$, $SD=0.7$). For the primary analysis, omnibus ($F(2,144.62)=18.354$, $p<.001$), PNBI v. PWAA ($p<.001$), and PWLA v. PWAA ($p<.01$) pairwise tests were significant. The PNBI v. PWLA pairwise comparison ($p<0.179$) did not reach statistical significance. For the secondary analysis, omnibus ($F(2,102.38)=8.021$, $p<.001$) and PNBI v. PWAA-top pairwise tests ($p<.001$) were significant. The PNBI v. PWLA ($p<0.179$) and PWLA v. PWAA-top ($p<0.116$) pairwise comparisons did not reach statistical significance.

Absent

PWLAs produced more absent MC codes (*M*=18.0, *SD*=5.8) than PNBIs (*M*=12.9, *SD*=6.0). When compared to PWAAs (*M*=22.2, *SD*=6.6) and PWAAs-top (*M*=20.8, *SD*=6.2), PWLAs produced fewer absent MC codes. For the primary analysis, omnibus (F(2,163.03)=68.156, p<.001) and all pairwise tests (p<.001) were significant. For the secondary analysis, omnibus (F(2,130.29)=37.235, p<.001) and all pairwise tests were significant (PNBI v. PWLA and PNBI v. PWAA-top, p<.001; PWLA v. PWAA-top, p<.05).

Table 5. Descriptive statistics for each MC code for Cinderella

Statistic	**Participant groups**			
	PNBI (N = 112)	**PWLA (N = 60)**	**PWAA (N = 139)**	**PWAA-top (N = 63)**
MC Code – Accurate/Complete (AC)				
Mean	18.6	11.2	5.9	7.5
SD	6.1	5.3	5.1	4.9
Median	20	11	5	7
Range	2 – 30	1 – 26	0 – 23	0 – 18
Skew	-0.611	0.425	0.962	0.373
Kurtosis	0.057	0.195	0.443	-0.753
MC Code – Accurate/Incomplete (AI)				
Mean	1.3	3.2	3.8	3.4
SD	1.4	2.0	2.8	2.6
Median	1	3	3	3
Range	0 – 6	0 – 8	0 – 13	0 – 13
Skew	1.235	0.357	0.999	1.284

Table 5. [cont.]				
Kurtosis	1.137	-0.474	0.889	2.828
MC Code – Inaccurate/Complete (IC)				
Mean	1.1	1.3	1.6	1.8
SD	1.3	1.3	1.6	1.6
Median	1	1	1	2
Range	0 – 7	0 – 5	0 – 7	0 – 6
Skew	1.883	1.117	1.056	0.764
Kurtosis	5.080	1.106	0.788	0.069
MC Code – Inaccurate/Incomplete (II)				
Mean	0.1	0.2	0.5	0.5
SD	0.3	0.5	0.8	0.7
Median	0	0	0	0
Range	0 – 2	0 – 2	0 – 4	0 – 3
Skew	3.488	2.266	1.569	1.514
Kurtosis	12.621	4.337	2.676	1.820
MC Code – Absent (AB)				
Mean	12.9	18.0	22.2	20.8
SD	6.0	5.8	6.6	6.2
Median	12	18	22	20
Range	2 – 31	4 – 31	5 – 33	9 – 33
Skew	0.730	-0.057	-0.183	0.247
Kurtosis	0.466	-0.261	-0.842	-0.753

Table 6. Welch analysis of variance (ANOVA) results for between group comparisons for MC codes – primary analysis

		Post hoc tests – Games-Howell		
MC code	**F Value**	**Participant Group Comparisons**		
		PNBI vs. PWLA	**PNBI vs. PWAA**	**PWLA vs. PWAA**
Accurate/Complete (AC)	$F(2,156.38)=155.757$***	***	***	***
Accurate/Incomplete (AI)	$F(2,151.55)=52.962$***	***	***	0.203
Inaccurate/Complete (IC)	$F(2,164.74)=3.533$*	0.561	*	0.417
Inaccurate/Incomplete (II)	$F(2,144.62)=18.354$***	0.179	***	**
Absent (AB)	$F(2,163.03)=68.156$***	***	***	***

Note: *$p < .05$; **$p < .01$; ***$p < .001$

Table 7. Welch analysis of variance (ANOVA) results for between group comparisons for MC codes – secondary analysis

		Post hoc tests – Games-Howell		
		Participant Group Comparisons		
MC code	**F Value**	**PNBI vs. PWLA**	**PNBI vs. PWAA-top**	**PWLA vs. PWAA-top**
Accurate/Complete (AC)	$F(2,137.37)=90.105$***	***	***	***
Accurate/Incomplete (AI)	$F(2,107.97)=33.519$***	***	***	0.898
Inaccurate/Complete (IC)	$F(2,125.7)=5.233$**	0.561	**	0.094
Inaccurate/Incomplete (II)	$F(2,102.38)=8.021$***	0.179	***	0.116
Absent (AB)	$F(2,130.29)=37.235$***	***	***	*

Note: *$p < .05$; **$p < .01$; ***$p < .001$

CHAPTER 4

DISCUSSION

This study replicated and extended previous MCA and MSSG research in the largest sample to date of persons with latent and anomic aphasia, including the largest sample of top-scoring persons with anomic aphasia. It is the first study to provide a detailed comparison of PWLAs to PNBIs, PWAAs, and PWAAs-top utilizing MSSG to examine macrostructural discourse characteristics. PWLAs had lower MSSG variable scores than PNBIs and higher MSSG variable scores than PWAAs and PWAAs-top. Our analyses revealed significant differences between all groups in the primary and secondary analyses for all MSSG variables, with the exception of PWLA v. PWAAs-top for total episodic components and episodic complexity. Taken together, these results indicate that PWLAs are a unique group separate from its two closest neighbors (i.e., PNBIs and PWAAs-top), with some overlap with PWAAs-top. Our additional analyses for MC codes (e.g., AC, AI, etc.) revealed more of an overlap between the groups under study, where the only codes that were consistently significantly different between PNBIs, PWLAs, PWAAs, and PWAAs-top were accurate/complete (AC) and absent (AB) productions.

Main Concept Composite - Primary Analysis

Previous research studies have examined performance differences between PNBIs, PWLAs, and PWAAs for MC composite scores. Fromm et al. (2017) found significant differences (*p*<.001) in performance between PNBIs and PWAAs (N=87) and between PNBIs v. PWLAs (N=27) for the MC composite score. Dalton and Richardson (2019) revealed significant differences (*p*<.001) in performance between PNBIs and PWAAs

(N=86) but found no significant differences between PNBIs and PWLAs (N=26). The reason for these inconsistent findings, despite the two studies sharing participant transcripts, may be related to different control sample sizes and/or different statistical tests (i.e., Fromm et al. [2017] used log-transformed data and Tukey HSD, Dalton and Richardson [2019] used nontransformed data and median tests). Though Dalton and Richardson (2019) did not directly compare PWLAs and PWAAs, their descriptive statistics showed that PWLAs produced higher MC composite scores (M=43.5, SD=17.7) than PWAAs (M=26.2, SD=16.6). Richardson et al. (2021) found significant differences (p<.001) between PNBIs and PWAAs (N=122) and PNBIs and PWLAs (N=54). As their study also did not provide a direct comparison, we examined their descriptive statistics which showed that PWLAs produced higher MC composite scores (M=43.5, SD=16.6) than PWAAs (M=27.7, SD=17.5). The primary analysis conducted here revealed significant differences in performance (p<.001) between all groups - PNBIs, PWLAs, and PWAAs - consistent with Fromm et al. (2017) and Richardson et al. (2021) findings for the three groups, and consistent with Dalton and Richardson (2019) findings for PNBIs v. PWAAs, but not their findings for PNBIs v. PWLAs.

Main Concept Composite - Secondary Analysis

Fromm et al. (2017) introduced the analysis examining differences between PWLAs (N = 27) and PWAAs-top (N = 36). They found that PWLAs and PWAAs-top did not significantly differ in performance for the MC composite score. This is inconsistent with our findings, as we found significant differences in performance between PWLAs v. PWAAs-top. This inconsistency with Fromm et al. (2017) could be due to 1) our study having a larger sample size than Fromm et al. (2017) (33 more PWLAs, 27 more PWAAs-top) that may more accurately represent these populations, and/or 2) different statistical tests (i.e., Fromm et al. [2017] used log-transformed data and Tukey HSD, while we conducted Welch's ANOVAs and Games-Howell tests).

Main Concept Codes - Primary Analysis

Dalton & Richardson (2019) reported on each of the 5 MC codes and found significant differences in performance between PNBI v. PWLA (N=26) and PNBI v. PWAA (N=86) across all MC codes, except for inaccurate/complete (IC) codes. Our study revealed significant differences (p<.001) between PNBIs, PWLAs (N=60), and PWAAs (N=139) for all MC codes, except between PNBIs v. PWLAs for inaccurate/complete (IC) and inaccurate/incomplete (II) codes. This is partially consistent with Dalton & Richardson's (2019) results as our study revealed no significant differences between PNBIs v. PWLAs for inaccurate/complete (IC) and inaccurate/incomplete (II) and found significant differences in performance between PNBIs and PWAAs for all MC codes.

Main Concept Codes - Secondary Analysis

Our secondary analysis revealed significant differences in performance between 1) PNBIs v. PWAAs-top for all MC codes, 2) PNBIs v. PWLAs for accurate/complete, accurate/incomplete, and absent codes, and 3) PWLAs v. PWAAs-top for accurate/complete and absent codes. There are no previously conducted studies for comparison, but based on our results, it seems that accurate/complete and absent MC codes appear to be the most promising codes to distinguish PWLAs from nearest neighbors PNBIs and PWAAs-top. This is somewhat inconsistent with previous research and interpretation by Nicholas and Brookshire (1995) and Dalton and Richardson (2019). Their studies suggested that accurate/incomplete, inaccurate/complete, and inaccurate/incomplete codes would be the most sensitive for detection of mild aphasic deficits, as PNBIs seldom produce these types of errors. While it is true that PNBIs and PWAAs-top differed significantly for the inaccurate and/or incomplete MC codes, the overlapping distributions for PWLAs and these groups led to our finding that PWLAs did not significantly differ from either group (Table 7). Though our study found that these error codes may not be as promising for identifying PWLAs from PNBIs or PWAAs-top, they still provide great information that can be used to direct the development of therapy goals.

MSSG Variables - Primary Analysis

Richardson et al. (2021) reported on each MSSG variable between PNBIs, PWLAs (N=54), and PWAAs (N=122). They found significant differences (p<.001) in performance between PNBIs v. PWLAs and between PNBIs v. PWAAs for all MSSG variables. Though Richardson et al. (2021) did not provide a direct comparison between PWLAs and PWAAs, their descriptive statistics showed that PWLAs produced higher MSSG variable scores than PWAAs: 1) MC composite scores - PWLAs (*M*=43.5, *SD*=16.6) and PWAAs (*M*=27.7, *SD*=17.5); 2) sequencing - PWLAs (*M*=47.1, *SD*=17.5) and PWAAs (*M*=32.8, *SD*=19.3); 3) MC composite + sequencing - PWLAs (*M*=90.6, *SD*=34) and PWAAs (*M*=60.5, *SD*=36.7); 4) essential story grammar components - PWLAs (*M*=16.2, *SD*=5.8) and PWAAs (*M*=11.4, *SD*=6.7); 5) total episodic components - PWLAs (*M*=9.2, *SD*=2.8) and PWAAs (*M*=6.7, *SD*=3.9); and 6) episodic complexity - PWLAs (*M*=3.2, *SD*=1.3) and PWAAs (*M*=2.2, *SD*=1.6). Our primary analysis also revealed significant differences (*p*<.001) between PNBIs v. PWLAs and between PNBIs v. PWAAs for all MSSG variables (i.e., MC composite, sequencing, MC composite + sequencing, essential story grammar components, total episodic components, and episodic complexity), consistent with Richardson et al. (2021) findings.

MSSG Variables - Secondary Analysis

PNBIs and PWLAs significantly differed for all MSSG variables as did PNBIs and PWAAs-top. Our secondary analysis revealed significant differences in performance between PWLAs and PWAAs-top for MC composite (*p*<.01), sequencing (*p*<.05), MC composite + sequencing (*p*<.05) and essential story grammar components (*p*<.05). There were no significant differences for total episodic components and episodic complexity, though descriptive statistics showed that PWLAs received higher total episodic components and episodic complexity scores than PWAAs-top (1) total episodic components - PWLAs (*M*=9.2, *SD*=2.9) and PWAAs-top (*M*=7.9, *SD*=3.6) and 2) episodic complexity - PWLAs (*M*=3.2, *SD*=1.3) and PWAAs-top (*M*=2.7, *SD*=1.5). Based on these results, all MSSG variables seemed to be sensitive enough to detect latent aphasia, while MC composite, sequencing, MC composite + sequencing, and essential story grammar components appeared to be the most informative MSSG variables for distinguishing PWLAs from PWAAs-top. These results indicate that MSSG analysis is sensitive enough to detect PWLAs higher-level discourse deficits, and provided further evidence that PWLAs should not be labeled as "recovered".

MSSG Classification

The use of MSSG classification allowed for visual representation of the relationship between each populations' ability to tell accurate, complete, and logically sequenced stories and their ability to maintain overall episodic structure (Greenslade et al., 2020). The results of the MSSG classification illustrated differences in performance between PNBIs, PWLAs, PWAAs, and PWAAs-top and revealed the extent and types of deficits found with each population. This analysis also highlighted a concern regarding standard deviation cutoff criteria, consistent with caution urged by Richardson et al. (2021), where moving from 1-*SD* to 2-*SD* cutoffs resulted in significantly different percentages of individuals in each population being categorized in each Quadrant (i.e., Quadrant 2; high sequenced content - high episodic structure and Quadrant 3; low sequenced content - poor episodic structure). For example, our results show that moving from the 1-*SD* to 2-*SD* cutoff for Quadrant 2 (i.e., high scores) resulted in an increase of 11.6% of PNBIs, 41.7% of PWLAs, 30.3% of PWAAs-top, and 28.1% of PWAAs being considered as having high sequenced content and episodic structure. A similar situation was observed with Quadrant 3 (i.e., low scores) where the change from the 1-*SD* to 2-*SD* cutoff resulted in a decrease of 6.2% of PNBIs, 20% of PWLAs, 23.8% of PWAAs-top, and 20.2% of PWAAs being considered as having low sequenced content and poor episodic structure. Our findings demonstrate the importance of avoiding setting arbitrary cutoffs on standardized diagnostic assessments until further research is conducted to improve our understanding of the deficits associated with these clinical populations. Certainly, use of a 2-*SD* cutoff would result in many PWLAs and PWAAs not being identified as having these narrative content and organization deficits that are illustrated by the classification visual.

Limitations

While the results from the current study and previous studies utilizing MSSG analyses demonstrate its sensitivity for detecting macrostructural discourse deficits in clinical populations, including PWLAs, there are several limitations worth mentioning. Although our study includes the largest and most diverse sample size to date, the majority of our participants were Caucasian, resulting in our data having race/ethnicity bias. Another limitation involves both a strength and a weakness of checklist-based scoring systems, whereby efficient and streamlined approaches may lead to ignored verbal and non-verbal participant productions. For example, when participant utterances do not directly align with a MC target, they are excluded from the entire MSSG analyses, even though these utterances could still provide useful information about the participants discourse abilities. A similar issue occurs when looking at a participant's non-verbal (i.e., gestures) productions, as they can still provide information that directly relates to a target MC but are ignored because the checklist relies on verbal productions only. Another limitation to note is that MSSG analyses utilizes a semi-spontaneous discourse task rather than conversational discourse, which allows for the data to be normed and replicated, but may not fully represent how that individual's functional, everyday communication is impacted. Finally, the results of this study may only be useful for researchers and clinicians utilizing the AphasiaBank protocol for Cinderella, as Cinderella is the only normed stimuli for MSSG analyses and strict adherence to the AphasiaBank protocol is essential for reliable outcome measures.

Future Directions

Discourse is now listed as a primary outcome measure for aphasia treatment research (Brady et al., 2016) and further work and research is needed to improve discourse measures and ensure that they are stable and reliable for clinical application. Although we had a large sample size overall (N=311), and each population examined included over the minimum 50 individuals (PNBI - N=112, PWLA - N=60, PWAA - N=139, and PWAA-top - N=63) to be considered adequate to represent each population, a future study that replicates this current study in an independent sample is essential to ensure accurate representation of PWLAs and other clinical populations. In addition to replicating this current study with an independent sample, designing a study to collect data on test-retest reliability would establish internal validity of MSSG analysis. In planning these future MSSG studies, a few issues should be taken into consideration. First, as we observed some difficulties with procedures for coding and scoring, additional training and/or the use of automated techniques that prevent data coding/entry and scoring errors is needed, particularly for sequencing. Second, creating further normed stimuli to analyze other types of discourse (i.e., description, recount, procedural, etc.) utilizing MSSG analyses may provide beneficial information regarding discourse deficits associated with PWLAs and other aphasia subtypes.

There is additional and potentially useful information available following MSSG analyses that has yet to be described or analyzed. MSSG analyses involves coding of story grammar elements but currently does not report the quantity of each element (i.e., setting, initiating events, attempts, direct consequences, etc.). Reporting on these individual story grammar variables may provide further characterization of these clinical populations and also help direct the development of treatment goals.

A final future direction involves improving upon the sensitivity of MSSG analyses for detecting the mild deficits experienced by PWLAs. Previous research conducted by DeDe and Salis (2020) found that temporal measures, such as percent of formulation time for utterances and speech rate, were sensitive to identifying individuals with latent aphasia. We suggest combining MSSG analyses with the afore mentioned temporal measures as it would provide a more detailed characterization of PWLAs and only require timing of the discourse sample and a simple calculation.

Summary

In summary, PWLAs did not differ significantly from: 1) PNBIs for inaccurate/complete and inaccurate/incomplete MC codes; 2) PWAAs for accurate/incomplete and inaccurate/complete MC codes; or 3) PWAAs-top for total episodic components and episodic complexity MSSG variables and accurate/incomplete, inaccurate/complete, and inaccurate/incomplete MC codes. PWLAs did differ significantly from: 1) PNBIs for all MSSG variables and accurate/complete, accurate/incomplete, and absent MC codes; 2) PWAAs for all MSSG variables and accurate/complete, inaccurate/incomplete, and absent MC codes; and 3) PWAAs-tops for MC composite, sequencing, MC composite + sequencing and essential story grammar components MSSG variables and accurate/complete, inaccurate/incomplete, and absent MC codes. The variables most useful for detection are those above that differ from PNBIs. The variables that help to identify PWLAs as a clinical population are those listed above where they differ from PNBIs and/or do not differ from PWAAs or PWAAs-top. The variables that help to further distinguish PWLAs as their own distinct subgroup are the variables where they significantly differed from all comparison groups - MC composite, sequencing, MC composite + sequencing, and essential story grammar components MSSG variables and accurate/complete and absent MC codes. These results provide confirmation of previous research findings and define macrostructural discourse differences for PWLAs that justify therapy services and highlight potential treatment targets. As persons with milder deficits have been shown to have greater response to treatment compared to persons with more severe deficits (e.g., Lambon Ralph et al., 2010; Quique, Evans, & Dickey, 2019; but see Fridriksson, Richardson, Fillmore, & Cai, 2012), development and clinical usage of new measures that help to identify deficits and inform treatment may help to quickly and effectively reduce disability and limitations experienced by this clinical group.

REFERENCES

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

Armes, E., Richardson, J., Arenas, R., & Nelson, L. (2020). The relationship between narrative informativeness and psychosocial outcomes in chronic stroke-induced aphasia. Masters Level Thesis, University of New Mexico, Albuquerque, NM.

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

Beeke, S., Wilkinson, R., & Maxim, J. (2003). Exploring aphasic grammar 2: Do language testing and conversation tell a similar story?. *Clinical Linguistics & Phonetics*, 17:2, 109-134

Boller, F., & Vignolo, L. A. (1966). Latent sensory aphasia in hemisphere-damaged patients: An experimental study with the token test. *Brain*, 89(4), 815-830. https://doi.org/10.1093/brain/89.4.815

Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews*, 6, CD000425.https://doi.org/10.1002/14651858.CD000425.pub4B

Cavanaugh, R., & Haley, K. L. (2020). Subjective communication difficulties in very mild aphasia. *American Journal of Speech-Language Pathology*, 29(1S), 437-448

Coelho, C., Le, K., Mozeiko, J., Hamilton, M., Tyler, E., Krueger, F., & Grafman, J. (2013. Characterizing discourse deficits following penetrating head injury: A preliminary model. *American Journal of Speech-Language Pathology*.

Coelho, C.A., Liles, B.Z., Duffy, R.J., Clarkson, J.V., & Elia, D. (1994). Longitudinal assessment of narrative discourse in a mildly aphasic adult. *Clinical Aphasiology, 22*, 145-155.

Cruice et al. (2020). UK speech and language therapists' views and reported practice of discourse analysis in aphasia rehabilitation. *International Journal of Language & Communication Disorder*, 55(3), 417-442.

Dalton, S.G. & Richardson, J.D. (2015). Core-Lexicon and Main-Concept Production During Picture-Sequence Description in Adults Without Brain Damage and Adults With Aphasia. *American Journal of Speech-Language Pathology,* 24, 923-938. https://pubmed.ncbi.nlm.nih.gov/26140462/

Dalton, S.G. & 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

Dalton, S. G. H., Hubbard, H. I., & Richardson, J. D. (2020). Moving toward non-transcription based discourse analysis in stable and progressive aphasia. In *Seminars in Speech and Language, 41*(1), 32.

DeDe, G., & Salis, C. (2020). Temporal and episodic analyses of the story of cinderella in latent aphasia. American Journal of Speech-Language Pathology, 29(1S), 449-462.

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.

Fabrigar, L. R., Wegener, D. T. MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272.

Fridriksson, J., Richardson, J. D., Fillmore, P., & Cai, B. (2012). Left hemisphere plasticity and aphasia recovery. *Neuroimage*, *60*(2), 854-863.

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.

Galetto, V., Andreetta, S., Zettin, M., & Marini, A. (2013). Patterns of impairment of narrative language in mild traumatic brain injury. *Journal of Neurolinguistics*, 26, 649-661.

Greenslade, K. J., Stuart, J. E., Richardson, J. D., Dalton, S. G., & Ramage, A. E. (2020). Macrostructural analyses of Cinderella narratives in a large nonclinical sample. *American Journal of Speech-Language Pathology*, *29*(4), 1923-1936.

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

Kertesz, A. (2007). *Wester Aphasia Battery* - R. New York, NY: Grune & Stratton.

Kong, A. P. H. (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.

Kong, A. P. H., Whiteside, J., & Bargmann, P. (2016). The main concept analysis: Validation and sensitivity in differentiating discourse produced by unimpaired English speakers from individuals with aphasia and dementia of Alzheimer type. *Logopedics Phoniatrics Vocology*, 41(3), 129-141

Lambon Ralph, M. A., Snell, C., Fillingham, J. K., Conroy, P., & Sage, K. (2010). Predicting the outcome of anomia therapy for people with aphasia post CVA: both language and cognitive status are key predictors. *Neuropsychological Rehabilitation, 20*(2), 289-305.

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.

Le, 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.

Marini, A., Zettin, M., Bencich, E., Marina Bosco, F., & Galetto, V. (2017). Severity effects on discourse production after TBI. *Journal of Neurolinguistics*, 44, 91-106

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 and Hearing Research, 36*, 338-350.

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, Language, and Hearing Research, 38*(1), 145-156.

Olsen, T. S., Bruhn, P., & Öberg, R. G. E. (1986). Cortical hypoperfusion as a possible cause of ‘subcortical aphasia’. *Brain, 109*(3), 393-410.

Pichot, P. (1955). Language disturbances in cerebral disease: Concept of latent aphasia

Power, E., Weir, S., Richardson, J., Fromm, D., Forbes, M., MacWhinney, B., & Togher, L. (2020). Patterns of narrative discourse in early recovery following severe Traumatic Brain Injury. *Brain injury, 34*(1), 98-109.

Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2017). Reviewing the quality of discourse information measures in aphasia. *International Journal of Language & Communication Disorders*, *52*(6), 689-732.

Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2018). Psychometric properties of discourse measures in aphasia: Acceptability, reliability, and validity. *International Journal of Language & Communication Disorders*, *53*(6), 1078-1093.

Quique, Y. M., Evans, W. S., & Dickey, M. W. (2019). Acquisition and generalization responses in aphasia naming treatment: A meta-analysis of semantic feature analysis outcomes. *American journal of speech-language pathology*, *28*(1S), 230-246.

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.

Richardson, J. D., & Dalton, S. G. (2019). Main concepts for two picture description tasks: An addition to Richardson and Dalton, 2016. *Aphasiology,* DOI: https://doi.org/10.1080/02687038.2018.1561417

Richardson, J. D., Hudspeth Dalton, S. G., Fromm, D., Forbes, M., Holland, A., & MacWhinney, B. (2018). The relationship between confrontation naming and story gist production in aphasia. *American Journal of Speech-Language Pathology*, *27*(1S), 406-422.

Richardson, J. D., Dalton, S. G., Greenslade, K. J., Jacks, A., Haley, K., & Adams, J. (2021). Main concept, sequencing, and story grammar (MSSG) analyses of Cinderella narratives in a large sample of persons with aphasia. *Brain Sci,* 11(1)

Silkes, J. P., Zimmerman, R. M., Greenspan, W., Reinert, L., Kendall, D., & Martin, N. (2020). Identifying Verbal Short-Term Memory and Working Memory Impairments in Individuals with Latent Aphasia. *American Journal of Speech-Language Pathology*, 1-16.

Stout, C. E., Yorkston, K. M., & Pimentel, J. I. (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.

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

Ulatowska, H. K., Freedman-Stern, R., Dovel, A. W., & Macaluso-Haynes, S. (1983). Production of narrative discourse in aphasia. *Brain and Language,* 19, 317-334.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications (*pp. 56-75). Thousand Oaks, CA: Sage Publications, Inc.

Whitworth, A. (2010). Using narrative as a bridge: Linking language processing models with real-life communication. *Seminars in speech and language,* (Vol. 31, No. 01, pp. 064-075).