Original Article

http://e-cacd.org/ eISSN: 2508-5948

Clinical Archives of Communication Disorders / Vol. 6, No. 3:172-186 / December 2021

Open Access

https://doi.org/10.21849/cacd.2021.00626

Proposing a quick, objective, comprehensive analysis of impaired spoken discourse in post-stroke aphasia

Anthony Pak-Hin Kong[1,2]*, Adam Reres[2,3]

[1]Academic Unit of Human Communication, Development, and Information Sciences (CDIS), The University of Hong Kong, Hong Kong; [2]School of Communication Sciences and Disorders, University of Central Florida, Orlando, United States; [3]School of Global Health Management and Informatics, University of Central Florida, Orlando, United States

Purpose: Aphasia is an acquired language disorder, most commonly caused by a stroke, that adversely affects one's ability to understand, speak, read, and write. Diagnosis of aphasia is typically done through administration of standardized aphasia batteries, many of which lack a detailed and adequate evaluation of spoken discourse. The aim of this paper was to propose a quick, objective, and comprehensive analytic system that addressed the micro- and macro-linguistic aspects of oral narratives in persons with aphasia (PWA).

Methods: Using a subgroup of unimpaired native speakers of English from the public corpus of AphasiaBank, task-specific normative data of three narrative tasks were constructed. Scoring criteria on the production of events, use of corresponding lexical items in these events, and the order of sequence of presenting the events were also developed. Twelve PWA were recruited and their spoken discourse was quantified using this newly proposed analytic system.

Results: Significant correlations were found between PWA's performance on formal aphasia batteries and selected indices on the number of events produced as well as informative words used. Fluent and non-fluent PWA performed significantly differently in terms of their use of informative words and event sequence. The analytic system also demonstrated excellent intra- and inter-rater reliabilities.

Conclusions: The clinical feasibility and diagnostic values of this proposed approach to quantify PWA's spoken output are confirmed. The present investigation also offered more evidence in supporting the needed psychometric properties of spoken discourse analysis that can supplement traditional formal aphasia batteries.

Keywords: Aphasia; Discourse production; Clinical assessment; Stroke; Speech-language pathology

Received: November 11, 2021
Revision: December 31, 2021
Accepted: December 31, 2021

Correspondence:
Anthony Pak-Hin Kong

Academic Unit of Human Communication, Development, and Information Sciences (CDIS), The University of Hong Kong, Room 764, Meng Wah Complex, The University of Hong Kong, Hong Kong
Tel: +852-39171567
Fax: +852-25170075
E-mail: akong@hku.hk

INTRODUCTION

Aphasia refers to the neurological disorder caused by brain damage specific to regions responsible for language processing. The most common etiology of aphasia is stroke and up to 40% of stroke survivors acquire aphasia, but other related neurological conditions such as traumatic brain injury or dementia can also lead to this impairment. According to National Aphasia Association (https://www.aphasia.org/), aphasia is currently affecting about two million Americans and nearly 180,000 Americans develop the disorder each year; this is more common than Parkinson's Disease, cerebral palsy, or muscular dystrophy. Persons with aphasia (PWA) are often older adults and demon-

strate common language deficits including problems in the modalities of understanding, speaking, reading and writing [1,2]. Diagnosing the presence of and subsequently classifying syndromes of aphasia can be a fatiguing process on the part of PWA. Apart from the use of standardized aphasia batteries (such as Western Aphasia Battery – revised; WAB-R [3]), clinicians also review one or multiple samples of spoken narratives to determine the severity of aphasia [4]. According to Labov [5], a fundamental principle in analyzing narrative discourse is how to apply an objective and rule-governed manner to determine the appropriateness on how one utterance is followed by another. Specifically, there are multiple approaches to measure the linguistic complexity and variability of macro-linguistic structure of narratives. Some examples include quantifying one's ability to properly transition from one topic to another or to provide adequate information in order to develop a complete and coherent output. When observing clients with language disorders such as aphasia, the linguistic symptoms (e.g. lexical errors, types of words used, and grammatical processes) will likely interfere even more with their ability to relay the information in a narrative to listeners.

Many studies have reported the categorization of language samples at the word and morphological level which compared the use, misuse (e.g., paraphasias or perseverations), and/or absence of content and function words within a narrative (see review by Linnik et al. [6]). These studies focused on the micro-structures of language and tended to require an extensive amount of time for clinicians to sort and measure aphasic characteristics due to their need to look at the smaller aspects of narrative data. While attempting to adopt a time effective option for diagnosis of PWA, other studies have analyzed a broader spectrum of their language use and identified the macro-structures of spoken output, such as the proportion and consistency of presented propositions as well as the order of information that are necessary for a cohesive narrative.

Andreetta et al. [7] examined the macro-linguistic impairments in ten PWA with anomic aphasia and compared their performance with ten controls matched in age and level of formal education. Each of the participants was introduced to one picture stimulus and two cartoon picture sequence stories (with six pictures per story), and was asked to provide a narrative description. The narrative samples were analyzed in terms of quantity of content units and words, speech rate, and mean length of utterances (MLU) to reflect the degree of cohesiveness. In addition, an error index that considered the misuse of gender agreement, function words or semantically related content words, sentence completion, and the variables of topic transition from one utterance to another was used. It was concluded that the PWA displayed more errors in global cohesiveness (defined as events at the beginning of a narrative concluding by the end of the narrative) than local cohesiveness (defined by events presenting a matching relationship from one event to another). Concerning the informativeness (i.e., the average quantity of details used in the narrative), there was a lack of significant difference between the two groups' language samples.

Another system to quantify the deficits of macro-structures in PWA was reported by Capilouto et al. [8] that determined a speaker's ability to recall the main events throughout a narrative. In this investigation, language samples were elicited from eight PWA and their age- and education-matched controls using two picture description and another two story telling tasks. Subsequently, based on the performance of the control speakers, a listing of 24 main events that reflected the most commonly mentioned events in each of the stimuli was developed. These events were treated as the target events for subsequent quantification of the aphasic samples (with one point being awarded when a participant mentioned one of the target events). Each participant was rated for the total number of events listed after the prompt. The results revealed that the PWA was significantly less capable of expressing main events than the control group. The authors suggested this approach of counting main events was a time efficient method to indicate one's macro-structures of spoken output, but did not adequately specify if the contents were organized and whether the propositions were presented in a cohesive manor. To overcome this limitation, a more comprehensive method of analyzing macro-linguistic structures in oral narratives based on three aspects, including search theme, local connectedness, and global connectedness, was presented by Ash et al. [9]. The study used language samples elicited from of 32 speakers with Lewy body spectrum disorder using a story telling task of a wordless picture book and then quantified their narrative organization. First, in terms of the search theme, which was related to one's ability to mention the necessary events of the story that would give a listener the main idea of the production, a set number of critical events were pre-determined by reviewing the broad plot of the story. These events used were critical to the flow and comprehensions of the story and, therefore, were expected to be mentioned. The participants were scored on the total number of events mentioned. Second, the participants' ability to connect one utterance to the

next utterance by providing some form of joining word(s) and/or statement(s) was measured. This aspect of local connectedness reflected how well events were connected to each other throughout one's narrative output. The third category was global connectedness which was defined as the narratives' consistency of information from events to events. To obtain a credit (i.e., present versus absent) for this measurement, the same characters, locations, and events would need to be consistently and correctly mentioned from the beginning of the story to the end.

Conceptually comparable to the above-mentioned methods is the Narrative Assessment Profile (NAP) put forth by McCabe and Bliss [10] to evaluate PWA's discourse coherence. Apart from measuring three micro-structural categories (referencing, conjunction cohesion, and fluency) of an output, the NAP also quantified three macro-structural categories: topic maintenance, event sequencing, and informativeness. To estimate one's ability to remain on a topic and avoid tangential rambling or complete loss of topic, the topic maintenance of narratives was marked as 'present', 'not-present', or 'needs further study'. Note that this topic maintenance marking could also reflect one's ability to relate all utterances to a central topic. Concerning the event sequencing of the output, which was defined as one's ability to logically or chronologically present the narrative in an organized manner, the same marking criteria of 'present', 'not-present', or 'needs further study' was adopted. Any tangentiality appeared in the discourse was not penalized for event sequencing as long as there was a logical or chronological order to the sequences mentioned. Finally, the degree of informativeness that could address the narrative content (i.e., whether the production made sense to a listener) was rated as being 'appropriate', 'inappropriate', 'variable', or 'needs further study'. The NAP was clinically useful given its ease of application using simple and quick rating scales. The authors have also concluded that it was an effective means to analyze narrative discourse produced by typical adults as well as those with aphasia across different demographics (i.e., PWA with European, Asian, Spanish, and African American descents) but realized that the subjectivity involved in the evaluation could be a potential drawback.

Aims

The current investigation proposed a quick, objective, and comprehensive analytic system covering both the micro- and macro-linguistic structures of narratives from PWA and examined its application. Specifically, this system addressed three aspects of spoken discourse in aphasia: (1) events – which allowed one to measure the amount of important propositions provided within the narrative that were necessary to portray the output in a coherent manor, (2) informativeness – which would quantify the use of appropriate and correct lexical items related to the events that allowed listeners to understand what was happening within each event and if each event related to others, and (3) event sequencing – which reflected the individual's ability to organize and arrange the events in a logical manner that would allow listeners to understand the order or process of the events. While the definitions of the currently proposed comprehensive system followed the earlier-reviewed studies, this investigation aimed to modify the scoring criteria specific to three narrative tasks that one can subsequently measure the overall coherence of PWA's narrative.

METHODS

Stage 1 – Establishing scoring methods and criteria

Stage 1 of the present study utilized unimpaired language data selected from a publicly accessible language corpus called the AphasiaBank (https://aphasia.talkbank.org/ [11]). Three narrative tasks from ten control transcripts[1] were randomly selected, including sequential picture description of 'Refused Umbrella', procedural narrative of making a 'Peanut Butter and Jelly Sandwich', and telling of 'Cinderella' story. These participants were all proficient in the English language and passed a hearing screening. None of them had previous reports of neurological or psychological illness that would affect their use of spoken language (see demographic information in Table 1).

This dataset was used to establish normative data for the basis of analysis, in terms of three aspects including (1) Events, (2) Informativeness, and (3) Event sequencing.

Events

In order to have a cohesive narrative, the output should contain the events, i.e., 'actions of sufficient importance to the narrative as a whole and are separate from other actions in the narrative' (p.207) [8], needed to relay the information. The ten unimpaired transcripts were used as the basis for (1) determining events that were most commonly presented (at least 70% of occurrence), and (2) identifying the target events (meeting the 70% criterion) for each genre. To quantify the

Table 1. Demographic information of PWA and Controls

Subject	Gender	Age	Years of education	Aphasia Type	Aphasia Quotient (out of 100)	Occupation	Time-post-onset (years)
Unimpaired speakers (in Stage 1)							
C1	M	64.5	20	-	-	College history professor	-
C2	F	76.0	12	-	-	Realtor	-
C3	M	71.5	16	-	-	Mechanical engineer	-
C4	M	79.5	16	-	-	Realtor	-
C5	F	71.7	13	-	-	n/a	-
C6	M	73.5	18	-	-	Teacher	-
C7	F	81.5	12	-	-	Graphic artist	-
C8	M	75.6	18	-	-	Social worker	-
C9	F	36.8	14	-	-	Banquet server	-
C10	M	41.0	12	-	-	Advertising	-
PWA (in Stage 2)							
F1	M	65.0	12	Anomic	84.0	n/a	1.0
F2	F	73.6	14	Conduction	60.7	Project manager	1.0
F3	F	29.9	14	Conduction	82.1	n/a	3.4
F4	M	49.3	14	Anomic	90.8	Manager	4.2
F5	M	66.2	16	Wernicke's	53.0	Engineer	3.0
F6	F	49.0	12	Anomic	54.3	Waitress	8.9
F7	F	55.2	16	Conduction	88.4	Real estate broker	5.4
F8	F	69.5	14	Anomic	91.9	Computer programmer	5.8
F9	M	62.0	12	Anomic	88.8	Mechanic	6.5
NF1	M	64.3	12	Broca's	48.8	Manager	22.8
NF2	M	37.7	16	Broca's	54.7	Geophysicist	1.0
NF3	F	30.7	14	Broca's	72.8	n/a	1.9

Following the Classification of Federal Data on Race and Ethnicity (Office of Management and Budget), all unimpaired speakers and 11 PWA were White. One of the PWA (F8) was Hispanic or Latino. There were no American Indian or Alaska Native, Asian, Black or African American, or Native Hawaiian or Other Pacific Islander.

events needed for a receiver to follow, a binary system similar to the one used by Capilouto et al. (i.e., assignment of 1 for correct or 0 for incorrect event) was adopted to determine the presentation of events needed to develop a cohesive output for each of the three genres (see Appendix A). The following indices were calculated:

- (e.total): It was the total number of all (correctly and incorrectly) mentioned events.
- (e.matched): It was the total number of events mentioned that matched the target list.
- (e.missed): It was the total of events in the target list that were not mentioned.
- (e.irrelevant): It was the total number of events mentioned by a speaker that were irrelevant to the narrative theme or content. In most cases, these were events with no sufficient importance to the narrative as a whole and were independent from other events in the narrative.
- (e.extra): This is another summary score that counted the total number of extra events mentioned in the narrative. These extra events must be of sufficient importance to the narrative as a whole and were independent from other events in the narrative, but were not the ones in the target event list.

Informativeness

The conveying of events with correct semantic information, with suitable location in a narrative, allows one to logically present a message. To fully gather a cohesive narrative, a di-

mension of consistent information must also be present. In order to quantify the narratives for informativeness, the normative narratives were analyzed for the presence and the correct employment of key lexical items required for each event, similar to the procedures used in the NAP system [10]. However, the range of these lexical items was expanded to additional categories of part-of-speech, including specific characters, objects, locations, actions, and times that corresponded to the content of the three narrative tasks. Appendix B displays a full list of target lexical items (i.e., Informative Words) for each genre that were correctly used by at least 70% of the speakers. The following criteria were considered for considering the lexical informativeness:

- Each lexical item was directly mentioned at least once within the event.
- A synonym or alternative form of the target item was acceptable.
- If a name or label was initially mentioned, subsequent use of correct pronouns would receive credit as one lexical item.
- If a self-correction occurred, only the final production in that event would be counted.
- Four measures tallying the total number of correctly produced Informative Words were computed: (i.pb&j), (i.umbr), (i.cind), that were the total for the three genres respectively, and (i.total), which was the sum of these three indices.

Event sequencing

For a narrative to be comprehensible, the events listed must be formed in a logical order for the listener to follow. This order can be chronological to the events or systematical to the procedures. To quantify the participants' ability to present the events of a narrative in a logical order, a scoring system was devised for event sequencing, based on how the unimpaired speakers conveyed the target events (in 2.1.1) for each of the three genres. Specifically, the sequential order of mentioning the Events, i.e., common order of events in 90% of the speakers, (s.total) was determined. The index of (s.total) was a percentage score of correct event order that considered the following criteria:

- A sequence was counted accurate if events were presented in the correct order (e.g., credit was given to the sequence of target event 1 to target event 2, or target event 2 to target event 3)
- A sequence was still counted if the events were presented in the correct order with missing or skipped events (e.g., target event 1 to target event 3)
- If an event was mentioned out of order, credit was not given towards the index of (s.total) (e.g., target event 2 to target event 1). The number of violated event order was not tallied.
- If an event was mentioned in repetition, no additional credit was awarded nor penalized.
- If a self-correction occurred, the scoring was based on the final order in the language sample.
- Speakers were not penalized for target events that were irrelevant or missed.

Stage 2 – PWA data collection and scoring

Twelve participants (six male and six female) with a previous diagnosis of stroke-induced aphasia were recruited[2]. As confirmed by the WAB-R [3], there were nine fluent and three non-fluent PWA. Eleven of them were Caucasian and one was Hispanic. All of them, aged between 29.9 and 73.6 years, were tested under the AphasiaBank protocol. Table 1 shows the demographic information of these participants. According to PWA's self-report, none of them had any previous reports of neurological or psychological illness prior to a single left hemispheric stroke. All of them have passed a hearing screening and were proficient in the English language.

Following the AphasiaBank protocol [11], all PWA were asked to provide several spoken language samples (including free speech, picture descriptions, and story narratives[3]). Three sets of samples from each participant were used in this study:

1) 'Refused Umbrella' story transcripts elicited from a sequential picture description task using a six-picture sequence of black-and-white line drawings. Specifically, the PWA were asked to tell a story that had a beginning middle and an end.
2) Cinderella story telling transcripts elicited using a wordless story book of the story. After reviewing the pictures that depicted the story, the book was removed and each participant was asked to tell the story of Cinderella from what they had observed and from what they previously knew about the story.
3) The procedural description transcripts, elicited by asking the participants to recite how to make a 'Peanut Butter and Jelly Sandwich.'

The PWA's output in the above three discourse tasks were orthographically transcribed and the performance was analyzed using the following quantification measures developed

in Stage 1: (e.total), (e.matched), (e.missed), (e.irrelevant), (e.extra), (i.pb&j), (i.umbr), (i.cind), (i.total), and (s.total)[4].

Furthermore, with the exception of the aphasia quotients of the WAB-R [3], information of the PWA's performance in the following tests that were not part of the AphasiaBank protocol were collected at the same time the discourse samples were elicited:

1) scores of the Object and Action Naming Battery (OAB) [12], which consisted of 162 pictures targeting nouns and 100 pictures that targeted verbs and provided an in-depth analysis of the participants' word retrieval ability,
2) scores of The Pyramids and Palm Trees Test (PPTT) [13] that reflected the participants' associative relationship skills. In particular, the PPTT was presented in to PWA in two forms. While the first form involved the use of three printed words with two of them having a semantic relation, the second presented the same stimuli in a picture format. This test, therefore, provided insight into how intact or hindered the PWA's ability was in recognizing semantic relationships from orthographic lexical and visual decoding, and
3) scores of a Spoken word-Picture Matching task and a Written Word-Picture Matching task, compiled using the picture stimuli in Snodgrass and Vanderwart [14]. These two matching tasks presented the participants with a spoken or written target word, respectively, and three pictures (one of the target words, a distracter, and one irrelevant word) and provided information about the participants' visual-lexical and auditory lexical abilities.

Statistical analysis

Pearson product-moment correlation coefficients were used to measure the relationships between PWA's scores on standardized tests and the index scores generated in this investigation. An independent t-test was also conducted to determine if the fluent and non-fluent PWA performed differently. Three PWA samples (25% of PWA data) were randomly chosen to be re-analyzed for inter- and intra-rater reliabilities.

RESULTS

Descriptive statistics (mean, standard deviation, and range) of the proposed indices for all PWA as a group, and by the two fluency groups, are summarized in Table 2. Results of Pearson product-moment correlations (Table 3) revealed significant relations between WAB spontaneous speech scores and the both the Events of (e.matched) and (e.missed). Significant correlations were also found between Informative Words and the aphasia test scores of WAB spontaneous speech scores, WAB AQ, as well as OAB scores.

Results of an independent t-test (Table 4) suggested significant differences between the fluent and non-fluent PWA in sequential order of Events for all genres (s.total) and the total number of correctly used Informative Words (i.total), suggesting sensitivity of this framework to differentiate between the two fluency groups. There were also excellent reliabilities (Table 5). The mean value of coefficients for intra-rater reliability was 0.992, with significant correlations for (e.matched), (e.missed), and (e.extra); this mean value was higher than that for inter-rater reliability (0.909).

DISCUSSION

Aphasia is a long lasting disorder that impaired one's commu-

Table 2. Descriptive Summary of Performance among PWAs

		e.total	e.matched	e.missed	e.irrelevant	e.extra	s.total	i.pb&j	i.umbr	i.cind	i.total
All	Mean	28.50	11.92	9.08	7.25	9.33	0.59	6.58	7.08	9.50	23.174
	Standard Deviation	14.18	5.42	5.42	5.38	6.69	0.31	4.21	3.37	6.74	11.65
	Range	7-54	4-19	2-17	1-17	1-21	0.00-0.84	0-15	2-13	2-21	8-41
F	Mean	31.00	14.11	6.89	5.89	11.00	0.65	7.67	8.44	11.67	27.78
	Standard Deviation	14.87	4.28	4.28	4.99	6.56	0.27	4.36	2.65	6.36	9.51
	Range	7-54	4-19	2-17	1-14	1-21	0.00-0.84	0-15	4-13	2-21	10-41
NF	Mean	21.00	5.33	15.67	11.33	4.33	0.38	3.33	3.00	3.00	9.33
	Standard Deviation	10.44	1.15	1.15	5.13	4.93	0.40	0.58	1.00	1.73	1.53
	Range	14-33	4-6	15-17	7-17	1-10	0.00-0.80	3-4	2-4	2-5	8-11

All, All PWA participants; F, Fluent PWA group (n=9); NF, Non-fluent PWA group (n=3).

Table 3. Results of Pearson Correlations between Scores of the Standardized Tests and PWA's Performance Scores

	e.total	e.matched	e.missed	e.irrelevant	e.extra	s.total	i.pb&j	i.umbr	i.cind	i.total
WAB-R: Aphasia Quotient	.107	.537	-.537	-.451	.154	-.080	.686*	.598*	.584*	.759**
WAB-R: Spontaneous Speech	.354	.815*	-.815**	-.391	.404	.128	.761**	.824**	.659*	.849**
WAB-R: Spontaneous Speech (Content)	.243	.605*	-.605*	-.279	.250	-.188	.700*	.676*	.512	.745**
WAB-R: Spontaneous Speech (Fluency, grammatical competence and paraphasia)	.383	.853**	-.853**	-.417	.456	.322	.708**	.821**	.675*	.884**
PPTT: picture	-.325	-.236	.236	-.279	-.273	-.125	-.049	-.058	-.113	-.100
PPTT: word	-.086	-.032	.032	-.161	-.026	.094	.023	-.030	.158	.091
SWPM	-.285	-.119	.119	-.326	-.246	-.161	.156	.064	-.023	.062
WWPM	-.284	-.188	.188	-.246	-.252	-.196	.126	.012	-.074	.006
OAB: action	.134	.469	-.469	-.335	.174	-.168	.700*	.508	.489	.683*
OAB: object	-.740	.328	-.328	-.520	-.005	-.119	.650*	.432	.445	.617*

WAB-R, Western Aphasia Battery – revised; PPTT, Pyramids and Palm Trees Test; SWPM, Spoken word-Picture Matching task; WWPM, Written Word-Picture Matching task; OAB, Object and Action Naming Battery; *p<0.05, **p<0.01.

Table 4. Results of Independent T-test between performances of F and NF PWA

Measures	T-test results
e.total	-0.273
e.matched	0.954
e.missed	0.954
e.irrelevant	-0.666
e.extra	-0.400
i.total	0.999*
s.total	-1.000**

*p<0.05, **p<0.01.

Table 5. Results of Inter-rater and Intra-rater Reliabilities

Measures	Inter-rater reliability	Intra-rater reliability
e.total	.864	.990
e.matched	.977	1.00**
e.missed	.977	1.00**
e.irrelevant	.592	.991
e.extra	.999*	.998*
i.total	.975	.979
s.total	.980	.989

*p<0.05, **p<0.01.

nication and social well-being. It is common in survivors of stroke, for which older people are at higher risk. The condition of aphasia can also be found in other aging-related neurogenic disorders such as dementia. The investigation attempted to propose a quick, objective, and combined analysis that renders clinicians (or speech-language pathologists, in particular) a convenient but comprehensive way to quantify characteristics of micro- and macro-linguistic deficits in aphasic spoken discourse. In particular, with reference to three selected narrative tasks in the AphasiaBank protocol [11], we constructed a set of task-specific normative data and scoring criteria that can be directly applied clinically. Unlike most previous studies of macro-linguistic deficits in aphasia that relied on subjective ratings [e.g., 15,16], we believe the present comprehensive analysis is more objective but still time efficient. We argue that the use of quantification-derived numeric values (in contrast to subjective ratings such as those in the NAP system [10]) is a more refined way to ensure sensitivity of this proposed analysis. As expected, the fluent PWA group, as a whole, scored superior to their non-fluent counterparts on all proposed indices (see Table 2), such as higher (e.matched) or (i.total), lower (e.irrelevant), and better (s.total).

In addition, this study illustrated how we had adapted and, in certain extent, improved and expanded existing researched methods of analyzing macro-structures. Our preliminary results also suggested the diagnostic values of this new approach of quantifying PWA's spoken output. Concerning scoring of events, we have expanded the scope of measurement described in Capilouto et al. [8] by also counting irrelevant or additional (i.e., extra) events mentioned. With categorization of the part-of-speech of lexical items, the present scoring (or tallying) method of spoken events was more systematic than

the original NAP system [10]; this echoed the clinical feasibility and ease of analysis in similar studies that focused on the linguistic aspects of spoken narratives [17,18]. Although our results did not reveal a significant correlation between event sequence and existing standardized aphasia tests, this tended to be an under-researched aspect of aphasic discourse that warrants further examination.

What needs to be highlighted is the excellent reliability of this proposed system. With only a small subset of data being examined, we were able to show a strong across- and within-rater consistency in identifying individual events of the spoken discourse and the corresponding lexical items of each event, as well as determining the sequence presentation of the events. This suggested that this simple but objective analysis demonstrated good potential of clinical application, which is worth of further investigation involving a larger clinical sample.

A recent report by Stark et al. [19] has emphasized the heterogeneous nature of quantifying measures, descriptive methods, and analytic options reported in the PWA spoken discourse literature. The present study can offer more evidence in supporting the needed psychometric properties of spoken discourse outcomes. Similar to other clinically-oriented discourse analyses for PWA [e.g., 20-24], it may be possible to use this proposed method to conduct post-sentential testing in a more informal manner which will bring several benefits from a clinical perspective. Apart from playing a supplementary role to traditional formal aphasia batteries that do not adequately address discourse productions, this method can improve efficiency of discourse assessment and its subsequent analysis would contribute to testing accuracy and minimize fatigue on the part of PWA. According to the experience of research personnel in this study, the actual amount of time needed to complete an analysis for a single client's transcript was between ten to fifteen minutes; this seemed to be consistent to what most clinicians are currently doing.

The present investigation was limited at least in the following three aspects. First, the PWA participants were recruited in a university intensive aphasia intervention program. They were not interviewed by the same examiner. In other words, there could have been some variabilities relative to how the narrative samples were actually collected and how the different aphasia assessments were conducted, but this effect was mitigated through consistent and direct instructions, training, and supervision given to the multiple examiners. Second, the small sample size (in both Stages 1 and 2) might have hindered the homogeneity of participants. Subsequently, the comparison of fluent and non-fluent PWA, as well as the examination of external validity of discourse output indices in relation to standardized aphasia tests may be of diminished implications. Future investigations should include an increased sample size with inclusion of comparable numbers of PWA with different aphasia syndromes; this will allow one to determine how well the proposed system can be used to classify types of aphasia. Additional data collection also applies to the control population to potentially reveal any age, gender, or education effects on the index measures. Finally, only three genres were focused in this study. By building on this current work, how well this proposed analysis can be applied to other narrative tasks, such as conversation, open- or close-ended monologue, or single picture description, can be further studied.

In conclusion, the clinical feasibility and diagnostic values of this proposed approach to quickly, objectively, and comprehensively quantify PWA's spoken output are confirmed. The present investigation also offered more evidence in supporting the needed psychometric properties of spoken discourse analysis that can supplement traditional formal aphasia batteries.

Notes:

1. The use of ten unimpaired speakers followed the methodology reported in Kong, Lau, and Cheng [16] (n = 10) for establishing normative data as the basis of analysis.
2. This study received ethics approval from the University of Central Florida Institutional Review Board (IRB Number: SBE-12-08810).
3. According to the results of t-tests, the PWA group was not statistically different from the group of ten unimpaired speakers in Stage 1 in terms of both age and years of education. These 12 PWA were also asked to provide additional language samples in the AphasiaBank protocol [11] that were not utilized in this present study. The complete language samples (elicited by other narrative tasks) are available in the AphasiaBank (https://aphasia.talkbank.org/).
4. Specifically, all test data were collected by student clinicians under the supervision of a CCC-SLP who is familiar with the AphasiaBank protocol. All subsequent data processing, scoring, intra-rater reliability, and statistical analysis was performed by author AR.

CONFLICT OF INTEREST

No potential conflict of interest was reported by the author.

Funding/Support Statement

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

ACKNOWLEDGMENTS

The authors are grateful to Dr. Janet Whiteside for her input to the study. Special thanks to Dr. Audrey Holland and Dr. Margie Forbes for offering support and training to AR at the beginning of this investigation.

REFERENCES

1. National Institute on Deafness and Other Communication Disorders (NIDCD). Aphasia. 2015;NIH Pub. Ni. 97-4257.
2. Worrall L, Simmons-Mackie N, Wallace SJ, et al. Let's call it "aphasia": Rationales for eliminating the term "dysphasia". Int J Stroke 2016;11:848-51.
3. Kertesz A. Western Aphasia Battery—Revised (WAB-R). Pearson; 2007.
4. Kong APH. Analysis of neurogenic disordered discourse production: From theory to practice. London: Routledge; 2016.
5. Labov W. Sociolinguistic patterns. Philadelphia: University of Pennsylvania Press; 1972.
6. Linnik A, Bastiaanse R, Höhle B. Discourse production in aphasia: a current review of theoretical and methodological challenges. Aphasiology. 2016;30:765-800.
7. Andreetta S, Cantagallo A, Marini A. Narrative discourse in anomic aphasia. Neuropsychologia. 2012;50:1787-93.
8. Capilouto GJ, Wright HH, Wagovich SA. Reliability of main event measurement in the discourse of individuals with aphasia. Aphasiology. 2006;20:205-16.
9. Ash S, McMillan C, Gross RG, Cook P, Morgan B, Boller A, Grossman M. The organization of narrative discourse in lewy body spectrum disorder. Brain Lang. 2011;119:30-41.
10. McCabe A, Bliss LS. Patterns of narrative discourse, a multicultural, life span approach. Place of publishing: Allyn& Bacon; 2003.
11. MacWhinney B, Fromm D, Forbes M, Holland A. AphasiaBank: Methods for studying discourse. Aphasiology. 2011;25:1286-1307.
12. Druks J, Masterson J. An Object and Action Naming Battery. London, England: Psychology Press; 2000.
13. Howard D, Patterson K. The Pyramids and Palm Trees Test: A test for semantic access from words and pictures. Bury St Edmunds: Thames Valley Test Company; 1992.
14. Snodgrass JG, Vanderwart M. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J Exp Psychol Hum Learn. 1980;6:174-215.
15. Van Leer E, Turkstra L. The effect of elicitation task on discourse coherence and cohesion in adolescents with brain injury. J Commun Disord. 1999;32:327-49.
16. Kong APH, Lau DKY, Cheng CYY. Analysing coherence of oral discourse among Cantonese speakers in Mainland China with traumatic brain injury and cerebrovascular accident. Int J Speech Lang Pathol. 2020;22:137-47.
17. Kong APH, Law SP. A Cantonese linguistic communication measure for evaluating aphasic narrative production: normative and preliminary aphasic data. J Multiling Commun Disord. 2004;2:124-46.
18. Kong APH, Law SP. A linguistic communication measure for monitoring changes in Chinese aphasic narrative production. Clin Linguist Phon. 2009;23:255-69.
19. Stark BC, Dutta M, Murray LL, et al. Standardizing assessment of spoken discourse in aphasia: A working group with deliverables. Am J Speech Lang Pathol. 2021;11:491-502.
20. Kong APH, Whiteside J, Bargmann P. The Main Concept Analysis: Validation and sensitivity in differentiating discourse produced by unimpaired English speakers from individuals with aphasia and dementia of Alzheimer type. Logoped Phoniatr Vocol. 2016;41:129-141.
21. Richardson JD, Hudspeth-Dalton SG, Shafer J, Patterson J. Assessment fidelity in aphasia research. Am J Speech Lang Pathol. 2016;25:S788-97.
22. Olness GS, Ulatowska HK. Personal narratives in aphasia: Coherence in the context of use. Aphasiology. 2011;25:1393-413.
23. Kong APH, Wong CWY. An integrative analysis of spontaneous storytelling discourse in aphasia: Relationship with listeners' rating and prediction of severity and fluency status of aphasia. Am J Speech Lang Pathol. 2018;21:1491-1505.
24. Hilger A, Ramsberger G, Gilley P, Menn L, Kong A. Analysing speech problems in a longitudinal case study of logopenic variant PPA. Aphasiology. 2014;28:840-861.

Appendix A. Target events (that appeared in at least 70% of the normative scripts) of the three genres

Genre	Target Events									
Refused umbrella	UE.1 Women warns boy	UE.2 Boy refuses umbrella	UE.3 It begins to rain	UE.4 Boy returns home	UE.5 Boy accepts the umbrella	UE.6 Coda				
Cinderella	CE.1 Mentioning of sisters	CE.2 Mentioning Cinderella is a servant	CE.3 Invitation to Ball	CE.4 God mother appears	CE.5 Cinderella goes to the ball	CE.6 Cinderella and prince interact	CE.7 Slipper is lost	CE.8 Search for slipper owner	CE.9 Fit check of slipper	CE.10 Coda
Peanut Butter and Jelly	PE.1 Get ingredients	PE.2 Apply peanut butter	PE.3 Apply jelly	PE.4 Fold bread together	PE.5 Coda					

Appendix B. Target informative words (lexical items that appeared in at least 70% of the normative scripts) of the three genres

1). i.umbr in the Refused umbrella genre

	Characters	Objects	Locations	Actions	Times	Others
UE1 (Women warns boy)	boy	backpack	school	gave	day	him
	mother	hat		get (ready)		rain
		umbrella		go (to)		
				head (to)		
				leave		
				take		
				warn		
UE2 (Boy refuses umbrella)	mother	umbrella		(doesn't) need		nope
	boy			get		
				insist		
				refuse		
				try		
				want		
				wave		
UE3 (It begins to rain)	mother	backpack	school	go		it
	boy			rain		he
				begin		she
				getting		wet
				start		
				trot		
				walking		
UE4 (Boy returns home)	mother	hands	house	head (to)		back
	boy			go		his
				run		wet
UE5 (Boy accepts the umbrella)	mother	umbrella		head (to)		her
	boy			prove		his
				succumb		
				take		
				told		
UE6 (Coda)	mother		school	know (best)		dry
	boy			send		his
				go		out
				walk		

(*Continued to the next page*)

Appendix B. Continued

2). i.cind in the Cinderella story genre

	Characters	Objects	Locations	Actions	Times	Others
CE1 (Mentioning of sisters)	daughters characters Cinderella father girl stepsister women			face remarried		family two
CE2 (Mentioning Cinderella is a servant)	Cinderella family girl maid servant	clothes dishes		clean (do) everything look(down) treat wash work		cruel
CE3 (Invitation to Ball)	bride king prince wife young man	letter	ball castle home palace party	decided find go invited looking not (go) received stay wanted work		(up)coming chores gala word
CE4 (God mother appears)	coachmen godmother fairy horses	(magic) wand carriage clothes dress stick trunk window	attic palace	appear discover find furnish go make read tell want	twelve midnight	upset
CE5 (Cinderella goes to the ball)	Cinderella horse	carriage dress pumpkin	ball party	ride take		(on) time everything there
CE6 (Cinderella and prince interact)	bride lady prince	(glass) slipper clock		dance found meets strike	twelve midnight	impressed problem

(*Continued to the next page*)

Appendix B. Continued

	Characters	Objects	Locations	Actions	Times	Others
CE7 (Slipper is lost)		(glass) slipper shoe	home	changed fall get back (home) leave lose run	twelve midnight	
CE8 (Search for slipper owner)	Cinderella damsel prince someone	(glass) slipper	countryside home house town	come find hunt look for scoured went (through)		match everybody
CE9 (Fit check of slipper)	(step) sisters Cinderella stepmother	(glass) slipper feet / foot toes shoe		accept cutting find fit found put try		nobody everybody
CE10 (Coda)	couple Cinderella prince			lived marry	ever(after)	everybody happy

3). i.pb&j in the Peanut Butter and Jelly Sandwich genre

	Characters	Objects	Locations	Actions	Times	Others
PE1 (Get ingredients)		basket bread drawer jar jelly peanut butter pieces plate refrigerator utensils	home kitchen store	buy go open slice take (out)		supplies
PE2 (Apply peanut butter)		bread finger knife peanut butter piece		put slice spread		on

(*Continued to the next page*)

Appendix B. Continued

	Characters	Objects	Locations	Actions	Times	Others
PE3 (Apply jelly)		bread		put		on
		finger		slice		
		jam		spread		
		jelly				
		knife				
		piece				
PE4 (Fold bread together)		finger	top	cut		on
		knife		fold (over)		together
		peanut butter		grill		
				put		
				slap		
				slice		
PE5 (Coda)		jelly		eat		happy
		peanut butter		give		your
		sandwich		serve		

Appendix C. Event sequence (that appeared in at least 90% of the normative scripts) of the three genres

Sequences	Current event	Next mentioned event
Refused umbrella		
US.1	UE.1	UE.2 or UE.3 or UE.4 or UE.5 or UE.6
US.2	UE.2	UE.3 or UE.4 or UE.5 or UE.6
US.3	UE.3	UE.4 or UE.5 or UE.6
US.4	UE.4	UE.5 or UE.6
Cinderella		
CS.1	CE.1	CE.2 or CE.3 or CE.4 or CE.5 or CE.6 or CE.7 or CE.8 or CE.9 or CE.10
CS.2	CE.2	CE.3 or CE.4 or CE.5 or CE.6 or CE.7 or CE.8 or CE.9 or CE.10
CS.3	CE.3	CE.4 or CE.5 or CE.6 or CE.7 or CE.8 or CE.9 or CE.10
CS.4	CE.4	CE.5 or CE.6 or CE.7 or CE.8 or CE.9 or CE.10
CS.5	CE.5	CE.6 or CE.7 or CE.8 or CE.9 or CE.10
CS.6	CE.6	CE.7 or CE.8 or CE.9 or CE.10
CS.7	CE.7	CE.8 or CE.9 or CE.10
CS.8	CE.8	CE.9 or CE.10
CS.9	CE.9	CE.10
Peanut Butter and Jelly		
PS.1	PE.1	PE.2 or PE.3 or PE.4 or PE.5
PS.2	PE.2	PE.3 or PE.4 or PE.5
PS.3	PE.3	PE.2 or PE.4 or PE.5
PS.4	PE.4	PE.5