

Global and local coherence in Mandarin Chinese aphasia spoken discourse

Yi Hu & Juqiang Chen

To cite this article: Yi Hu & Juqiang Chen (22 Aug 2025): Global and local coherence in Mandarin Chinese aphasia spoken discourse, *Aphasiology*, DOI: [10.1080/02687038.2025.2549740](https://doi.org/10.1080/02687038.2025.2549740)

To link to this article: <https://doi.org/10.1080/02687038.2025.2549740>



Published online: 22 Aug 2025.



Submit your article to this journal [↗](#)



Article views: 38



View related articles [↗](#)



View Crossmark data [↗](#)



Global and local coherence in Mandarin Chinese aphasia spoken discourse

Yi Hu ^{a,b} and Juqiang Chen ^{a,b}

^aSchool of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China; ^bNational Research Centre for Language and Well-being, Shanghai, China

ABSTRACT

Background: The overall quality of spoken discourse production has been increasingly recognized as an important part of aphasia assessment. Coherence analysis offers crucial information about the discourse organization and has been used in examining English aphasia discourse. Very few studies so far have investigated coherence in aphasia discourse of Mandarin, a topic-prominent language in which speakers rely more on implicit semantic connectivity via contextual information than explicit lexical and grammatical devices to organize discourse.

Aims: The current study aims to examine global and local coherence in different types of discourse produced by Mandarin persons with aphasia (PWA) and test their validity in assessing Mandarin PWA discourse.

Methods & Procedures: Discourse transcripts of 18 Mandarin PWA and 18 age- and education-matched health controls were extracted from Mandarin section of AphasiaBank for two discourse task types, i.e. picture descriptions and story narratives. Global and local coherence were rated using four-point rating scales.

Outcomes & Results: Both global and local coherence scores of aphasia discourse were significantly lower than those of healthy people across most tasks. However, discourse coherence was influenced by task types, with performance varying depending on the nature of the tasks. Global and local coherence scores were positively correlated with aphasia severity and discourse informativeness measures for all tasks, but only significantly correlated with lexical diversity for story narrative tasks.

Conclusions & Implications: Global and local coherence reliably differentiate discourse production between Mandarin PWA and the control group, highlighting the value of coherence measures in assessing discourse functionality in Mandarin PWA.

ARTICLE HISTORY

Received 10 March 2025

Accepted 14 August 2025

KEYWORDS

coherence; Mandarin Chinese; aphasia; discourse analysis

1. Introduction

People with aphasia (PWA) experience difficulties in discourse production, which could disrupt their ability to express coherent thoughts and ideas (Pritchard et al., 2017). Spoken discourse production has been increasingly recognized as a clinically valuable measure of language competence in aphasia (J. Chen & Chang, 2024,

J. Chen & Chang, 2025; Deng, Gao, et al., 2024; Deng, Liang, et al., 2024). Although discrete linguistic measures at the word and sentence levels, such as sample length, type-token ratios and syntactic complexity, can reveal important information of word-finding and sentence-producing difficulties experienced by PWA, these measures often fall short in capturing the overall complexity and functionality of discourse, which is beyond single clauses or individual sentences (Armstrong, 2000). Recent studies have investigated coherence, i.e., the speaker's ability to maintain thematic unity throughout discourse organization (Glosser & Deser, 1990), to understand how discourse-level information is structured and organized in PWA discourse (e.g., Christiansen, 1995, Glosser & Deser, 1990; Hazamy & Obermeyer, 2020; Wright & Capilouto, 2012).

Coherent discourse requires the selection of appropriate lexical representations and integration of these units to denote related facts in the "real world" (Marini & Urgesi, 2012). An important distinction should be made between coherence (the focus of the present study) and a related but distinct concept: cohesion, namely the lexical and grammatical means, such as coreference, substitution, and conjunction, that facilitate meaning connectedness between discourse elements (c.f., Halliday & Hasan, 2014; Zhang et al., 2021). Additionally, coherence can be further differentiated into global and local coherence. Global coherence is the general relatedness of each utterance/sentence to a central topic, whereas local coherence refers to the maintenance of association between adjacent utterances within a text (Coelho & Flewellyn, 2003). English speakers with aphasia often fail to properly construct a coherent discourse that is reasonably understandable for discourse tasks. In a picture-assisted story narrative, for example, the healthy people produced utterances that were explicitly related to the pictures by mentioning main characters, actions, or objects, whereas PWA neglected substantive information and required more inferences in the listeners (Wright & Capilouto, 2012). Apart from the overall organization, relations between adjacent utterances were also disrupted by abrupt topic switches and ambiguous referents in PWA discourse (Andreetta et al., 2012).

There are two main approaches in measuring coherence: the error-based and rating-based approach. The error-based approach counts the number of coherence violations such as information gaps, repetitions, and tangential or erroneous propositions, which has proven to be reliable in evaluating global coherence in aphasia. People with fluent aphasia produced significantly more errors and diverse error types compared with healthy people. These errors, particularly coherence violations such as incorrect, non-specific, and off-topic utterances, were identified as impairing global coherence in fluent aphasia (Christiansen, 1995; Hazamy & Obermeyer, 2020). In addition, people with non-fluent aphasia also struggled to maintain global coherence, primarily due to the production of fragmented or incomplete utterances (Hazamy & Obermeyer, 2020). Even without prominent aphasia syndromes, post-stroke patients made significantly more errors in global and local coherence maintenance (Barker et al., 2017). However, the error-based analysis cannot fully capture the complexity of coherence, because they mainly focus on textual coherence without considering extralinguistic elements (e.g., shared knowledge, visual context). These external factors transform linguistically limited content into coherent and situationally appropriate content (Linnik et al., 2016). In addition, transcript-based error counting is time-consuming and labour-intensive and thus has limited application in clinical settings.

On the other hand, the rating-based approach via five-point (Coelho & Flewellyn, 2003; Glosser & Deser, 1990; Van Leer & Turkstra, 1999) or four-point (Hazamy & Obermeyer, 2020; Wright, Capilouto, et al., 2013; Wright & Capilouto, 2012) coherence scales quantifies coherence as a general property of discourse and is more clinically practicable (Linnik et al., 2016). PWA demonstrated significantly lower global coherence scores for story narratives (Wright & Capilouto, 2012), picture descriptions (Hazamy & Obermeyer, 2020), and interactive conversations (Leaman & Edmonds, 2021). In addition, a longitudinal study observed that PWA showed measurable improvements in global coherence over 12 months, likely due to spontaneous recovery, as the study did not report any discourse-specific treatment. However, their performance remained significantly below that of healthy individuals (Coelho & Flewellyn, 2003). While most studies generally supported coherence deficits in PWA discourse production, a few studies found comparable global and local coherence scores between PWA and control groups (Glosser & Deser, 1990; Van Leer & Turkstra, 1999). In these studies, participants recounted familiar personal experiences, which required less cognitive efforts to structure a logical discourse (Marini & Urgesi, 2012; Van Leer & Turkstra, 1999; Wright, Koutsoftas, et al., 2013). The absence of a fixed goal or theme further minimized the differences in coherence scores between the two groups (E. C. Li et al., 1995).

Coherence disruptions in PWA discourse have been found to be related with lexical retrieval deficits (Andreetta & Marini, 2015; Andreetta et al., 2012), reduced information (Hazamy & Obermeyer, 2020; Wright & Capilouto, 2012), as well as disrupted logical temporal-causal order (Ulatowska et al., 2013) and event sequences (Kong et al., 2020). Reduced lexical productivity was correlated with global coherence disturbance, because word retrieval deficits contributed to difficulties in maintaining semantic unity (Andreetta & Marini, 2015; Hazamy & Obermeyer, 2020; Wright & Capilouto, 2012).

In addition, there exist correlations between coherence and discourse informativeness measures, such as Correct Information Unit (CIU; Carlomagno et al., 2011) and Main Concepts (MC; Kong et al., 2016). CIU has been used to quantify the intelligible, relevant, and informative contents in the aphasia discourse. PWA produced significantly decreased amount of CIUs compared to healthy controls (Carlomagno et al., 2011; Deng, Liang, et al., 2024; Nicholas & Brookshire, 1993). MC is the essential information about a picture or topic that evaluates the presence, completeness and accuracy of main ideas in a discourse task. PWA demonstrated significant declines in the completeness and correctness of MC production (Kong et al., 2016; Nicholas & Brookshire, 1995). Thus, the correlations between coherence and CIU and MC suggest that the coherent presentation of a discourse facilitates listeners' understanding and improves their perception of informativeness (Hazamy and Obermeyer, 2020; Kong et al., 2020; Nicholas and Brookshire, 1995; Ulatowska et al., 2013). Importantly, although CIU and MC provides relevancy of discourse production to the central information required by the task, which is closely tied to coherence, global/local coherence ratings provide more and different insights into discourse compared to CIU and MC. For example, MC analysis focuses on the accurate and complete presence of ideas based on individual utterance, whereas coherence ratings specifically evaluate how multiple utterances are logically organized and connected to construct meaningful discourse. Thus, coherence ratings can identify subtle discourse deficits in individuals who may produce most main concepts but present them in a disorganized or disconnected manner. While previous research has examined these

relationships separately, the present study integrates both lexical- and discourse-level measures to provide a more comprehensive understanding of how they interact with global and local coherence in PWA discourse.

Most PWA discourse research so far has mainly examined coherence in English, a subject-prominent language structured around grammatical relationships. Coherence is even more relevant in evaluating topic-prominent languages such as Chinese, in which discourse is organized primarily through contextual and semantic relations (C. N. Li & Thompson, 1989). In other words, Chinese speakers rely more on implicit semantic connectivity through contextual cues and logical relationships rather than explicit lexical and grammatical markers to organize discourse. Consequently, a comprehensive assessment of Chinese discourse requires analysis of coherence to provide a more direct and reliable measure of discourse organization, as it captures the underlying meaning structure that governs communication in the absence of overt cohesive devices. Only a few PWA discourse studies have examined the issue of coherence in Chinese (i.e., Chow et al., 2016; Kong et al., 2018, 2020, 2023). Picture descriptions and story narratives by Cantonese PWA showed impaired global and local coherence as measured by a five-point scale (Kong et al., 2020), suggesting problematic event sequence and impaired language integrity. Similarly, Kong et al. (2018) conducted a coherence analysis of Cantonese PWA discourse using the Rhetorical Structure Theory (Mann & Thompson, 1988) and found decreased coherence in storytelling and procedural descriptions by PWA. This reduced coherence was a result of reduced content, shallow elaboration, and structural disruptions. However, a more recent study involving a Chinese-speaking PWA group with anomic aphasia found no significant differences in global and local coherence errors between PWA and healthy controls (Kong et al., 2023). It should be noted that both the PWA and control groups included speakers of Mandarin and Cantonese, but the study did not compare coherence performance between the two dialect groups. Given that the existing research on coherence in Cantonese and Mandarin PWA discourse has yielded mixed findings, possibly due to varied coherence measures (Kong et al., 2023), there is a need for a more comprehensive investigation of coherence in Mandarin discourse using valid rating scales.

In the present study, we report an analysis of global and local coherence of Mandarin PWA discourse production as rated with four-point scales. To ensure the validity of our coherence analysis, we implemented two types of validity tests: first, we examined known-groups validity to verify the sensitivity of our measure to distinguish between two groups that are known to differ in language abilities (Davidson, 2023); second, we assessed convergent validity to ensure our tests exhibit a strong relationship with other established and conceptually similar tests (Eignor, 2013). These two tests are standard psychometric approaches in speech-language pathology assessment field (e.g., Blomert et al., 1994; McNeil et al., 2001). For the known-groups validity, we compared discourse produced by Mandarin PWA with control participants in terms of global and local coherence. We hypothesize that discourse production of PWA should be lower in both global and local coherence scores than control group. For the convergent validity, we correlated global and local coherence scores with aphasia severity, lexical diversity, and discourse informative measures. It is expected that both global and local coherence scores should correlate strongly with aphasia severity scores, i.e., AQs of the WAB (Kertesz, 2012), and informativeness measures, i.e., CIU (Nicholas & Brookshire, 1993)

and MC (Kong et al., 2016), but only moderately with lexical diversity measures because coherence primarily reflects discourse organization and meaning connectedness rather than lexical variability alone.

2. Materials and methods

2.1. Data

The data of the present study were extracted from the Mandarin section (Z.-Z. Chen et al., 2018) of the AphasiaBank database (MacWhinney et al., 2011). In the original English AphasiaBank, each participant was asked to perform five types of discourse tasks, including greetings, picture description, story narrative, procedural discourse and free speech. The Mandarin section maintained all five task types with some modifications. The *Cinderella* story narrative was substituted with two stories that were commonly known in China, i.e., *The Tortoise and The Hare* and *Cry Wolf*. The *Peanut Butter and Jelly Sandwich* procedural discourse task was replaced by *Egg Ham Fried Rice* description (refer to the AphasiaBank website <https://aphasia.talkbank.org/protocol/> for detailed information). We excluded the greeting task due to the limited amount of data available for each participant. In addition, the free speech task was not selected because it exhibited excessive variability, making it difficult to ensure comparability across participants. The procedural discourse task, in which participants described how to cook fried eggs and rice, was excluded because it lacked a clear and consistent goal or expected outcome, making it challenging to effectively assess global coherence. A total of four tasks were selected for the final analysis: Picture Description (A. *Broken Window* and B. *Refused Umbrella*) and Story Narrative (C. *The Tortoise and The Hare* and D. *Cry Wolf*).

2.2. Participants

Participants were selected based on the following criteria: (i) aphasia resulted from a stroke that was verified via neuroimaging or a clinical diagnosis, (ii) Mandarin Chinese as the primary language, and (iii) no history or presence of dementia or comorbidities associated with serious cognitive consequences. All participants were from Mainland China. Eighteen PWA (8 females, $M_{age} = 44.6$) and eighteen healthy controls (9 females, $M_{age} = 47.9$) were included in the present study. No significant differences were found between the two groups in age ($t = 0.78$, $df = 33.93$, $p = 0.44$) or education years ($t = -0.86$, $df = 25.85$, $p = 0.40$, see Table 1). PWA group, classified using WAB-R (Kertesz, 2012), included anomic ($n = 15$), Broca's ($n = 1$), Wernicke's ($n = 1$), and conduction ($n = 1$) aphasia. PWA group demonstrated a mean AQ of 78.4 ($SD = 12$; range: 54–92.9) on the WAB-R.

Linguistic productivity, i.e., the number of characters, word types, and word tokens, was calculated to provide an overview of the differences between PWA and healthy

Table 1. Demographic information for control and PWA groups.

Group	n	Age (yrs)		Education (yrs)		WAB AQ		Post-onset Month	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Control	18	47.9	12.7	12.8	3.1	–	–	–	–
PWA	18	44.6	13.3	14.2	5.9	78.4	12	6.3	5.8

Table 2. Number of characters, word types, and tokens for each task per group.

Tasks	Groups	Characters		Types		Tokens	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Pic-Umbrella	Control	131.6	47.3	53.2	14.5	100.1	32.8
	PWA	74.8	40.6	31.8	13.6	58.0	32.7
Pic-Window	Control	120.4	64.5	48.2	14.0	90.2	48.7
	PWA	51.1	30.9	23.6	10.8	39.1	24.2
S-Crywolf	Control	246.9	126.3	78.3	29.1	189.4	97.2
	PWA	96.5	73.0	37.7	23.2	73.6	57.3
S-Tortoise	Control	240.6	113.2	84.2	29.4	168.3	80.0
	PWA	96.0	60.9	33.5	18.1	68.3	42.9

Note: Pic = picture description; S = story narrative.

controls (see Table 2). Mann-Whitney U Tests indicated that language production of healthy controls and PWA were significantly different in all tasks (see the online supplementary material Table A.1 for statistical details available at https://osf.io/cvx7t/?view_only=298198e3fc894703a793871ec3dce554).

2.3. Data processing

The Computerized Language Analysis (CLAN) program (Version 05-Mar-2022) was employed to extract transcripts from each task using the command (e.g., for Broken Window task, `gem + swindow + n + d1 + t*PAR + t%mor + t%gra + f.cha.`). Subsequently, clear speech was extracted (via `flo + cr + t@`), systematically excluding annotated repetitions, repairs, and other forms of disfluency. Utterance boundaries were manually determined by the transcribers of the original corpus and were used for coherence analysis. All stories were randomly ordered and analyzed by the first author with no prior knowledge of participants' demographic information.

The global and local coherence of each task and participant were rated by the first author. Global coherence refers to the relevance of an utterance to the general topic and was quantified via a 4-point scale by Wright et al. (2010) for the present study (see Table 3). This 4-point scale was reported to have stronger correlation with aphasia severity than the 5-point scale (Wright, Koutsoftas, et al., 2013). Each utterance within a story received a score for global coherence.

Table 3. Four-point rating scales for global coherence (Wright et al., 2010).

Score	Scoring Criteria
4	The utterance is overtly related to the stimulus as defined by mention of actors, actions, and/or objects present in the stimulus which are of significant importance to the main details of the stimulus. In the case of procedural descriptions and reactions when a designated topic acts as the stimulus, overt relation is defined by provision of substantive information related to the topic so that no inference is required by the listener.
3	The utterance is related to the stimulus or designated topic but with some inclusion of suppositional or tangential information that is relevant to the main details of the stimulus; or substantive information is not provided so that the topic must be inferred from the statement. In recounts, appropriate elaborations that are not essential but are related to the main topic are scored a 3.
2	The utterance is only remotely related to the stimulus or topic, with possible inclusion of inappropriate egocentric information; may include tangential information or reference some element of the stimulus that is regarded as non-critical.
1	The utterance is entirely unrelated to the stimulus or topic; the utterance may be a comment on the discourse or tangential information is solely used.

Local coherence refers to the relevance of an utterance to that of the preceding utterance. We adapted the 5-point scale for assessing local coherence, originally developed by Glosser and Deser (1990) and Van Leer and Turkstra (1999), into a 4-point scale. We first identified the conceptual ambiguity and inconsistency among the original scores of 2 (one clause possibly relates to the content in the preceding utterance but another may not), 3 (one clause generally relates to that of the preceding utterance), and 4 (one definitely relates to the content of the preceding utterance but the other (s) may not). Our adaptation preserved the distinct endpoints (original scores 5 and 1) as our new scores 4 and 1, respectively, and the original midpoint (score 3) as our score 3. We developed a more precise operationalization definition for score 2 that clarifies the partial relationship category. This restructuring eliminated problematic categorical overlap while maintaining discriminatory validity across the rating continuum. Negotiation and iterative revisions have been completed by the two authors until consensus was built. Employing 4-point scales for both global and local coherence allowed the raters to apply comparable judgement criteria across both coherence dimensions, enhancing scoring consistency. Moreover, identical scale structures also facilitated direct statistical comparisons between global and local coherence performance. In the final 4-point scale, local coherence scores for each utterance could range from being strongly and clearly related (scored 4) to being entirely unrelated (scored 1) to the preceding utterance (See Table 4). The relation may be achieved via elaboration, exemplification, repetition, and other kinds of continuation of the preceding utterance. The first utterance of each story was not rated for local coherence since there was no preceding utterance. After the coherence rating was finished, the average scores for global and local coherence were calculated for each participant in each task.

To ensure the reliability of coherence ratings, we tested the inter-rater reliability. Another rater was first familiarized with the scoring rubrics and then asked to complete the ratings for all stories independently. Agreement between two raters was assessed using Krippendorph's alpha (Krippendorff, 2019). The inter-rater correlation coefficients were 0.90 for global coherence and 0.92 for local coherence, indicating near perfect agreement between the two raters.

The coherence scores of the PWA group were validated in two ways. Known-group validity was evaluated by the differences between healthy control and PWA groups in

Table 4. Four-point rating scales for local coherence (adapted from Glosser & Deser, 1990; Van Leer & Turkstra, 1999).

Score	Scoring Criteria
4	The utterance is strongly and clearly related to the preceding utterance by continuing the topic through elaboration, temporal sequencing, enumeration of related examples, or maintaining the same actor, subject, action, or argument as the focus. The utterance provides substantive information clearly related to the preceding utterance, requiring no inference by the listener.
3	The utterance generally relates to the topic of the preceding utterance but may include a shift in focus from the subject or activity of the preceding utterance. The utterance may be vague or incomplete, requiring the listener to infer the relationship to the preceding utterance.
2	The utterance is only partially related to the elements of the preceding utterance or the connection may be so ambiguous or fragmented that the listener can barely infer. In the case where it contains multiple clauses, one clause relates to the preceding utterance but others do not.
1	The utterance is entirely unrelated to the topic of the preceding utterance. It may introduce an irrelevant topic or contain no substantive information related to the preceding utterance.

coherence scores. We hypothesize that Mandarin PWA should perform significantly worse in both global and local coherence than the control group. On the other hand, convergent validity was evaluated by the correlation between coherence scores and aphasia severity (AQ in WAB-R), lexical diversity, and two informativeness measures (i.e., Correct Information Unit and Main Concept scores). We hypothesize that coherence scores should have strong correlations with AQs.

Lexical diversity pertains to a speaker's range and breadth of vocabulary (Fergadiotis & Wright, 2011). We utilized Guiraud's R (Herdan & Guiraud, 1961), the root transformation of the type-token ratio to measure lexical diversity, which has proven to be valid for Mandarin aphasia discourse production (J. Chen & Chang, 2024). We do not anticipate strong correlations between lexical diversity and coherence scores because diverse vocabulary does not inherently indicate more coherent or logical discourse.

Discourse informativeness was quantified by Correct Information Unit percent (percent CIU), following Nicholas and Brookshire (1993). CIU refers to words that are intelligible in context, relevant to the content, and informative about the pictures or topic. To account for the variations in text length, we calculated percent CIUs, i.e., the total number of words that are CIUs divided by the total number of count words, for each story produced by each participant, which has been demonstrated to be stable across tasks (Deng, Liang, et al., 2024; Nicholas & Brookshire, 1993). We expect a strong correlation between percent CIU and both global and local coherence scores.

Another measure of discourse informativeness is Main Concept score (MC score) that evaluates the completeness and accuracy of main ideas in a discourse task (Kong et al., 2016). The main concepts for each discourse task were first listed by two coders, and disagreements were solved through discussion until a consensus was reached by both coders (see Table A.2 for the main concepts for four tasks in the present study). Each discourse sample was analyzed in terms of the presence, accuracy, and completeness of all main concepts, which were then grouped into four categories: (i) accurate and complete (AC), (ii) accurate but incomplete (AI), (iii) inaccurate (IN), and (iv) absent (AB). A composite MC score was calculated using the formula " $3 \times AC + 2 \times AI + 1 \times IN + 0 \times AB$ " per person per task. We anticipate a strong correlation between MC scores and global coherence because both measures assess the overall organization and clarity of discourse, but a medium to strong correlation between MC scores and local coherence because the two concepts are less closely aligned.

3. Results

3.1. Comparison of coherence scores between PWA and healthy control

Global and local coherence measures (see Figure 1) were each selected as dependent variables and fitted with a separate linear mixed-effects model using the *lme4* package (Bates et al., 2015) in R. Participant groups (healthy controls vs. PWA), tasks (Pic-Umbrella, Pic-Window, S-Crywolf, S-Tortoise) were used as fixed factors. Two models were built to test all main effects and interactions. For each model, the analysis was run with participants as random intercepts, using the Kenward-Roger degrees of freedom approximation (as suggested by Halekoh & Højsgaard, 2014) to estimate *p*-values for the fixed effects of interest (See Tables A.3 and A.4 for model details). Main effects for multilevel factors or

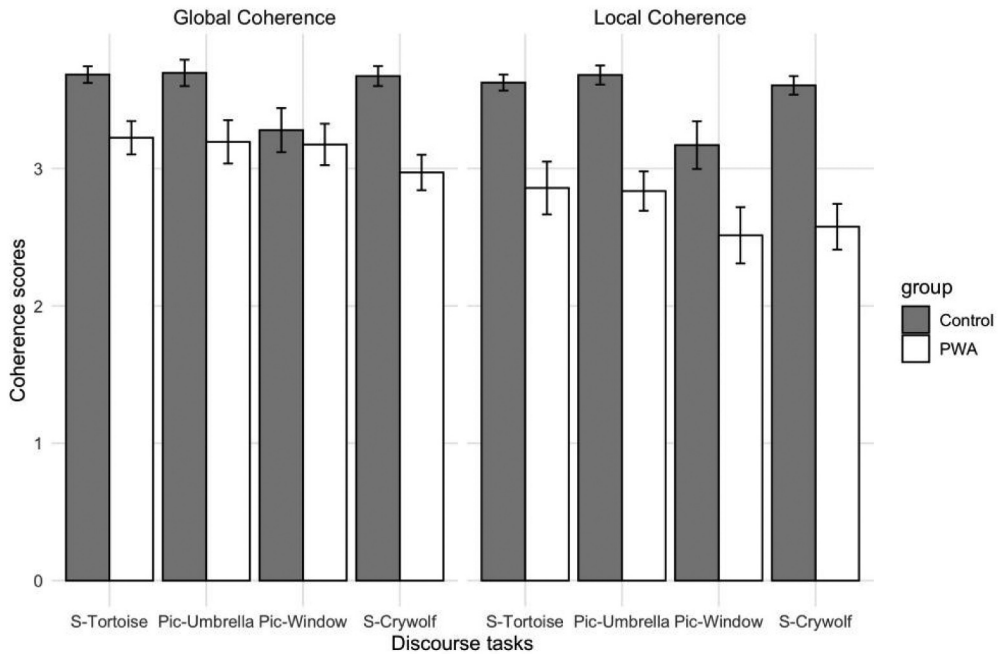


Figure 1. Coherence scores of different tasks by Mandarin healthy controls and PWA. Note: pic = picture description; S = story narrative.

interactions were reported using F-tests via the *Anova* function from the *car* package (Weisberg & Fox, 2010) in R, rather than comparisons with the baseline level using t-tests.

For global coherence, there were significant main effects of groups but no main effect of task types. In addition, there was a two-way interaction (group $\times$ task) (see Figure 1). The analysis revealed significant main effects of group ($F(1, 33.82) = 11.20, p < 0.001$), no main effects of task ($F(3, 99.61) = 2.69, p = 0.05$), and a significant interaction between task and group ($F(3, 99.63) = 3.93, p = 0.01$). To further examine the main effects and the interaction, we conducted pairwise multiple *t*-tests with Tukey adjustments. For the main effect of groups, healthy controls ($M = 3.58, 95\% \text{ CI } [3.39, 3.77]$) generally had higher scores in global coherence than PWA ($M = 3.13, 95\% \text{ CI } [2.94, 3.32]$; $t(33.8) = 3.39, p = 0.002$).

Post-hoc pairwise comparisons of the group by task interaction (see Table 5) showed that healthy controls had significantly better performance in global coherence than PWA in Pic-Umbrella ($M_{PWA} = 3.18, 95\% \text{ CI } [2.93, 3.43]$; $M_{Control} = 3.70, 95\% \text{ CI } [3.45, 3.94]$),

Table 5. Group differences for global coherence scores within each task.

Task	<i>df</i>	<i>t</i>	<i>p</i>
Pic-Umbrella	84.07	2.96	0.004*
Pic-Window	82.50	0.60	0.547
S-Crywolf	84.07	4.10	<0.001*
S-Tortoise	84.07	2.72	0.008*

Note: Pic = picture description; S = story narrative.

S-Crywolf ($M_{PWA} = 2.96$, 95% CI [2.71, 3.21]; $M_{Control} = 3.67$, 95% CI [3.43, 3.92]), and S-Tortoise ($M_{PWA} = 3.21$, 95%CI [2.96, 3.46]; $M_{Control} = 3.68$, 95% CI [3.44, 3.93]) tasks. However, their global coherence performances did not significantly differ in the Pic-Window task ($M_{PWA} = 3.18$, 95% CI [2.93, 3.42]; $M_{Control} = 3.28$, 95% CI [3.04, 3.52]).

Within each group we examined the differences in global coherence scores between tasks (see Table 6). For the healthy controls, their global coherence scores were significantly lower in Pic-Window ($M = 3.28$, 95% CI [3.04, 3.52]) than in S-Tortoise ($M = 3.68$, 95% CI [3.44, 3.93]), Pic-Umbrella ($M = 3.70$, 95% CI [3.45, 3.94]), and S-Crywolf ($M = 3.67$, 95% CI [3.43, 3.92]). In contrast, for the PWA group, none of the task comparisons were statistically significant. The results indicate that task-related differences in global coherence were evident in healthy control but not observed in the PWA group. See Table A.5 in the online supplementary materials for means and confidence intervals of all global and local coherence scores by each group and task.

For local coherence, we found a significant main effect of group ($F(1, 33.83) = 27.23$, $p < 0.001$) and of task ($F(3, 99.59) = 6.71$, $p < 0.001$) but no interaction between task and group ($F(3, 99.61) = 1.16$, $p = 0.33$). For the main effect of group, the control group ($M = 3.52$, 95% CI [3.29, 3.75]) performed significantly better in local coherence than PWA ($M = 2.69$, 95% CI [2.46, 2.92]; $t(33.8) = 5.24$, $p < 0.0001$) in all tasks (see Table 7). For the main effect of task, participants generally received lower local coherence scores in Pic-Window ($M = 2.84$, 95% CI [2.64, 3.04]) than in S-Tortoise ($M = 3.24$, 95% CI [3.04, 3.44]; $t(100.1) = -3.78$, $p = 0.002$), Pic-Umbrella ($M = 3.26$, 95% CI [3.05, 3.46]; $t(100.1) = -3.94$, $p < 0.001$), and in S-Crywolf ($M = 3.09$, 95% CI [2.88, 3.29]; $t(100.1) = -2.34$, $p = 0.096$). This indicated that participants performed significantly worse in Pic-Window task compared with S-Tortoise and Pic-Umbrella tasks. Comparisons between other tasks were not significant.

Table 6. Task differences for global coherence scores within each group.

Group	Task comparisons		<i>df</i>	<i>t</i>	<i>p</i>
Control	S-Tortoise	Pic-Umbrella	99.17	−0.10	1.000
	S-Tortoise	Pic-Window	99.17	3.17	0.011*
	S-Tortoise	S-Crywolf	99.17	0.08	1.000
	Pic-Umbrella	Pic-Window	99.17	3.26	0.008*
	Pic-Umbrella	S-Crywolf	99.17	0.18	0.998
	Pic-Window	S-Crywolf	99.17	−3.09	0.014*
PWA	S-Tortoise	Pic-Umbrella	99.17	0.23	0.996
	S-Tortoise	Pic-Window	100.98	0.26	0.994
	S-Tortoise	S-Crywolf	99.17	1.93	0.224
	Pic-Umbrella	Pic-Window	100.98	0.03	1.000
	Pic-Umbrella	S-Crywolf	99.17	1.70	0.331
	Pic-Window	S-Crywolf	100.98	1.68	0.339

Note: Pic = picture description; S = story narrative.

Table 7. Group differences for local coherence scores within each task.

Task	<i>df</i>	<i>t</i>	<i>p</i>
Pic-Umbrella	81.22	4.17	< 0.001*
Pic-Window	79.64	3.25	0.002*
S-Crywolf	81.22	5.07	< 0.001*
S-Tortoise	81.22	3.79	< 0.001*

Note: Pic = picture description; S = story narrative.

We also compared global and local coherence scores in the two groups by fitting a linear mixed-effects model. Coherence score was taken as the dependent variable, while group and coherence types were taken as fixed factors, and participants as random intercept. We found significant main effects of group, ($F(1, 33.94) = 20.89, p < 0.001$) and of coherence types ($F(1, 244.08) = 25.16, p < 0.001$). Importantly, there was a significant interaction between group and coherence types ($F(1, 244.08) = 14.35, p < 0.001$). Within each group we examined the differences between global and local coherence scores. There was no significant differences between global ($M = 3.58, 95\% \text{ CI } [3.37, 3.80]$) and local ($M = 3.52, 95\% \text{ CI } [3.31, 3.73]$) coherence scores in healthy controls ($t(244) = 0.88, p = 0.38$). PWA group, by contrast, showed significantly lower scores in local coherence ($M = 2.68, 95\% \text{ CI } [2.47, 2.90]$) than in global coherence ($M = 3.13, 95\% \text{ CI } [2.91, 3.34]; t(244) = 6.16, p < 0.0001$), suggesting that PWA had more difficulties in maintaining local coherence than in global coherence.

3.2. Correlations of coherence score with other test results

Spearman's correlation tests were conducted between global and local coherence scores and Aphasia Quotients (AQ in WAB-R), lexical diversity R, CIU, and Main Concept scores (see Table 8) for each task and each group. Correlation strength was decided based on Spearman's r values ($r > 0.70$, very strong; $0.40\text{--}0.69$, strong; $0.30\text{--}0.39$, moderate; $0.20\text{--}0.29$, weak; $0.01\text{--}0.19$, no or negligible), following Dancey and Reidy (2017).

Significant correlations between AQ scores and both measures of coherence were observed across three tasks (see Table 9), suggesting that both global and local coherence are effective indicators for estimating the severity of aphasia.

A significant correlation between global coherence and lexical diversity R was observed in the S-Tortoise task, but not in the other tasks (see Table 9). Similarly, local coherence showed significant correlations with R in the S-Crywolf and S-Tortoise tasks but not in others. In picture description tasks, both types of coherence scores were only moderately or weakly related with lexical diversity. These findings suggest that coherence have less consistent degree of correlations with lexical diversity across different tasks.

To examine the relationship between coherence and informativeness, we conducted Spearman's correlations between coherence scores and percent CIU. Results indicate significant and strong correlations between both types of coherence and percent CIU in all tasks (see Table 9), suggesting that coherence can reflect informativeness of discourse.

Table 8. Lexical diversity R, percent CIU, and MC scores of two groups in each task.

Tasks	Groups	R		% CIU		MC Scores	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Pic-Umbrella	Control	5.30	0.72	0.84	0.13	19.56	5.11
	PWA	4.19	0.79	0.69	0.25	10.71	5.49
Pic-Window	Control	5.20	0.44	0.74	0.22	10.00	4.20
	PWA	3.79	0.78	0.65	0.22	5.56	3.50
S-Crywolf	Control	5.75	0.81	0.82	0.11	18.72	5.52
	PWA	4.46	1.10	0.59	0.29	7.06	5.29
S-Tortoise	Control	6.52	0.88	0.83	0.11	18.28	4.73
	PWA	4.08	1.05	0.58	0.28	9.53	4.26

Note: R = Guiraud's R; % CIU = percent Correct Information Unit; MC Score = Main Concept score; Pic = picture description; S = story narrative.

Table 9. Correlations of coherence scores with lexical diversity R, percent CIU, and main Concept scores for PWA data.

Types	Tasks	AQ		R		% CIU		MC Scores	
		<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Global	Pic-Umbrella	0.70	0.002*	0.04	0.885	0.8	< 0.001*	0.50	0.042*
	Pic-Window	0.71	0.001*	0.20	0.428	0.59	0.010*	0.74	< 0.001*
	S-Crywolf	0.48	0.052	0.40	0.108	0.57	0.017*	0.79	< 0.001*
	S-Tortoise	0.74	0.001*	0.53	0.030*	0.84	< 0.001*	0.65	0.004*
Local	Pic-Umbrella	0.76	< 0.001*	0.26	0.316	0.68	0.003*	0.63	0.007*
	Pic-Window	0.58	0.012*	0.34	0.162	0.47	0.047*	0.82	< 0.001*
	S-Crywolf	0.53	0.028*	0.56	0.020*	0.78	< 0.001*	0.71	0.001*
	S-Tortoise	0.73	0.001*	0.69	0.002*	0.84	< 0.001*	0.67	0.003*

AQ = aphasia quotient; R = Guiraud’s R; % CIU = percent Correct Information Unit; MC Score = Main Concept score; Pic = picture description; S = story narrative.

Additionally, to investigate the relationship between coherence and the relevance to core content of a discourse task, we conducted correlation analyses between coherence scores and MC scores for each task. The results revealed significant and strong correlations between both coherence measures and MC scores across all discourse tasks (see Table 9), suggesting that both global and local coherence are strongly linked to the completeness and accuracy of main concepts in aphasic discourse.

4. Discussion

The current study investigated global and local coherence of discourse produced by Mandarin PWA and control group. PWA discourse received lower global and local coherence scores than that of the control group. A significant group by task interaction was found for global coherence, while a significant main effect for task was found for local coherence. In addition, global and local coherence were significantly correlated with aphasia severity and discourse informativeness in most tasks, but only significantly correlated with lexical diversity in story narratives.

First, discourse produced by Mandarin PWA was generally less coherent than that by the control group in both global and local coherence measures. Our findings are consistent with previous studies on English (Coelho & Flewellyn, 2003; Olness & Ulatowska, 2011; Wright & Capilouto, 2012) and Cantonese PWA discourse (Kong et al., 2018, 2020), suggesting that PWA experienced greater difficulties than the control group in organizing content around the central topic and maintaining effective connections of adjacent utterances within a discourse. Specifically, for global coherence, the discourse produced by PWA involved frequent inclusion of tangential information or non-essential elaborations that required the listeners’ efforts to infer. For local coherence, PWA discourse was characterized by ambiguous or doubtful relations between contiguous concepts, rendering confusions in listeners. It should be noted that our findings differ from some other studies (e.g., Glosser & Deser’s, 1990; Van Leer & Turkstra’s, 1999) that found no group differences. In these studies, PWA were asked to recount their personal experiences, e.g., family, work, and trauma stories, which lacked a consistent goal or theme for both groups, leading to minimized differences between the two groups (Wright & Capilouto, 2012). In addition, PWA seemed to show more deficits in local coherence than in global coherence, consistent with previous studies (Van Leer & Turkstra, 1999). This may be a result of the

different cognitive demands of global and local coherence (Rogalski et al., 2010). The relative preservation of conceptual knowledge and planning abilities may partially preserve global coherence, allowing PWA to maintain some topical focus. However, local coherence requires moment-to-moment integration of adjacent linguistic and conceptual information, which may place higher demands on attention and working memory.

Second, we found a main effect of task on local coherence, with significantly lower scores in the Pic-Window task compared with other tasks, indicating that the Pic-Window task was more challenging for participants in maintaining local coherence. These findings supported the previous studies arguing that elicitation tasks and stimuli can affect discourse production since they posed varying linguistic and cognitive demands (Leaman & Edmonds, 2023; Van Leer & Turkstra, 1999). Several factors could contribute to the low local coherence of the Broken Window task. First, topic familiarity could facilitate the clear elaboration, as pre-existing knowledge can be conceptualized as a cognitive schema that is activated when facing a more familiar topic (E. C. Li et al., 1995). The story narrative tasks are relatively familiar to the participants, whereas the content of stimuli in the picture description tasks is new to them. The prior knowledge about the stories could facilitate the instantaneous organization of discourse structures, enhancing the coherence between consecutive sentences. On the other hand, the unfamiliar pictures required additional interpretation during discourse organisation. In other words, difficulties in interpreting the pictures might interfere with generating the adjacent sentences, which hampered the connections between adjacent utterances in discourse production. This was particularly evident with certain pictures, such as those in the Broken Window task, where even the controls received relatively low global and local coherence scores and the differences between PWA and controls became non-significant in terms of global coherence.

Related to this, the significant group by task interaction for global coherence suggests that the control group displayed variations in global coherence scores across different tasks whereas PWA demonstrated consistently low scores across tasks. Global coherence processing involves accessing semantic representations of real-world facts, concepts, and relationships, as well as online integration and organization of ideas. These processes place high demands on cognitive functions such as attention, working memory, and concentration (Rogalski et al., 2010). Given that PWA exhibit cognitive disruptions in these cognitive domains, their ability to maintain global coherence remained uniformly low, regardless of task difficulty.

Third, both global and local coherence scores of Mandarin PWA discourse were positively correlated with their AQs in all tasks, partially consistent with a previous Cantonese aphasia study (Kong et al., 2020), suggesting that coherence measures can effectively reflect aphasia severity in Chinese population. However, Kong et al. (2020) did not find correlation between global coherence and AQ, which may be attributed to the smaller AQ variation in their aphasia group ($M_{AQ} = 87.3$, $SD = 9.22$) than that in the current study ($M_{AQ} = 78.4$, $SD = 12$). A lower level of heterogeneity of the sample may obscure the potential correlation between these factors.

Fourth, there existed correlations between global and local coherence and lexical diversity, though with variations across different tasks. Coherence pertains to the logical structure and meaning of a discourse (i.e., ideas are well-connected and easy to follow), whereas lexical diversity refers to the richness and variety of vocabulary used. These two

aspects of discourse can function independently: a discourse can be highly coherent using a limited set of words, or can exhibit high lexical diversity but lack logical connections between ideas (Coelho & Flewellyn, 2003). We speculate this dynamic can be modulated by task types. The strong correlations in story narratives, as in line with previous studies (Andreetta & Marini, 2015; Marini et al., 2011; Wright & Capilouto, 2012), reflected the role of lexical selection in producing a well-formed and smoothly flowing narrative. For PWA, difficulties in lexical retrieval led to compensatory strategies such as repetitions and nonsensical fillers that directly disrupt global and local coherence (Marini & Urgesi, 2012). Compared with story narratives, the correlations between lexical diversity and coherence were weaker in picture description tasks. Both *Broken Window* and *Refused Umbrella* present visually explicit elements and typically involve fewer narrative components, requiring speakers to describe a static scene rather than construct a dynamic narrative. These tasks can be accomplished with a more limited range of lexical items while still maintaining structural coherence, as speakers can effectively communicate by using basic vocabulary to denote the main characters, objects, and spatial relationships depicted in the images (Fergadiotis & Wright, 2011). Furthermore, since picture descriptions allow a more straightforward listing of visible elements without complex causal or temporal transitions, speakers can produce coherent descriptions even with restricted vocabulary, further weakening the relationship between lexical diversity and discourse coherence. This task-dependent variation reflects the fundamental differences in cognitive-linguistic demands: story narratives require sophisticated vocabulary for complex plot development, while picture descriptions can achieve coherence through simpler lexical choices focused on visible elements.

Fifth, global and local coherence were significantly correlated with discourse informativeness, as measured by percent CIUs and MC scores, for all tasks, suggesting that coherence scores can reflect the informativeness of PWA discourse. This is in line with the previous studies reporting significant relations between coherence and informativeness (Carlomagno et al., 2011; Hazamy & Obermeyer, 2020; Kong et al., 2018; Wright & Capilouto, 2012) and main events sequencing (Kong et al., 2020; Ulatowska et al., 2013). Considering this, we argue that coherence disruptions may be largely attributed to decreased content and incomplete inclusions of events rather than reduced lexical diversity. To be more specific, PWA tended to produce unintelligible, inaccurate, and irrelevant information (as measured in CIUs, e.g., comments on task performance and the task itself, or reiterating points; Hazamy & Obermeyer, 2020). Such tangential information hampered the connection of one utterance to the central themes and to the adjacent utterance in the context, thus influencing global and local coherence. In addition, PWA were inclined to produce incorrect main concepts (e.g., the essential characters, locations, and objects) or skipped them in their discourse, disrupting logical flows and reducing both local and global coherence.

The present study provides valuable insights into coherence in Mandarin PWA discourse. In clinical settings, coherence rating provides a reliable assessment of discourse organization without requiring labour-intensive transcription procedures. The systematic application of these rating scales may also facilitate efficient monitoring of therapeutic progress and enables evidence-based treatment planning that targets higher-level deficits in discourse functionality, particularly in cases when basic linguistic competencies have been restored. This methodological efficiency enhances clinical practicality for

routine assessment protocols, especially when time and resources are limited. However, some limitations warrant consideration. First, the study cohort primarily comprised PWA with anomic aphasia, with all cases classified as mild to moderate in severity, i.e., PWA with AQ above 50. Future research should explore potential differences in discourse coherence across various aphasia types and severity to gain a more comprehensive understanding. Additionally, longitudinal studies are needed to investigate how discourse coherence changes throughout the course of rehabilitation, shedding light on the long-term effects of therapeutic interventions.

5. Conclusion

The present study examined global and local coherence in Mandarin PWA discourse production. The findings indicate that PWA experienced significant difficulties in maintaining both global and local coherence relative to the control group. Task-induced variations differed between global and local coherence. Task type affected global coherence of discourse production by the control group, whereas the PWA group exhibited consistent difficulties in maintaining global coherence across all tasks. Local coherence was particularly compromised for the Broken Window picture description task. Moreover, global and local coherence were significantly correlated with aphasia quotient, Correct Information Units (CIUs), and Main Concept scores across all tasks, suggesting that coherence measures can serve as indicators of aphasia severity and discourse informativeness. These findings underscore the importance of incorporating discourse-level assessments into aphasia diagnosis and rehabilitation, enhancing the evaluation and treatment of communication impairments in individuals with aphasia.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Project of Humanities and Social Sciences Foundation of the Ministry of Education of PRC [23YJC740002].

ORCID

Yi Hu  <http://orcid.org/0009-0001-1206-7937>

Juqiang Chen  <http://orcid.org/0000-0001-8151-3964>

References

- Andreetta, S., Cantagallo, A., & Marini, A. (2012). Narrative discourse in anomic aphasia. *Neuropsychologia*, 50(8), 1787–1793. <https://doi.org/10.1016/j.neuropsychologia.2012.04.003>
- Andreetta, S., & Marini, A. (2015). The effect of lexical deficits on narrative disturbances in fluent aphasia. *Aphasiology*, 29(6), 705–723. <https://doi.org/10.1080/02687038.2014.979394>
- Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology*, 14(9), 875–892. <https://doi.org/10.1080/02687030050127685>

- Barker, M. S., Young, B., & Robinson, G. A. (2017). Cohesive and coherent connected speech deficits in mild stroke. *Brain and Language*, 168, 23–36. <https://doi.org/10.1016/j.bandl.2017.01.004>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Blomert, L., Kean, M. L., Koster, C., & Schokker, J. (1994). Amsterdam-Nijmegen everyday language test: Construction, reliability and validity. *Aphasiology*, 8(4), 381–407. <https://doi.org/10.1080/02687039408248666>
- Carlomagno, S., Giannotti, S., Vorano, L., & Marini, A. (2011). Discourse information content in non-aphasic adults with brain injury: A pilot study. *Brain Injury*, 25(10), 1010–1018. <https://doi.org/10.3109/02699052.2011.605097>
- Chen, J., & Chang, H. (2024). Core lexicon analysis of spoken discourse production by Mandarin Chinese speakers with aphasia. *Aphasiology*, 38(11), 1766–1785. <https://doi.org/10.1080/02687038.2024.2325169>
- Chen, J., & Chang, H. (2025). Lexical network properties in aphasic discourse production. *Clinical Linguistics and Phonetics*, 1–19. <https://doi.org/10.1080/02699206.2025.2507049>
- Chen, Z.-Z., Lin, F., Deng, B.-M., Zhang, Q., Wu, Y.-C., Zhou, L., & Jiang, Z.-L. (2018). Construction of Chinese Mandarin corpus with anomic aphasia. *Chinese Journal of Rehabilitation Medicine*, 33(6), 669–674. <https://doi.org/10.3969/j.issn.1001-1242.2018.06.010>
- Chow, W., Kong, A. P. H., & Lau, K.-Y. D. (2016). An investigation of global and local coherence of spontaneous personal versus descriptive narratives in native Chinese speakers with traumatic brain injury: Preliminary data. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/conf.fpsyg.2016.68.00018>
- Christiansen, J. A. (1995). Coherence violations and propositional usage in the narratives of fluent aphasics. *Brain and Language*, 51(2), 291–317. <https://doi.org/10.1006/brln.1995.1062>
- Coelho, C., & Flewellyn, L. (2003). Longitudinal assessment of coherence in an adult with fluent aphasia: A follow-up study. *Aphasiology*, 17(2), 173–182. <https://doi.org/10.1080/729255216>
- Dancey, C., & Reidy, J. (2017). *Statistics without maths for psychology* (7th ed.). Pearson.
- Davidson, M. (2023). Known-groups validity. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 3764–3764). Springer. https://doi.org/10.1007/978-3-031-17299-1_1581
- Deng, B.-M., Gao, J., Liang, L.-S., Zhao, J.-X., Lin, F., Yin, M.-Y., Zheng, H.-Q., & Hu, X.-Q. (2024). Discourse task type-specific linguistic characteristics in anomic aphasia and healthy controls: Evidence from Mandarin-Chinese AphasiaBank. *American Journal of Speech-Language Pathology*, 33(2), 937–951. https://doi.org/10.1044/2023_AJSLP-23-00078
- Deng, B.-M., Liang, L.-S., Zhao, J.-X., Zheng, H.-Q., & Hu, X.-Q. (2024). Correct information unit analysis in different discourse tasks among persons with anomic aphasia based on Mandarin AphasiaBank. *American Journal of Speech-Language Pathology*, 33(2), 800–813. https://doi.org/10.1044/2023_AJSLP-23-00217
- Eignor, D. R. (2013). The standards for educational and psychological testing. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology* (Vol. 1, pp. 245–250). American Psychological Association. <https://doi.org/10.1037/14047-013>
- Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology*, 25(11), 1414–1430. <https://doi.org/10.1080/02687038.2011.603898>
- Glosser, G., & Deser, T. (1990). Patterns of discourse production among neurological patients with fluent language disorders. *Brain and Language*, 40(1), 67–88. [https://doi.org/10.1016/0093-934X\(91\)90117-J](https://doi.org/10.1016/0093-934X(91)90117-J)
- Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models-the R package pbrtest. *Journal of Statistical Software*, 59(9). <https://doi.org/10.18637/jss.v059.i09>
- Halliday, M. A. K., & Hasan, R. (2014). *Cohesion in English*. Routledge. <https://doi.org/10.4324/9781315836010>

- Hazamy, A. A., & Obermeyer, J. (2020). Evaluating informative content and global coherence in fluent and non-fluent aphasia. *International Journal of Language & Communication Disorders*, 55(1), 110–120. <https://doi.org/10.1111/1460-6984.12507>
- Herdan, G., & Guiraud, P. (1961). Problemes et methodes de la statistique linguistique. *Language*, 37(1), 120. <https://doi.org/10.2307/411255>
- Kertesz, A. (2012). *Western aphasia Battery-revised [dataset]*. <https://doi.org/10.1037/t15168-000>
- Kong, A. P.-H., Lau, D. K.-Y., & Cheng, C.-Y.-Y. (2020). Analysing coherence of oral discourse among Cantonese speakers in mainland China with traumatic brain injury and cerebrovascular accident. *International Journal of Speech-Language Pathology*, 22(1), 37–47. <https://doi.org/10.1080/17549507.2019.1581256>
- Kong, A. P.-H., Lau, D. K.-Y., & Lai, D. H.-Y. (2023). Measuring pragmatic competence of discourse output among Chinese-speaking individuals with traumatic brain injury. *Brain Impairment: A Multidisciplinary Journal of the Australian Society for the Study of Brain Impairment*, 24(3), 660–678. <https://doi.org/10.1071/IB23048>
- Kong, A. P.-H., Linnik, A., Law, S.-P., & Shum, W. W.-M. (2018). Measuring discourse coherence in anomic aphasia using rhetorical structure theory. *International Journal of Speech-Language Pathology*, 20(4), 406–421. <https://doi.org/10.1080/17549507.2017.1293158>
- 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. <https://doi.org/10.3109/14015439.2015.1041551>
- Krippendorff, K. (2019). *Content analysis: An introduction to its methodology*. SAGE. <https://doi.org/10.4135/9781071878781>
- Leaman, M. C., & Edmonds, L. A. (2021). Measuring global coherence in people with aphasia during unstructured conversation. *American Journal of Speech-Language Pathology*, 30(1S), 359–375. https://doi.org/10.1044/2020_AJSLP-19-00104
- Leaman, M. C., & Edmonds, L. A. (2023). Analyzing language in the picnic scene picture and in conversation: The type of discourse sample we choose influences findings in people with aphasia. *American Journal of Speech-Language Pathology*, 32(4), 1413–1430. https://doi.org/10.1044/2023_AJSLP-22-00279
- Li, C. N., & Thompson, S. (1989). *Mandarin Chinese: A functional reference*. University of California Press. <https://www.semanticscholar.org/paper/Mandarin-Chinese%3A-A-Functional-Reference-Grammar-Li-Thompson/cf37fc17c2364837b0b2b3185c358e62a825f9cd>
- Li, E. C., Williams, S. E., & Volpe, A. D. (1995). The effects of topic and listener familiarity on discourse variables in procedural and narrative discourse tasks. *Journal of Communication Disorders*, 28(1), 39–55. [https://doi.org/10.1016/0021-9924\(95\)91023-Z](https://doi.org/10.1016/0021-9924(95)91023-Z)
- Linnik, A., Bastiaanse, R., & Höhle, B. (2016). Discourse production in aphasia: A current review of theoretical and methodological challenges. *Aphasiology*, 30(7), 765–800. <https://doi.org/10.1080/02687038.2015.1113489>
- MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, 25(11), 1286–1307. <https://doi.org/10.1080/02687038.2011.589893>
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text - Interdisciplinary Journal for the Study of Discourse*, 8(3), 243–281. <https://doi.org/10.1515/text.1.1988.8.3.243>
- Marini, A., Andreetta, S., Del Tin, S., & Carlomagno, S. (2011). A multi-level approach to the analysis of narrative language in aphasia. *Aphasiology*, 25(11), 1372–1392. <https://doi.org/10.1080/02687038.2011.584690>
- Marini, A., & Urgesi, C. (2012). Please get to the point! A cortical correlate of linguistic informativeness. *Journal of Cognitive Neuroscience*, 24(11), 2211–2222. https://doi.org/10.1162/jocn_a_00283
- McNeil, M. R., Doyle, P. J., Fossett, T. R. D., Park, G. H., & Goda, A. J. (2001). Reliability and concurrent validity of the information unit scoring metric for the story retelling procedure. *Aphasiology*, 15(10–11), 991–1006. <https://doi.org/10.1080/02687040143000348>

- Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech, Language, & Hearing Research*, 36(2), 338–350. <https://doi.org/10.1044/jshr.3602.338>
- 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, & Hearing Research*, 38(1), 145–156. <https://doi.org/10.1044/jshr.3801.145>
- Olness, G. S., & Ulatowska, H. K. (2011). Personal narratives in aphasia: Coherence in the context of use. *Aphasiology*, 25(11), 1393–1413. <https://doi.org/10.1080/02687038.2011.599365>
- 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. <https://doi.org/10.1111/1460-6984.12318>
- Rogalski, Y., Altmann, L. J. P., Plummer D'Amato, P., Behrman, A. L., & Marsiske, M. (2010). Discourse coherence and cognition after stroke: A dual task study. *Journal of Communication Disorders*, 43(3), 212–224. <https://doi.org/10.1016/j.jcomdis.2010.02.001>
- Ulatowska, H. K., Reyes, B., Santos, T. O., Garst, D., Vernon, J., & McArthur, J. (2013). Personal narratives in aphasia: Understanding narrative competence. *Topics in Stroke Rehabilitation*, 20(1), 36–43. <https://doi.org/10.1310/tsr2001-36>
- Van Leer, E., & Turkstra, L. (1999). The effect of elicitation task on discourse coherence and cohesion in adolescents with brain injury. *Journal of Communication Disorders*, 32(5), 327–349. [https://doi.org/10.1016/S0021-9924\(99\)00008-8](https://doi.org/10.1016/S0021-9924(99)00008-8)
- Weisberg, S., & Fox, J. (2010, November 29). *An R companion to applied regression*. <https://www.semanticscholar.org/paper/An-R-Companion-to-Applied-Regression-Weisberg-Fox/2fc546a340101eb20d834da96c827d7e83090129>
- Wright, H. H., & Capilouto, G. J. (2012). Considering a multi-level approach to understanding maintenance of global coherence in adults with aphasia. *Aphasiology*, 26(5), 656–672. <https://doi.org/10.1080/02687038.2012.676855>
- Wright, H. H., Capilouto, G. J., & Koutsoftas, A. (2013). Evaluating measures of global coherence ability in stories in adults. *International Journal of Language & Communication Disorders*, 48(3), 249–256. <https://doi.org/10.1111/1460-6984.12000>
- Wright, H. H., Koutsoftas, A., Capilouto, G., & Fergadiotis, G. (2013). Global coherence in younger and older adults: Influence of cognitive processes and discourse type. *Neuropsychology, Development, and Cognition*, 21(2), 174–196. <https://doi.org/10.1080/13825585.2013.794894>
- Wright, H. H., Koutsoftas, A., Fergadiotis, G., & Capilouto, G. (2010). Coherence in stories told by adults with aphasia. *Procedia-Social and Behavioral Sciences*, 6, 111–112. <https://doi.org/10.1016/j.sbspro.2010.08.056>
- Zhang, M., Geng, L., Yang, Y., & Ding, H. (2021). Cohesion in the discourse of people with post-stroke aphasia. *Clinical Linguistics and Phonetics*, 35(1), 2–18. <https://doi.org/10.1080/02699206.2020.1734864>