Stimulus Complexity and Retelling Output in Aphasia: An Exploratory Single-Case Analysis

Hyunsoo Yoo[1*]

[1]Baylor University, Waco, United States

Submitted to Journal:
Frontiers in Human Neuroscience

Specialty Section:
Speech and Language

Article type:
Original Research Article

Manuscript ID:
1679562

Received on:
05 Sep 2025

Journal website link:
www.frontiersin.org

Scope Statement

This manuscript aligns with the scope of Frontiers in Psychology – Language Sciences by addressing the structural properties of narrative stimuli used in aphasia therapy and their influence on discourse-level language production. The study introduces a quantifiable profiling method for analyzing stimulus complexity and explores its relevance to treatment responsiveness in a single-case design. The work integrates theoretical models of discourse processing with clinical application, contributing to the understanding of language rehabilitation at the intersection of psycholinguistics, cognitive processing, and therapy design.

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Credit Author Statement

Hyunsoo Yoo: Writing – original draft, Writing – review & editing.

Keywords

Aphasia, story retelling, Narrative discourse, Structural complexity, discourse profiling

Abstract

Word count: 246

Discourse-based aphasia therapies often rely on narrative stimuli, yet the structural characteristics of these materials remain largely unexamined. While stimulus complexity may influence language production and cognitive load, few studies have proposed quantifiable methods for profiling discourse-level structure. This study aimed to develop and apply a replicable structural profiling method for narrative stimuli used in story-retelling therapy. A secondary aim explored the applicability of the profiling framework to narrative output in a single-case exploratory analysis. Twelve stimuli were analyzed using a Python-based pipeline to extract sentence count, average sentence length, complex sentence ratio, and estimated information units. Composite scores were calculated using z-score normalization and used to classify stimuli into Low, Medium, and High complexity groups. IU (Information Unit) gains were measured from pre-and post-treatment retellings of nine treated stimuli and three untrained evaluation stimuli. Stimuli varied in structural complexity. Some high-complexity stimuli yielded substantial IU gains when supported with treatment, though this pattern was not consistent. In contrast, similarly complex stimuli used only for evaluation (e.g., Loan) showed reduced or negative gains. Exploratory correlations showed a weak positive trend between complexity and IU gain for treated stimuli ($r = +0.13$), but this should be interpreted cautiously given the limited sample. This study introduces a quantifiable method for structural profiling of narrative stimuli and demonstrates its feasibility in treatment design. Structural complexity may function as a modifiable input variable for stimulus calibration in discourse-based aphasia therapy, with potential applications for tailoring task demands to individual readiness.

Funding information

This work was supported by start-up research funds provided by Baylor University.

Funding statement

The author(s) declare that financial support was received for the research and/or publication of this article.

Ethics statements

Studies involving animal subjects

Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects

Generated Statement: The studies involving humans were approved by This study involving human participants was reviewed and approved by the Institutional Review Board at Baylor University (IRB# 2219727-9). Written informed consent was obtained from the participant prior to participation in accordance with institutional guidelines and the Declaration of Helsinki.. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Inclusion of identifiable human data

Generated Statement: No potentially identifiable images or data are presented in this study.

Data availability statement

Generated Statement: The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Generative AI disclosure

The author(s) verify and take full responsibility for the use of generative AI in the preparation of this manuscript. Generative AI was used

ChatGPT was used for language refinement in the early drafts. All content was verified and finalized by the author.

Stimulus Complexity and Retelling Output in Aphasia: An Exploratory Single-Case Analysis

Hyunsoo Yoo

Department of Communication Science and Disorders

Baylor University, Waco, TX

Correspondence to:

Hyunsoo Yoo*

Department of Communication Science and Disorders

Robbins College of Health and Human Sciences

Baylor University, Waco, TX 76798

Email: Hyunsoo_Yoo@Baylor.edu.

Abstract

Discourse-based aphasia therapies often rely on narrative stimuli, yet the structural characteristics of these materials remain largely unexamined. While stimulus complexity may influence language production and cognitive load, few studies have proposed quantifiable methods for profiling discourse-level structure. This study aimed to develop and apply a replicable structural profiling method for narrative stimuli used in story-retelling therapy. A secondary aim explored the applicability of the profiling framework to narrative output in a single-case exploratory analysis. Twelve stimuli were analyzed using a Python-based pipeline to extract sentence count, average sentence length, complex sentence ratio, and estimated information units. Composite scores were calculated using z-score normalization and used to classify stimuli into Low, Medium, and High complexity groups. IU (Information Unit) gains were measured from pre- and post-treatment retellings of nine treated stimuli and three untrained evaluation stimuli. Stimuli varied in structural complexity. Some high-complexity stimuli yielded substantial IU gains when supported with treatment, though this pattern was not consistent. In contrast, similarly complex stimuli used only for evaluation (e.g., Loan) showed reduced or negative gains. Exploratory correlations showed a weak positive trend between complexity and IU gain for treated stimuli ($r = +0.13$), but this should be interpreted cautiously given the limited sample. This study introduces a quantifiable method for structural profiling of narrative stimuli and demonstrates its feasibility in treatment design. Structural complexity may function as a modifiable input variable for stimulus calibration in discourse-based aphasia therapy, with potential applications for tailoring task demands to individual readiness.

Keywords: aphasia, story retelling, narrative discourse, structural complexity, discourse profiling

Introduction

In conversational contexts, the ability to listen, comprehend, and rephrase a partner's message plays a central role in functional communication (Clark & Brennan, 1991; Glosser & Deser, 1992; Kagan & Simmons-Mackie, 2007). This process involves more than passive reception of information; it engages integrated cognitive-linguistic operations including listening, comprehension, reconstruction, and verbal output (Ulatowska, North, & Macaluso-Haynes, 1983). At the heart of this ability is story retelling, which supports essential discourse functions such as appropriate response, topic maintenance, and contextual alignment (Marini et al., 2005).

Unlike spontaneous narrative production, story retelling places heavy demands on verbal working memory, as it requires the temporary storage and reorganization of aurally presented linguistic input. Yoo and McNeil (2019) demonstrated that in individuals with aphasia, performance on story retelling tasks is significantly correlated with measures of verbal working memory (r = .50–.75). When story retelling abilities are impaired due to aphasia, individuals often struggle to follow conversational turns and maintain contextual coherence, which can negatively impact not only the efficiency of daily interaction but also broader aspects of social participation and relationship-building.

Story retelling is more than a sequence of sentence repetitions. It requires the integration of narrative structure, temporal sequencing, and causal coherence, placing complex demands on both linguistic and cognitive processing resources (Stein & Glenn, 1979; Trabasso & van den Broek, 1985). This task is often conceptualized as narrative retelling in the discourse literature, where it is regarded as a discourse-level function commonly used in both assessment and treatment contexts (Liles, 1993).

A range of studies has identified five categories of factors influencing story retelling performance. First, cognitive factors such as working memory capacity and processing speed have been shown to significantly influence discourse-level comprehension and organization. For example, See and Ryan (1995) demonstrated that individual differences in these cognitive capacities mediate age-related variations in discourse processing performance. From a theoretical perspective, discourse comprehension models (e.g., van Dijk & Kintsch, 1983; Kintsch, 1988) highlight the central role of working memory in supporting macrostructural integration, coherence building, and the real-time coordination of incoming and stored information across discourse segments.

Second, linguistic factors contribute to discourse production through several interrelated mechanisms. According to Levelt (1989), lexical retrieval involves a multi-stage process that includes conceptual preparation, lemma selection, and phonological encoding. When any part of this process is disrupted, speakers may experience breakdowns in fluency and lexical cohesion, which can negatively impact the continuity of narrative expression. Syntactic comprehension also plays an essential role, particularly when processing structurally complex sentences such as those with embedded clauses or noncanonical word orders. Gibson's (1998) Syntactic Prediction Locality Theory explains that increased syntactic complexity elevates both integration and memory costs, which can interfere with the real-time construction and maintenance of coherent discourse. Building on this account, expectation-based models of sentence processing suggest that the difficulty of comprehension is influenced by how well upcoming structures align with probabilistic syntactic expectations (Levy, 2008). The use of cohesive devices is equally important for sustaining both local and global coherence across utterances. Proper use of linguistic markers such as pronouns, conjunctions, and lexical ties helps maintain the semantic flow of discourse. In a study comparing children with and without language disorders, Liles (1993) reported that limited or inappropriate use of cohesive devices was associated with poorer narrative organization and reduced listener comprehension. Collectively, findings on lexical retrieval, syntactic comprehension, and cohesion highlight the critical role of linguistic processing in discourse production and demonstrate that deficits in any of these domains may compromise performance at the discourse level.

Third, discourse-level factors also play a critical role in narrative performance, particularly through schematic knowledge and causal event representation. Mandler and Johnson (1977) first demonstrated that narrative comprehension and recall are guided by underlying story grammar structures. Building on this foundation, Stein and Glenn (1979) proposed that a well-formed narrative is organized around an internalized story schema, typically consisting of elements such as setting, initiating event, goal, attempt, and outcome. This mental framework guides both comprehension and production by helping individuals structure events in a temporally and causally coherent sequence. Marini et al. (2005) found that individuals with language impairments often struggle to construct causal relationships between events, resulting in fragmented or poorly integrated narratives. These findings highlight the importance of macrostructural organization in discourse-level processing.

Fourth, stimulus-level characteristics have also been shown to influence language performance, particularly through structural complexity and modality of presentation. Thompson and Shapiro (2005), in their work on Treatment of Underlying Forms (TUF), demonstrated that training on syntactically complex sentences led to generalization to simpler, untrained structures. Although this work was conducted at the sentence level, it underscores a key principle: structurally complex input imposes greater cognitive load but can also foster broader learning and generalization. When complex sentences are used as treatment stimuli, they place a greater cognitive burden on the language processing system. Research within the Treatment of Underlying Forms (TUF) framework has shown that training on complex sentence structures leads to generalization to untrained, linguistically related simpler structures (Thompson & Shapiro, 2005). In other words, such stimuli not only improve performance on the trained complex forms but also facilitate better performance on simpler, untrained tasks, likely by strengthening integrative processing mechanisms. While not focused on discourse per se, these findings provide conceptual justification for considering stimulus complexity as an active design element in discourse-based aphasia therapy.

However, despite the growing recognition of stimulus-level effects, structural complexity has rarely been examined in discourse-based intervention. While structural complexity at the sentence level has been well studied, particularly through treatment approaches such as Treatment of Underlying Forms (TUF), which shows that training on complex syntax can promote generalization (Thompson & Shapiro, 2005), this principle has seldom been extended to the discourse level. Discourse-based therapies continue to rely heavily on narrative stimuli, and while clinicians may informally consider aspects of narrative structure such as temporal ordering and causal relationships, stimuli are typically selected based on topic familiarity, cultural relevance, or perceived appropriateness for the client (Dipper et al., 2020). Structural characteristics of the stimuli, including syntactic complexity or information density, are rarely quantified or systematically manipulated as experimental variables. This lack of quantification and calibration of discourse-level structure represents a methodological gap that limits both empirical validation and stimulus design in current aphasia therapy.

This framework builds on prior findings but shifts the analytic focus from output to input. To address this gap, the present study introduces a quantifiable method for structural profiling of narrative stimuli used in aphasia intervention. Using four linguistic indicators, sentence count, average sentence length, subordination ratio, and estimated information units, we applied a Python-based pipeline to analyze twelve story stimuli (Python Software Foundation, 2023). Composite complexity scores were calculated using z-score normalization and used to classify stimuli into low, medium, and high complexity tiers.

As a proof-of-concept application, we explored how this profiling framework might be used to characterize stimuli and inform stimulus selection for treatment planning. By shifting analytic focus from narrative output to stimulus input, this study offers a preliminary demonstration of data-informed stimulus profiling for potential use in discourse therapy.

Method

Design

This study employed a single-case pre-post design to (1) quantify the structural complexity of narrative stimuli and (2) examine the relationship between structural complexity and treatment responsiveness in story retelling therapy (SRT). The study consisted of two components: a structural analysis of 12 narrative stimuli and an intervention phase. Structural features including sentence count, average sentence length, subordination ratio, and estimated information units were extracted using a Python-based pipeline. *Z*-score normalization was applied, and composite structural complexity scores were calculated and used to group stimuli into Low, Medium, and High Load Blocks based on tertile distribution.

Participant

A single participant took part in this study: a right-handed, monolingual English-speaking woman, aged 70 at the time of participation. She experienced a left hemisphere stroke approximately 34 months prior and was diagnosed with aphasia by a neurologist in 2022. Her educational background included 17 years of formal schooling, and she had previously worked in a professional administrative role at a university before retirement. Based on Western Aphasia Battery-Revised (WAB-R) scores, she presented with mild anomic aphasia (AQ = 92.7). She did not exhibit signs of moderate or severe apraxia of speech. Eligibility was determined according to the following inclusion criteria: (1) diagnosis of mild to moderate aphasia; (2) stroke onset of more than six months prior; (3) age 18 or older; (4) right-handed premorbidly; (5) English as the primary language; (6) normal or corrected-to-normal vision; and (7) a minimum of 12 years of education. Exclusion criteria included global or severe aphasia, left-handedness prior to stroke, multiple or recurrent stroke events, current enrollment in similar therapy studies, and any co-occurring neurological conditions unrelated to stroke. All procedures were approved by the Institutional Review Board (IRB) at Baylor University, and informed consent was obtained prior to participation.

Procedure

The participant completed nine face-to-face story retelling treatment (SRT) sessions over a three-week period. Each session was conducted by the Principal Investigator (PI) and lasted approximately one hour. A total of 12 narrative stimuli were used: nine were assigned as treatment stimuli, and three were used exclusively as evaluation-only stimuli. Each treatment session consisted of an auditory presentation of a story stimulus, a pre-treatment retelling, a multimodal therapy phase (including repetition, reading, writing, sequencing, summarization, and guided retelling), and a post-treatment retelling. All retellings were based solely on auditory input; no visual cues were provided. Three formal assessments were conducted: a pre-test (Test 1), a post-test immediately after the final treatment session (Test 2), and a delayed post-test seven weeks later (Test 3). The delay between Test 2 and Test 3 was extended from four to seven weeks due to a fall experienced by the participant. Evaluation-only stimuli were administered at all three assessment timepoints to examine generalization and maintenance. Performance was measured using Information Units (IUs) produced in retellings.

Structural Feature Extraction

To quantify stimulus-level complexity, four structural features were analyzed for each of the twelve story stimuli. These structural features were selected not only for their theoretical relevance to syntactic and propositional complexity, but also for their practical utility in automated discourse analysis. Selection was guided by four key criteria:

(1) conceptual grounding in prior research on syntactic load and information density,

(2) extractability using standard Python-based string processing methods (e.g., str.split(), str.count(), re.search()),

(3) objectivity and consistency through rule-based implementation, and

(4) applicability across all 12 stimuli to support reliable comparison.

This approach supported that the resulting composite score would be both theoretically interpretable and computationally reproducible.

Stimulus Analysis

All computations were conducted in Python 3.10 (Python Software Foundation, 2023) using standard libraries including pandas (McKinney, 2010), numpy (Harris et al., 2020), and re (Python Software Foundation, 2023) for data processing, and matplotlib (Hunter, 2007) for visualization. Linguistic features were extracted using basic string handling methods in Python 3.10, including str.split() for word segmentation, len() for word count calculation, and str.count() for detecting conjunctions and punctuation-based information units. All subsequent calculations, normalization, and block assignment were performed using pandas, numpy, and re.

1. Sentence Count

Total number of syntactically complete sentences per stimulus, determined through manual segmentation based on independent clause boundaries rather than punctuation alone. Initial segmentation was estimated using regular expression-based punctuation splits, but final sentence counts were manually corrected to reflect functional sentence boundaries and clause completeness.

2. Complex Sentence Ratio

The proportion of sentences containing at least one subordinating conjunction (e.g., because, although, when, if, since, though, unless, while). Each sentence was flagged as complex if it included any of these conjunctions, detected using regular expression searches.

Complex Sentence Ratio=Total Sentence CountNumber of sentences with subordinating conjunctions /Total Sentence Count

3. Average Sentence Length

The mean number of words per sentence was calculated by dividing the total word count by the manually verified sentence count. Word counts were extracted using Python's string splitting and length functions.

Average Sentence Length= Total Word Count/Sentence Count

4. Estimated Information Units (IUs)

A proxy for informational density, estimated as the sum of all commas, occurrences of 'and', and occurrences of 'but' in each stimulus. These elements were treated as proxies for clause-level segmentation and cohesion markers.

Estimated IUs = count of commas + count of "and" + count of "but"

Raw values for each feature were z-score normalized to allow for cross-feature comparison. A composite structural load score was calculated for each stimulus by averaging the four z-scores. Based on tertile distribution of composite scores, stimuli were categorized into Low**,** Medium**,** or High Load Blocks to facilitate load-based comparisons of treatment response.

5. Normalization and Composite Score

Each raw feature was converted to a z-score where x is the raw score for a given stimulus, μ is the mean, and σ is the standard deviation across all stimuli:

$$z = x - \mu / \sigma$$

A Composite Structural Load Score was then computed for each stimulus as the mean of its four z scores of normalized features: sentence count, complex ratio, sentence length, and estimated IUs.

$$\text{Composite Score} = z\text{Sentence Count} + z\text{Complex Ratio} + z\text{Sentence Length} + z\text{IUs}/4$$

6. Block Assignment (Load Level)

Stimuli were categorized into Low, Medium, or High Load Blocks based on their Composite Structural Load Scores, using tertile distribution. Specifically, the top third of scores were classified as High Load, the middle third as Medium Load, and the bottom third as Low Load. This classification was implemented using the qcut() function from the Python pandas library.

Exploratory Application: IU Response Profiling

To examine the relationship between structural complexity and treatment response,
four structural features—sentence count, complex sentence ratio, average sentence length, and estimated information units—were quantified for each narrative stimulus and standardized using z-scores.
The mean of these z-scores was used to calculate a composite structural complexity score for each story.
IU (Information Unit) gain was computed as the difference in the number of IUs produced before and after treatment. A total of 12 narrative stimuli (n = 12) were included in the analysis: nine were used as active treatment stimuli, and three were used solely for evaluation to assess generalization and retention effects. The nine treated stories were presented within therapy sessions and evaluated at both pre- and post-treatment to assess direct treatment effects.
The three evaluation-only stories were not used in therapy but were retold at three time points: pre-treatment, immediately after completing all treatment sessions (Post1), and seven weeks post-treatment (Post2). These stimuli were used to assess the generalization and maintenance of treatment gains.
IU gains were calculated for each comparison interval (Post1–Pre, Post2–Pre, and Post2–Post1), and the

same structural profiling procedure was applied to these evaluation stories as for the treated ones. The relationship between stimulus complexity and IU gain was examined using Pearson correlation analysis and visualized using scatterplots with regression lines. This allowed comparison of structural responsiveness between treated and untreated stimuli. Given the limited sample size, the analysis was exploratory in nature and aimed to identify directional trends rather than statistical significance.
The correlation analysis was conducted using the scipy.stats.pearsonr() function in Python.

Results

Treatment and IU Measurement

Of the twelve narrative stimuli developed for this study, nine were implemented in structured story-retelling treatment sessions with a single participant diagnosed with mild anomic aphasia. The remaining three stimuli were designated as evaluation-only and were not used during treatment. During treatment sessions, each stimulus was presented auditorily, followed by a pre-treatment retelling. A multimodal therapy protocol was then conducted, consisting of repetition of words and phrases, reading, writing, picture-sequence ordering, sentence ordering, true/false decision-making, story summarization (beginning–middle–end), and retelling at progressively larger units. At the end of each session, a post-treatment retelling was elicited. All retellings were based solely on auditory input without visual aids. Retellings were transcribed and analyzed for Information Units (IUs), with IU gain calculated as the difference between post- and pre-treatment scores.

Stimulus Analysis: Structural Complexity of 12 Stories

Each of the twelve stimuli was analyzed for four structural indicators: (1) manually verified sentence count, (2) average sentence length, (3) proportion of complex sentences (based on the presence of subordinating conjunctions), and (4) estimated information units (approximated by counts of commas, 'and', and 'but'). Each measure was z-score normalized and averaged to compute a composite structural complexity score. Stimuli were divided into Low, Medium, and High Load Blocks based on tertile distribution of these composite scores (see Table 1, Figure 1).

Table 1. Structural Analysis of Narrative Stimuli

Story	Sentence Count	Word Count	Avg Sentence Length	Subordination Ratio	Estimated IUs	Composite Score	Load Block
Airport	14	220	15.71	0.50	21	0.698	High
Gas*	14	217	15.50	0.43	21	0.550	High
Tickets	13	212	16.31	0.85	13	0.513	High
Fire	13	223	17.15	0.54	16	0.474	High
Water	14	220	15.71	0.36	16	0.131	Medium
Paint	14	224	16.00	0.29	16	0.091	Medium
Garage Sale	14	214	15.29	0.43	13	-0.077	Medium
Library*	13	202	15.54	0.62	10	-0.199	Medium
Loan*	16	200	12.50	0.25	16	-0.261	Low
Sandwich	14	202	14.43	0.50	11	-0.304	Low
Tightrope	14	217	15.50	0.14	12	-0.511	Low
Baseball	11	152	13.82	0.45	12	-1.106	Low

Table 1. Structural features of the twelve narrative stimuli. Sentence count values were manually verified. Composite scores were calculated as the mean of z-scored structural variables (sentence count, average sentence length, subordination ratio, and estimated IUs). Load Block assignments were based on tertile split of composite scores.

Figure 1. Composite Structural Complexity Scores of 12 Stories

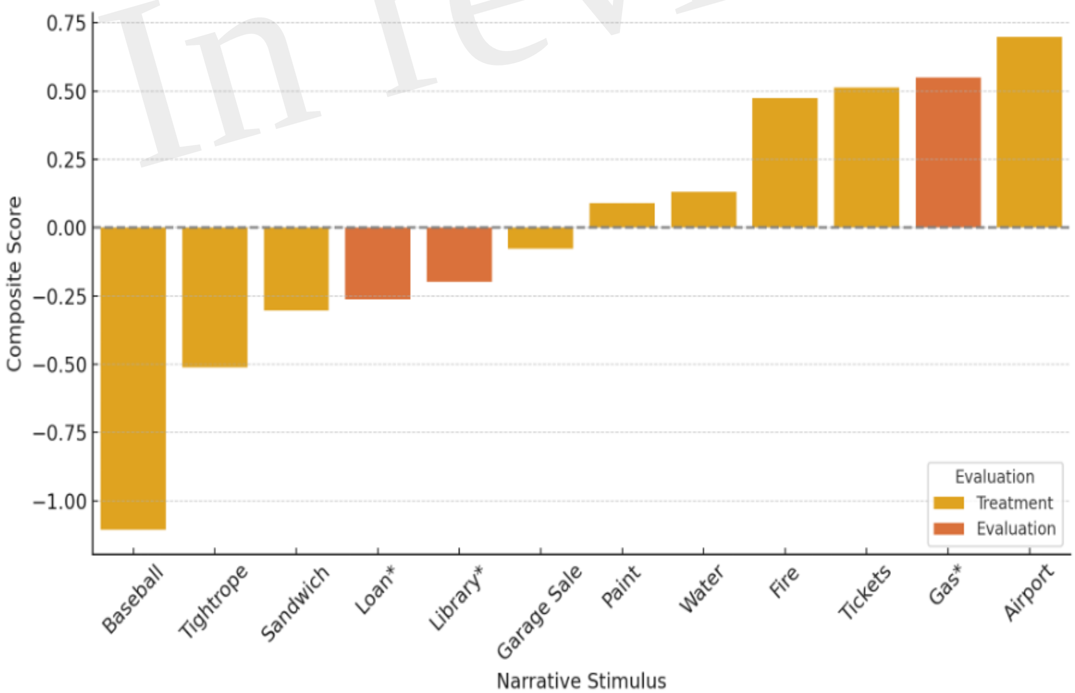

Figure 1. Composite Structural Complexity Scores of 12 Stories. Evaluation stimuli are marked with an asterisk (*) and highlighted in a different color.

The stimuli were evenly distributed across complexity tiers, permitting exploratory comparisons of structural complexity and treatment responsiveness.

Exploratory Aim: Relationship Between Structural Complexity and IU Gain

Treatment Stimuli (n = 9)

For the nine treatment stimuli, IU gain was calculated as the difference between post-treatment and pre-treatment retellings. In this case, high-load stories such as *Airport* (Composite = +0.70, Gain = +25), *Tickets* (+0.51, +30), and *Fire* (+0.47, +41) were associated with relatively higher IU gains. In contrast, low-load stimuli such as *Tightrope* (–0.51, +13) and *Sandwich* (–0.30, +11) showed smaller or inconsistent gains. Table 2 presents the composite complexity scores and corresponding IU gains for all stimuli, while Figures 2 and 3 visualize the relationship between complexity and treatment response. The average IU gain for High Block treatment stimuli was +31.5, compared to +28.7 for Low Block stories. A weak-to-moderate positive correlation was observed between composite structural complexity and IU gain ($r = +0.54$), though this exploratory result should be interpreted with caution given the small sample size and the context-specific nature of the data.

Table 2. Composite Scores and IU Gain for All 12 Narrative Stimuli

Story	Composite Score	IU Gain	Load Block
Airport	0.698	25	High
Gas*	0.55	4	High
Tickets	0.513	30	High
Fire	0.474	41	High
Water	0.131	79	Medium
Paint	0.091	27	Medium
Garage Sale	-0.077	31	Medium
Library*	-0.199	38	Medium
Loan*	-0.261	-17	Low
Sandwich	-0.304	46	Low
Tightrope	-0.511	13	Low
Baseball	-1.106	31	Low

Table 2. Composite Scores and IU Gain for All 12 Narrative Stimuli. IU Gain for treatment stimuli reflects the difference between post-treatment and pre-treatment retellings. For evaluation-only stimuli (marked with *), IU Gain reflects the difference between Post2 and Pre sessions. Load Blocks were assigned based on tertile distribution of composite scores.

Figure 2. Relationship between composite structural complexity scores and IU gain

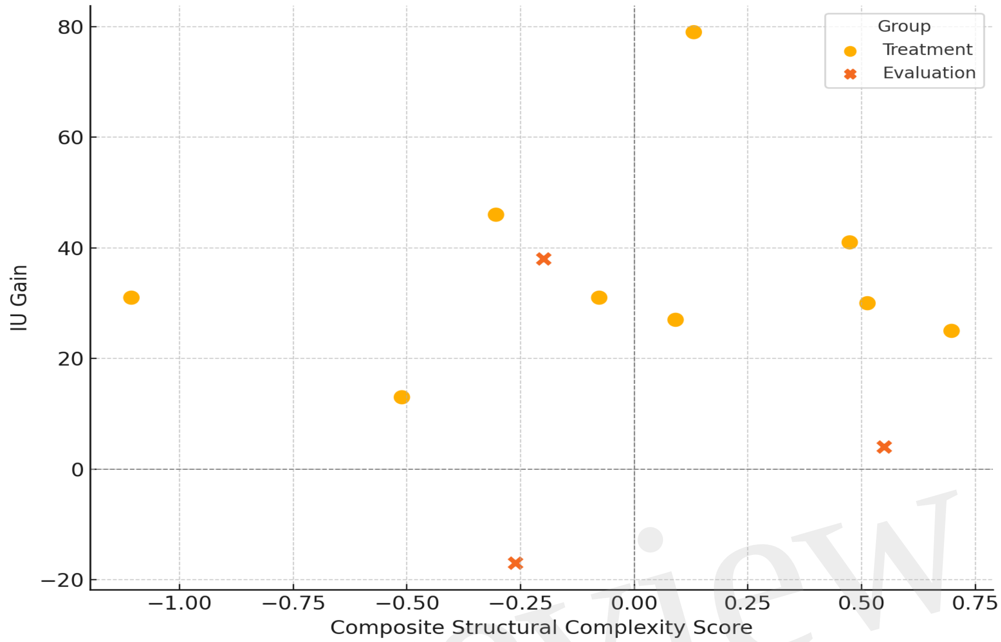

Figure 2. Relationship between composite structural complexity scores and IU gain. Each point represents a stimulus. Treatment stimuli are shown as orange circles and evaluation-only stimuli (marked with *) are shown as red stars. IU gain for evaluation stimuli is calculated as Post2 – Pre.

Figure 3. IU gain by narrative stimulus, grouped by structural complexity (Load Block)

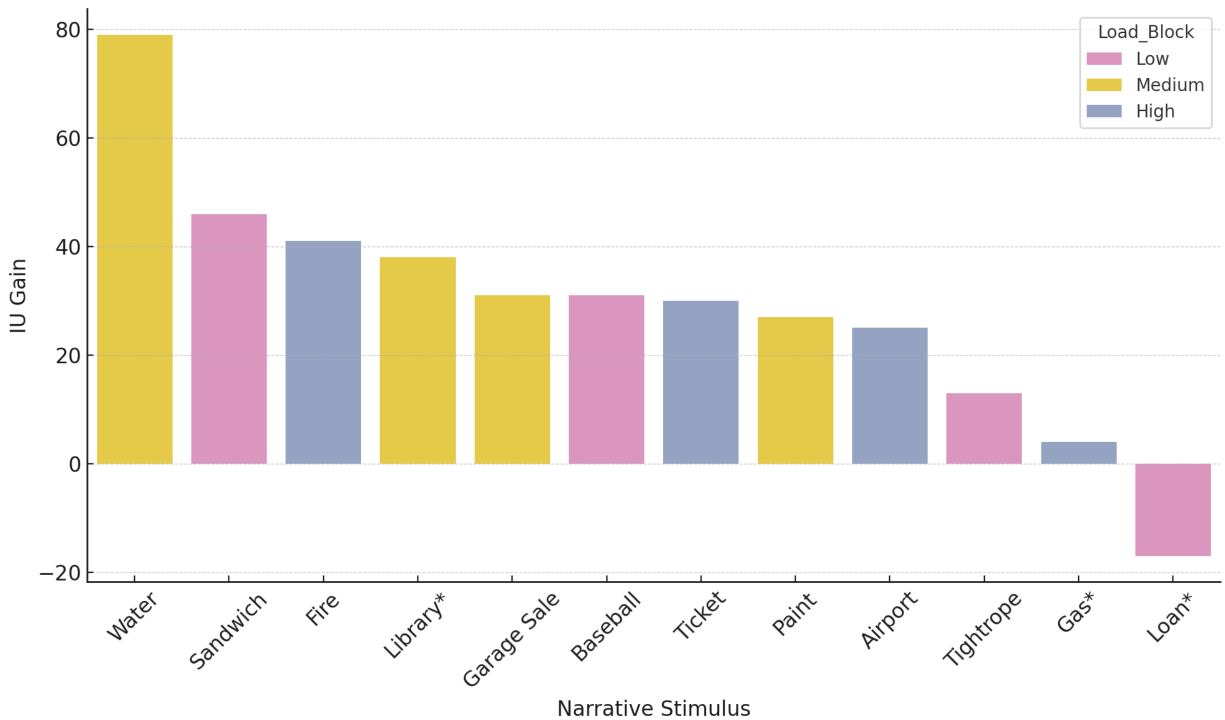

Figure 3. IU gain by narrative stimulus, grouped by structural complexity (Load Block). Colors represent Load

Block classification: Low (pink), Medium (yellow), and High (blue). Evaluation-only stimuli are marked with an asterisk (*); IU gain was calculated as the difference between post- and pre-treatment (or Post2 – Pre for evaluation-only stimuli)

Evaluation-Only Stimuli (n = 3)

Among the three evaluation-only stimuli, IU gain was calculated as the difference between Post2 and Pre sessions. Although all three stimuli were classified within the Low Load Block, they showed divergent patterns: Loan (Composite = –0.26) yielded a decrease in IU score (–17), while Library (Composite = –0.20) showed the highest gain across all stimuli (+38). A correlation analysis suggested a strong negative association between structural complexity and IU gain (r = –0.99); however, this result should be interpreted with caution given the very small sample size (n = 3) and the substantial variability across individual items. In addition to overall IU gains, temporal trajectories for the evaluation-only stimuli were examined across three timepoints: Pre, Post1, and Post2 (Figure 4). *Library* showed a steady increase from 19 to 44 to 57 IUs, indicating sustained growth even without direct treatment. *Gas* remained relatively stable across timepoints (32 → 37 → 36), while *Loan* exhibited a sharp decline from 40 to 21 and finally to 23, indicating possible decline or limited generalization effects. These results highlight the variability in maintenance and generalization patterns among untreated stimuli.

Figure 4. IU count for evaluation-only stimuli across Pre, Post1, and Post2

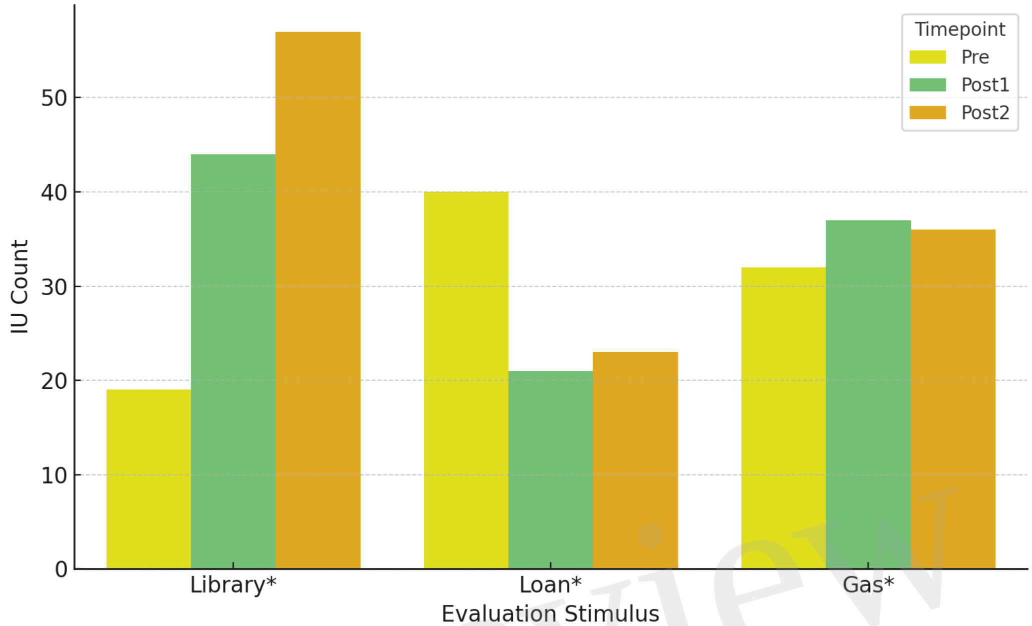

Figure 4. IU count for evaluation-only stimuli across Pre, Post1, and Post2 timepoints using actual data. Library* shows continuous improvement, Gas* remains stable, and Loan* demonstrates a sharp decline after treatment.

Discussion

This study quantitatively analyzed the structural complexity of narrative stimuli used in aphasia therapy and explored how this complexity related to discourse-level treatment responsiveness. While previous studies have focused on features of patient-produced discourse (Brookshire & Nicholas, 1994; Glosser & Deser, 1992), this study uniquely investigated the structural properties of the stimulus materials themselves and examined how those properties interacted with treatment outcomes. Several key findings emerged.

First, the correlation between composite structural complexity scores and IU gain across all twelve stories was minimal (r = 0.05), indicating no consistent linear relationship. However, when the analysis was restricted to treated stimuli only (n = 9), a moderate positive trend was observed (r = +0.54), indicating that structural complexity may interact with treatment responsiveness under certain conditions**.** Rather than dismissing the role of complexity, this pattern suggests that structural complexity may function as a moderator, with its influence likely interacting with other factors, such as auditory comprehension, topic familiarity, task repetition, or individual cognitive profiles (Caplan & Waters, 1999; Daneman & Carpenter, 1980). As shown in Table 2, the lack of consistent patterns across stimuli highlights the influence of contextual factors beyond structural load alone.

Second, when stimuli were grouped by complexity level (Low, Medium, High), Medium Load stories yielded the highest average IU gains (43.75), exceeding those of both High Load (25.0) and Low Load (18.25) groups. While this trend emerged from a small number of items per group, this tentative pattern may reflect principles from Cognitive Load Theory (Sweller et al., 2011) and Vygotsky's Zone of Proximal Development (Vygotsky, 1978), which suggest that materials of moderate difficulty can support optimal learning when appropriately matched to the learner's capabilities. This observation may also be consistent with foundational models of arousal and performance, such as the Yerkes-Dodson Law (Yerkes & Dodson, 1908), although further research is needed to determine whether this pattern is replicable and theoretically robust.

Third, a clear contrast was observed between treatment stimuli and evaluation-only stimuli. Treatment stimuli, delivered with repetition, multimodal support, and clinician feedback, led to overall higher IU gains. Evaluation stimuli, by contrast, had a lower average gain (8.33) and showed much greater variability across items. Notably, the evaluation-only stimulus *Library* (Composite = –0.20) yielded a substantial IU gain (+38) despite not being directly trained. This suggests that treatment-induced strategies or language processing mechanisms may have generalized to untrained materials, producing positive effects. It is also plausible that the stimulus benefited from thematic familiarity or alignment with pre-existing narrative schemas (Bartlett, 1932; van Dijk & Kintsch, 1983). For example, situations involving libraries often include familiar actions such as reading, borrowing books, or interacting in quiet, structured settings, which are widely shared and reinforced in cultural experience (Nelson, 1986). These schema-consistent elements may have facilitated access to relevant linguistic and conceptual representations during retelling. The observed gain in this case may therefore reflect not only generalization effects from therapy, but also the cognitive advantage afforded by the use of familiar and semantically coherent content. Taken together, these findings indicate that such generalization effects may arise not solely from structural complexity, but from the interaction of multiple factors, including treatment conditions, thematic properties of the stimuli, and individual background knowledge and cognitive state (Kay & Ellis, 1987; Edmonds et al., 2009).

Additionally, because evaluation stimuli were presented only once without feedback, they may have more directly reflected the raw cognitive burden imposed by structural complexity. In contrast, the fact that several high-complexity treatment stimuli yielded strong IU gains suggests that the therapeutic process itself—through repetition, scaffolding, and feedback—may have enabled the participant to manage or even overcome the processing demands posed by more complex narratives. This aligns with the logic of Treatment of Underlying Forms (Thompson & Shapiro, 2005; Thompson et al., 2003), which supports the use of complex structures to promote generalization when accompanied by appropriate support.

Furthermore, the generalization of treatment effects to some untrained stimuli suggests the possibility that the participant internalized strategies or representations that transferred beyond the trained material (Ulatowska et al., 1981). It is also possible that generalization resulted from implicit learning or schema-based facilitation, such as thematic familiarity or narrative scaffolding. However, this interpretation remains speculative given the limited data available.

These observations suggest that structural complexity is better conceptualized not as a binary factor but as a modifiable design parameter that can be adjusted in relation to therapeutic goals and individual needs. Stimuli can be sequenced or adapted not only based on thematic content, but also on structural characteristics that align with a learner's readiness and cognitive resources.

To further clarify these patterns, Table 3 summarizes how each level of structural complexity—low, medium, and high—interacted with stimulus context (treatment vs. evaluation) to influence responsiveness. This contextual view highlights that medium-complexity stimuli may be better suited for evaluation tasks, where no feedback or support is provided, while high-complexity stimuli may perform better in supported treatment environments.

Table 3. Contextual Suitability of Narrative Stimuli by Structural Complexity Level

Complexity Level	Suitability in Treatment Context	Suitability in Evaluation Context	Interpretation	Example Stimuli
Low	Limited effectiveness due to excessive simplicity	Low burden but reduced responsiveness	May lack cognitive stimulation	Sandwich, Tightrope, Baseball (T), Loan* (E)
Medium	Stable and consistently effective	Elicits adequate response even without feedback	Potentially suitable for evaluation stimuli	Water, Paint, Garage Sale (T), Library* (E)
High	High responsiveness when supported with repetition and feedback	Variable or burdensome under single exposure	Potentially suitable for treatment stimuli	Airport, Tickets, Fire (T), Gas* (E)

*Note: (T) = Treatment stimulus, (E) = Evaluation-only stimulus.

Table 3. Contextual suitability of narrative stimuli by structural complexity level.
The table summarizes the interpretive roles of stimuli with low, medium, and high structural complexity across treatment and evaluation contexts. Example stimuli reflect the original stimulus names used in the study, with (T) indicating treatment stimuli and (E) indicating evaluation-only stimuli.

This mapping reinforces the notion that structural complexity is neither inherently beneficial nor detrimental, but rather dependent on context. It underscores the importance of calibrating stimulus properties in relation to both linguistic characteristics and therapeutic goals, such as supporting generalization through untutored retellings or fostering growth through structured intervention. This perspective aligns with principles from cognitive load theory and the zone of proximal development, highlighting the value of tailoring task difficulty to individual readiness. Taken together, these considerations support a precision-based design framework in discourse-level aphasia therapy (Sweller et al., 2011; Vygotsky, 1978; Thompson & Shapiro, 2005).

Clinical Relevance and Theoretical Implications

Taken together, the findings suggest that structural complexity is not a fixed facilitator or barrier to progress, but a dynamic feature whose impact depends on task demands and treatment context. Rather than assuming that greater complexity is always helpful or harmful, clinicians and researchers may benefit from treating complexity as a modifiable design parameter, one that can be calibrated according to individual profiles and therapy goals (Thompson & Shapiro, 2005; Sweller et al., 2011). Structurally informed stimulus design may facilitate alignment of input difficulty with individual readiness, support generalization through graded exposure, and promote adaptive sequencing in discourse-based aphasia therapy. Rather than establishing a fixed treatment framework, this approach may inform the development of precision-based, cognitively guided strategies for stimulus selection and calibration in future research (Thompson & Shapiro, 2005; Sweller et al., 2011; Marini et al., 2011; Boyle, 2010). In this view, structural complexity becomes a modifiable design parameter—one that can be explicitly measured, systematically manipulated, and flexibly matched to evolving therapeutic goals.

Conclusion

This single-case study demonstrates that the structural complexity of narrative stimuli can be systematically measured and categorized using quantifiable features. It also introduces a replicable tool for input calibration in therapy design. Although no consistent linear relationship was observed between complexity and treatment gains, certain patterns suggest that complexity may interact with treatment responsiveness under specific conditions. For example, complex stimuli appeared to support language production when actively trained but may have hindered generalization when untrained. These preliminary findings indicate that structural complexity may function as a context-sensitive variable, rather than a fixed facilitator or barrier, with potential to be modulated according to therapeutic goals and task demands. Further empirical research is needed to validate these observations and to explore the broader applicability of structural profiling in discourse-level intervention.

Limitations and Future Directions

This study is not without limitations. It relied on a single-case design and included only twelve stimuli, limiting generalizability. Additionally, while four structural features were quantified, other potentially influential linguistic variables were not examined. These include lexical sophistication (Kyle & Crossley, 2015), imageability (Paivio, 1991), verb argument structure complexity (Levelt, 1989; Thompson & Shapiro, 2007), and referential cohesion (Halliday & Hasan, 1976), which may affect discourse processing and production in nuanced ways. Their omission limits the explanatory power of the current structural profile, and future studies may benefit from incorporating a broader range of linguistic metrics to capture additional sources of variability in narrative responsiveness. Moreover, external variables such as topic familiarity, emotional salience, and cognitive state at time of testing were not controlled, particularly for evaluation-only stimuli. These factors may explain the variability observed across stories with similar complexity scores. Furthermore, the inherent clinical heterogeneity in aphasia profiles suggests that stimulus responsiveness may vary across individuals, warranting multi-case designs to assess generalizability. Future research should incorporate larger samples and more diverse stimulus sets, ideally using model-based approaches to capture interactions among stimulus complexity, participant profiles, and treatment conditions. Integrating real-time discourse data from individuals with aphasia and longitudinal follow-up could offer additional insight into the role of stimulus features in supporting long-term functional outcomes. Taken together, these limitations underscore the need of developing structured, multidimensional frameworks that systematically integrate structural, cognitive, and contextual factors in discourse-level aphasia rehabilitation.

References

1. Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). American Psychological Association. https://doi.org/10.1037/10096-006
2. Glosser, G., & Deser, T. (1992). A comparison of changes in macrolinguistic and microlinguistic aspects of discourse production in normal aging. *Journal of Gerontology, 47*(4), P266–P272. https://doi.org/10.1093/geronj/47.4.P266
3. Kagan, A., & Simmons-Mackie, N. (2007). Beginning with the end: Outcome-driven assessment and intervention with life participation in mind. *Topics in Language Disorders, 27*(4), 309–317. https://doi.org/10.1097/01.TLD.0000299885.39488.bf
4. Ulatowska, H. K., North, A. J., & Macaluso-Haynes, S. (1983). Production of narrative discourse in aphasia. *Brain and Language, 19*(2), 317–334. https://doi.org/10.1016/0093-934X(83)90074-3

5. Marini, A., Boewe, A., Caltagirone, C., & Carlomagno, S. (2005). Age-related differences in the production of textual descriptions. *Journal of Psycholinguistic Research, 34*(5), 439–463. https://doi.org/10.1007/s10936-005-6203-z
6. Yoo, H. & McNeil, M. R. (2019). Story retelling and verbal working memory in people with aphasia. *Clinical Archives of Communication Disorders, 4*(3), 223–235.
7. Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing* (pp. 53–120). Ablex Publishing.
8. Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language, 24*(5), 612–630. https://doi.org/10.1016/0749-596X(85)90049-X
9. Liles, B. Z. (1993). Narrative discourse in children with language disorders and children with normal language: A critical review of the literature. *Journal of Speech and Hearing Research, 36*(5), 868–882. https://doi.org/10.1044/jshr.3605.868
10. See, K., & Ryan, E. B. (1995). Cognitive mediation of age differences in discourse processing. *Journal of Speech-Language Pathology and Audiology, 20*(2), 85–168.
11. Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension.* New York: Academic Press.
12. Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163–182. https://doi.org/10.1037/0033-295X.95.2.163
13. Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.
14. Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1
15. Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9*(1), 111–151. https://doi.org/10.1016/0010-0285(77)90006-8
16. Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing* (Vol. 2, pp. 53–120). Ablex Publishing.
17. Marini, A., Boewe, A., Caltagirone, C., & Carlomagno, S. (2005). Age-related differences in the production of textual descriptions. *Journal of Psycholinguistic Research, 34*(5), 439–463. https://doi.org/10.1007/s10936-005-6200-z
18. Thompson, C. K., & Shapiro, L. P. (2005). Treating agrammatic aphasia within a linguistic framework: Treatment of underlying forms. *Aphasiology, 19*(10–11), 1021–1036. https://doi.org/10.1080/02687030444000696
19. Dipper, L., Marshall, J., Boyle, M., Botting, N., Hersh, D., Pritchard, M., & Cruice, M. (2020). Treatment for improving discourse in aphasia: a systematic review and synthesis of the evidence base. *Aphasiology*, *35*(9), 1125–1167. https://doi.org/10.1080/02687038.2020.1765305
20. Python Software Foundation. (2023). *Python (Version 3.11) [Programming language]*. https://www.python.org
21. McKinney, W. (2010). Data structures for statistical computing in Python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 51–56). https://doi.org/10.25080/Majora-92bf1922-00a
22. Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array programming with NumPy. *Nature, 585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2
23. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering, 9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55
24. Brookshire, R. H., & Nicholas, L. E. (1994). Speech sample size and test-retest stability of connected speech measures for adults with aphasia. *Journal of Speech, Language, and Hearing Research*, 37(2), 399–407. https://doi.org/10.1044/jshr.3702.399

25. Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, 22(1), 77–94. https://doi.org/10.1017/S0140525X99001788
26. Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6
27. Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4
28. Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.
29. Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology*, 18(5), 459–482. https://doi.org/10.1002/cne.920180503
30. Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.
31. Nelson, K. (1986). *Event knowledge: Structure and function in development*. Erlbaum.
32. Kay, J., & Ellis, A. W. (1987). A cognitive neuropsychological case study of anomia: Implications for psychological models of word retrieval. *Brain*, 110(3), 613–629. https://doi.org/10.1093/brain/110.3.613
33. Edmonds, L. A., Nadeau, S. E., & Kiran, S. (2009). Effect of Verb Network Strengthening Treatment (VNeST) on lexical retrieval of content words in sentences in persons with aphasia. *Aphasiology*, 23(3), 402–424. https://doi.org/10.1080/02687030802291339
34. Thompson, C. K., Shapiro, L. P., Kiran, S., & Sobecks, J. (2003). The role of syntactic complexity in treatment of sentence deficits in agrammatic aphasia: The complexity account of treatment efficacy (CATE). *Journal of Speech, Language, and Hearing Research*, 46(3), 591–607. https://doi.org/10.1044/1092-4388(2003/047)
35. Ulatowska, H. K., North, A. J., & Macaluso-Haynes, S. (1981). Production of narrative and procedural discourse in aphasia. *Brain and Language*, 13(2), 345–371. https://doi.org/10.1016/0093-934X(81)90100-0
36. Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly, 49*(4), 757–786. https://doi.org/10.1002/tesq.194
37. Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology / Revue canadienne de psychologie, 45*(3), 255–287. https://doi.org/10.1037/h0084295
38. Thompson, C. K., & Shapiro, L. P. (2007). Complexity in treatment of syntactic deficits. *American Journal of Speech-Language Pathology, 16*(4), 321–335. https://doi.org/10.1044/1058-0360(2007/005)
39. Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.
40. 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
41. Boyle, M. (2010). Semantic Feature Analysis Treatment for Aphasic Word Retrieval Impairments: What's in a Name? *Topics in Stroke Rehabilitation*, *17*(6), 411–422. https://doi.org/10.1310/tsr1706-411

Funding Statement
This work was supported by start-up research funds provided by Baylor University.

Conflict of Interest
The author declares no commercial or financial conflicts of interest.

Data Availability Statement
The data and materials supporting the conclusions of this article are available from the author upon reasonable request.
Portions of the treatment data reported here are also included in a separate manuscript currently under review, which focuses on overall treatment effects. The present study exclusively analyzes stimulus structure and its relationship to treatment responsiveness.

Ethics Statement
This study involving human participants was reviewed and approved by the Institutional Review Board at Baylor University (IRB# 2219727-9). Written informed consent was obtained from the participant prior to participation in accordance with institutional guidelines and the Declaration of Helsinki.

Author Contributions
The author confirms sole responsibility for the conceptualization, design, data collection, analysis, interpretation, and writing of this study.