Available Online: 1 September 2026 DOI: 10.54254/2753-7080/2026.36483

The explanatory value of multidimensional measures of linguistic complexity for aphasia severity: evidence from AphasiaBank speech data

Lvxin Wang

Beijing Institute of Technology, Beijing, China

wanglx265358@163.com

Abstract. Aphasia is an acquired language disorder marked by substantial individual variability. Quantitative measures of connected speech may clarify how linguistic performance relates to aphasia severity. Semi-spontaneous speech samples from 22 English-speaking adults with aphasia in the APROCSA subcorpus of AphasiaBank were analyzed. Eight measures represented three domains: Mean Length of Utterance (MLU), clause density, and embedded clause ratio for syntax; type–token ratio, verb diversity, and content-word ratio for lexis; and verb marking and noun marking for morphology. Pearson correlations examined associations with Quick Aphasia Battery Overall scores. Three domain-specific regression models and one comprehensive model containing all eight measures were compared using adjusted R^2, AIC, and BIC.MLU ($r = 0.734$), clause density ($r = 0.592$), embedded clause ratio ($r = 0.488$), and verb marking ($r = 0.753$) were significantly positively associated with QAB Overall scores, with higher scores indicating milder aphasia. The lexical measures and noun marking were not significantly associated with QAB Overall. Among the domain-specific models, the morphological model showed the greatest explanatory power (adjusted $R^2 = 0.523$), followed by the syntactic model (adjusted $R^2 = 0.494$); the lexical model was nonsignificant. The comprehensive model provided the strongest explanatory power and best fit (adjusted $R^2 = 0.805$; AIC = 64.82; BIC = 75.73). Syntactic and morphological measures, particularly MLU and verb marking, were more sensitive to variation in aphasia severity than the lexical measures examined. Combining measures across domains provided complementary information, supporting a multidimensional approach to connected-speech assessment in aphasia.

Keywords: aphasia, linguistic complexity, discourse analysis, aphasia severity, AphasiaBank

1. Introduction

Aphasia is an acquired language disorder caused by brain damage. It most commonly occurs following a stroke in the language-dominant hemisphere, typically the left hemisphere. Individuals with aphasia may experience varying degrees of impairment in spoken language production, language comprehension, reading, and writing, with substantial consequences for everyday communication, social participation, and quality of life [1-3]. A growing body of research has shown that aphasia not only markedly compromises language and communication abilities but is also closely associated with long-term functional disability, psychological distress and depression, and reduced quality of life [4, 5]. Accurate assessment of language abilities is therefore essential for clinical diagnosis, evaluation of aphasia severity, prognostic decision-making, and the development of appropriate rehabilitation programs.

Discourse impairment is among the most prominent clinical manifestations of aphasia. As Prins and Bastiaanse [6] observed, aphasia is fundamentally a disorder of language rather than pragmatics. The assessment of language production in individuals with aphasia should therefore prioritize theoretically grounded linguistic analyses, including quantitative measures of such fine-grained linguistic domains as syntax, lexis, and morphology. Fromm et al. [7], in their comparison of manual and automated analyses of connected speech, subsequently demonstrated that quantitative language measures can objectively characterize language production in aphasia and have considerable potential for clinical application. Casilio et al. [8] reduced more than 20 salient features of natural language production in individuals with aphasia to four interpretable dimensions—paraphasia, logopenia, agrammatism, and motor speech impairment—indicating that language impairment spans multiple levels of processing. A systematic review and meta-analysis by Beber et al. [9] further showed that morphosyntactic processing deficits are primarily associated with damage to the left frontal regions, particularly the inferior frontal gyrus, whereas impairments in thematic-role processing are mainly linked to temporoparietal regions. These findings provide substantial evidence for the neural basis of impaired morphosyntactic abilities in aphasia. Rashmi et al. [10] likewise emphasized that morphosyntactic impairment is a core characteristic of agrammatic aphasia and is of considerable clinical relevance. Systematic quantification of lexical, syntactic, and morphological features in the language production of individuals with aphasia may therefore support a more objective assessment of their language abilities, provide evidence for evaluating aphasia severity, and help account for individual differences in the extent of impairment.

Previous studies have examined the relationships between language measures, language ability, and aphasia severity at several fine-grained levels of linguistic organization. Some of these measures have shown considerable potential for predicting and explaining aphasia severity. At the syntactic level, Gordon [11] found that Mean Length of Utterance in words (MLU) may be influenced by lexical retrieval ability but primarily reflects grammatical complexity. Gordon [12] further demonstrated that MLU effectively distinguished nonfluent aphasia subtypes, including Broca's aphasia and transcortical motor aphasia, from fluent or milder profiles, including individuals classified as not aphasic by the Western Aphasia Battery (NABW) and those with conduction or anomic aphasia. This distinction reflects differences in syntactic production across aphasia profiles and suggests that MLU is a sensitive and useful measure for characterizing aphasia severity, particularly the extent of syntactic impairment. Kuvač Kraljević et al. [13] found that individuals with aphasia had significantly lower clause density—defined as the mean number of clauses per T-unit—than neurotypical controls. This finding indicates that clause density can distinguish individuals with aphasia from neurologically healthy adults and may help explain or predict variation in language ability and aphasia severity. At the lexical level, the Type–Token Ratio (TTR) is commonly used to measure lexical diversity, with higher values indicating the use of a broader range of vocabulary [14]. Lexical diversity has long been shown to distinguish individuals with aphasia from neurotypical adults [15]. Bastiaanse [16] noted that most individuals with aphasia experience difficulty retrieving verbs and that this difficulty is often more pronounced than noun-retrieval impairment. Quantitative measures of verb use, such as verb diversity, may therefore help distinguish individuals with aphasia from neurotypical speakers and may also have value in predicting or explaining aphasia severity. Alyahya et al. [17] found that individuals with poststroke aphasia produced significantly fewer and less diverse content words in connected discourse, although the lexical–semantic complexity of the words they produced was relatively preserved. These findings point to the potential value of content-word measures, including the content-word ratio, for characterizing language ability and aphasia severity. At the morphological level, Gordon [12] investigated noun marking and verb marking in individuals with aphasia and found that verb marking, together with other language measures, captured aspects of linguistic formulation ability. This finding suggests that the ability to produce verbal inflectional markings is closely related to overall language ability and the extent of language impairment.

Although previous studies have provided preliminary evidence that some language measures may help predict aphasia severity, several important questions remain unresolved. First, among the many available measures of linguistic complexity, which are most strongly associated with aphasia severity? Answering this question may help identify a core set of measures with the greatest potential for clinical assessment. Second, which of the three linguistic domains—syntax, lexis, or morphology—has the greatest explanatory power and predictive potential for aphasia severity? In other words, which dimension of linguistic complexity most clearly reflects variation in the extent of aphasic impairment? Addressing this question may indicate which level of language should receive priority in linguistic assessment. Third, does combining measures from all three domains yield a more accurate account of aphasia severity than using measures from any single domain? More specifically, does a multidimensional model provide incremental predictive validity? An answer to this question would offer empirical evidence regarding the need to integrate measures from multiple linguistic domains in clinical language assessment.

To address these gaps, the present study draws on spontaneous speech samples from 22 individuals with aphasia in AphasiaBank. It systematically evaluates the explanatory power and predictive potential of eight measures of linguistic complexity across three domains in relation to aphasia severity, as measured by the QAB Overall score. The syntactic measures comprise Mean Length of Utterance (MLU), clause density, and embedded clause ratio; the lexical measures include Type–Token Ratio (TTR), verb diversity, and content-word ratio; and the morphological measures consist of verb marking and noun marking. The study addresses the following questions:

(1) Which of the eight language measures are most strongly associated with QAB Overall scores?

(2) To what extent do the syntactic, lexical, and morphological dimensions of linguistic complexity account for individual variation in QAB Overall scores, and what potential value does each dimension have as a linguistic indicator of aphasia severity?

(3) Does a combined model incorporating measures from all three domains provide greater explanatory power than each domain-specific model and demonstrate stronger potential for predicting aphasia severity?

The findings are expected to provide empirical guidance for selecting language measures in the clinical assessment of aphasia and quantitative evidence supporting a multidimensional approach to the evaluation of linguistic complexity.

2. Method

2.1. Participants

The data analyzed in this study were drawn from the spoken-language samples of adults with aphasia in the APROCSA corpus, a subcorpus of AphasiaBank. The corpus comprises 22 participants with aphasia, five of whom underwent longitudinal follow-up assessments. All participants completed the standardized AphasiaBank protocol and the Quick Aphasia Battery [18, 19]. English was the primary language of every participant, and aphasia in the great majority of cases resulted from stroke. Further participant characteristics are presented in Table 1.

Table 1. Demographic characteristics of participants with aphasia

Participants	Age (months)	Sex	Handedness	Years of education	QAB Overall
1	58	male	right	12	9
2	65	male	right	16	9.1
3	60	female	right	14	9.2
4	67	male	right	15	4.8
5	70	male	ambidextrous	20	5.4
6	51	female	right	14	9.3
7	51	female	right	14	9.5
8	54	female	right	12	2.1
9	59	male	right	14	6.5
10	59	male	right	14	7.2
11	63	female	left	12	7.4
12	63	female	left	12	7.2
13	52	female	left	18	7.7
14	55	female	right	20	5.2
15	55	female	right	20	6.4
16	55	female	right	20	7.2
17	65	male	right	13	5.3
18	55	male	right	14	4.9
19	55	male	right	14	6.8
20	55	male	right	14	7.5
21	68	male	right	16	4.2
22	41	female	right	12	4.1

Note: QAB = Quick Aphasia Battery; QAB Overall = Overall Aphasia Score.

2.2. Tasks

Under the AphasiaBank protocol, discourse samples are elicited through several tasks, including personal narratives, descriptions of single pictures and picture sequences, story retelling (*Cinderella*), and procedural discourse. The present study primarily analyzed picture-based narratives, retellings of *Cinderella*, and a series of structured interview exchanges between the examiner and the participant. During these exchanges, participants were asked about topics such as their name, current location, medical condition, what happened on the day of their stroke, and matters they considered important in life. These tasks were selected because narrative elicitation generally produces longer language samples and closely reflects many discourse activities encountered in everyday life, such as recounting events that occurred during the day [20].

2.3. Variable analysis

Variables were selected from the data available in AphasiaBank to provide as comprehensive an assessment as possible of participants' abilities in lexical selection and sentence formulation. The selected measures covered three major dimensions of language production: syntactic, lexical, and morphological complexity.

Previous studies have suggested that the core linguistic impairments associated with aphasia are primarily manifested at the microlinguistic level [21-23]. Representative microlinguistic measures therefore offer considerable value for language assessment and clinical interpretation. The following sections describe all variables included in the study and the procedures used to calculate them. All measures were obtained using the Computerized Language Analysis program (CLAN), with Microsoft Excel used to assist with data organization and calculation. Nonlinguistic fillers and unintelligible words or utterances were excluded from the analyses.

2.3.1. Sentence-level measures

Syntactic complexity at the sentence level was assessed using three measures. Mean Length of Utterance in words (MLU-w) represents the average number of words produced per utterance and was calculated as Equation (1):

$$MLU\text{-}w = \frac{Total\ number\ of\ words}{Total\ number\ of\ utterances} \tag{1}$$

A higher $MLU\text{-}w$ generally indicates an ability to produce longer and more structurally complex utterances and is therefore commonly regarded as a measure of overall syntactic complexity [24].

Clause density measures the average number of clauses contained within each syntactic unit. Following previous research, all clauses were identified and annotated. $Clause\ density$ was calculated as Equation (2):

$$Clause\ Density = \frac{Total\ number\ of\ clauses}{Total\ number\ of\ T\text{-}units} \tag{2}$$

A $T\text{-}unit$ was defined as one independent main clause together with all subordinate clauses and nonclausal structures attached to it. This measure captures the extent of clausal elaboration in language production. A higher value indicates that each $T\text{-}unit$ contains more clauses and consequently exhibits greater syntactic complexity [13].

The embedded clause ratio measures the proportion of all clauses that are embedded and thus reflects a participant's ability to produce syntactically embedded structures. Object clauses, relative clauses, complement clauses, and other embedded structures were identified and annotated. The embedded clause ratio was calculated as Equation (3):

$$Embedded\ Clause\ Ratio = \frac{Total\ number\ of\ embedded\ clauses}{Total\ number\ of\ clauses} \tag{3}$$

A higher ratio indicates more frequent production of embedded clauses, which represent relatively advanced syntactic structures, and therefore reflects a higher level of syntactic complexity.

2.3.2. Lexical-level measures

Three lexical-level measures were included to examine lexical use and lexical complexity in individuals with aphasia.

First introduced by Johnson [25], the Type–Token Ratio (TTR) measures lexical diversity and was calculated as Equation (4):

$$TTR = \frac{Number\ of\ different\ words\ (types)}{Total\ number\ of\ words\ (tokens)} \tag{4}$$

A higher TTR indicates that the speaker uses a broader and more diverse vocabulary [14].

Verb diversity was used to assess the richness of the verbs produced by each participant. In the present study, it was calculated as Equation (5):

$$Verb\ Diversity = \frac{Number\ of\ different\ verb\ types}{Total\ number\ of\ verb\ tokens} \tag{5}$$

A higher value indicates a more diverse verb repertoire. Because verbs play a central role in syntactic organization and event representation, verb diversity can provide a sensitive measure of lexical retrieval and semantic expression in individuals with aphasia. It has therefore become an important measure in recent aphasia research [26].

The content-word ratio reflects the proportion of information-bearing words in a language sample. Content words included nouns, main verbs, adjectives, and adverbs. It was calculated as Equation (6):

$$Content\ Word\ Ratio = \frac{Number\ of\ content\ word\ tokens}{Total\ number\ of\ word\ tokens} \tag{6}$$

A higher *content-word ratio* indicates that a greater proportion of the language sample consists of information-bearing lexical items [17, 27].

2.3.3. Morphological-level measures

Morphological complexity was assessed using two measures: verb marking and noun marking. Verb marking represents the proportion of verbs inflected for number and tense and was calculated as Equation (7):

$$Verb\ Marking = \frac{Number\ of\ verbs\ inflected\ for\ number\ and\ tense}{Total\ number\ of\ verbs} \tag{7}$$

Noun marking represents the proportion of nouns marked for plurality and possession and was calculated as Equation (8):

$$Noun\ Marking = \frac{Number\ of\ nouns\ marked\ with\ plural\ and\ possessive\ morphemes}{Total\ number\ of\ nouns} \tag{8}$$

Both measures ranged from 0 to 1. Higher values indicate a greater ability to use inflectional morphology in language production [28].

2.4. Data analysis

After data preparation, Pearson correlation analyses were first conducted to examine the relationships between the demographic variables—age and years of education—and aphasia severity, as indexed by the QAB Overall score. This preliminary analysis was used to determine whether either variable should be included as a covariate in the subsequent analyses.

Descriptive statistics were then calculated for the QAB Overall score and the eight measures of linguistic complexity. These included the Mean (*M*), Standard Deviation (*SD*), and range of each variable.

To examine the bivariate relationships between individual language measures and aphasia severity, Pearson correlations were calculated between QAB Overall scores and each of the eight measures: MLU, clause density, embedded clause ratio, TTR, verb diversity, content-word ratio, verb marking, and noun marking.

Because the combined model included all eight predictors, multicollinearity and residual diagnostics were conducted before model fitting to determine whether the principal assumptions of multiple linear regression were met. Multicollinearity was assessed using the Variance Inflation Factor (VIF), with values below 10 taken to indicate the absence of severe multicollinearity [29]. The normality of model residuals was assessed using the Shapiro–Wilk test.

Four multiple linear regression models were subsequently fitted to compare the explanatory power of different linguistic-complexity models for aphasia severity. The first three were domain-specific models. The syntactic complexity model included MLU, clause density, and embedded clause ratio as predictors; the lexical complexity model included TTR, verb diversity, and content-word ratio; and the morphological complexity model included verb marking and noun marking. The fourth model was a combined linguistic complexity model incorporating all eight language measures as predictors.

Model performance was compared primarily using the coefficient of determination (R^2), adjusted coefficient of determination (adjusted R^2), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). All statistical analyses were conducted in *R* [30], with the significance level set at $\alpha = 0.05$.

3. Results

Before conducting the main analyses, the relationships between the demographic variables and QAB Overall scores were examined. Pearson correlation analyses showed that neither age ($r = -0.050$, $p = 0.826$) nor years of education ($r = -0.030$, $p = 0.895$) was significantly associated with QAB Overall scores. Given the absence of statistically significant associations, neither variable was included as a covariate in the subsequent regression analyses. The remaining analyses therefore focused on the relationships between QAB Overall scores and measures of syntactic, lexical, and morphological complexity.

3.1. Descriptive statistics

Descriptive statistics were first calculated for the measures of linguistic complexity and QAB Overall scores among participants with aphasia. The results are presented in Table 2.

As shown in Table 2, the mean QAB Overall score was 6.64 (*SD* = 1.97), with scores ranging from 2.10 to 9.50. This range indicates that the sample included participants with varying degrees of aphasic impairment. The measures of linguistic complexity also displayed considerable variability. At the syntactic level, the mean values for Mean Length of Utterance (MLU), clause density, and embedded clause ratio were 4.49 (*SD* = 1.58), 1.00 (*SD* = 0.46), and 0.28 (*SD* = 0.12), respectively. At the lexical level, the mean values for Type–Token Ratio (TTR), verb diversity, and content-word ratio were 0.33 (*SD* = 0.10), 0.24 (*SD* = 0.13), and 0.42 (*SD* = 0.07), respectively. At the morphological level, the mean verb-marking and noun-marking rates were 0.53 (*SD* = 0.16) and 0.16 (*SD* = 0.10), respectively.

Because these descriptive statistics provide information only about the central tendency and dispersion of the data, correlation and regression analyses were subsequently conducted to examine the relationships between the language measures and QAB Overall scores, as well as their explanatory power. The results are reported in the following sections.

Table 2. Descriptive statistics for the language measures

Variable	*M*	*SD*	Minimum	Maximum
QAB Overall	6.64	1.97	2.10	9.50
MLU	4.49	1.58	2.09	7.29
Clause Density	1.00	0.46	0.00	2.02
Embedded Clause Ratio	0.28	0.12	0.00	0.48
Type–Token Ratio (TTR)	0.33	0.10	0.05	0.49
Verb Diversity	0.24	0.13	0.09	0.56
Content Word Ratio	0.42	0.07	0.22	0.58
Verb Marking	0.53	0.16	0.08	0.78
Noun Marking	0.16	0.10	0.02	0.38

Note: *M* = Mean; *SD* = Standard Deviation.

3.2. Correlations between linguistic complexity and aphasia severity

To examine the relationships between the individual language measures and aphasia severity, Pearson correlation analyses were conducted between the QAB Overall score and the eight language measures. The results are presented in Table 3.

At the syntactic level, MLU (r = 0.734, p < 0.001), clause density (r = 0.592, p = 0.004), and embedded clause ratio (r = 0.488, p = 0.021) were all significantly and positively correlated with QAB Overall scores, with MLU showing the strongest association. At the morphological level, only verb marking was significantly and positively correlated with QAB Overall scores (r = 0.753, p < 0.001), whereas the correlation involving noun marking did not reach statistical significance (r = 0.281, p = 0.205). At the lexical level, none of the three measures—TTR (r = 0.046, p = 0.840), verb diversity (r = 0.110, p = 0.626), or content-word ratio (r = 0.020, p = 0.929)—was significantly correlated with QAB Overall scores.

Overall, four of the eight language measures—MLU, clause density, embedded clause ratio, and verb marking—were significantly and positively correlated with QAB Overall scores. The remaining four measures showed no statistically significant associations.

Table 3. Pearson correlations between QAB Overall scores and the language measures

Language Measure	Pearson's r	p
MLU	0.734***	< 0.001
Clause Density	0.592**	0.004
Embedded Clause Ratio	0.488*	0.021
Type–Token Ratio, TTR	0.046	0.840
Verb Diversity	0.110	0.626
Content Word Ratio	0.020	0.929
Verb Marking	0.753***	< 0.001
Noun Marking	0.281	0.205

Note: r = Pearson's correlation coefficient; QAB Overall = Overall score on the Quick Aphasia Battery. Asterisks indicate statistical significance: ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels respectively.

3.3. Comparison of unidimensional and multidimensional linguistic complexity models

Before fitting the regression models, multicollinearity and residual normality were examined for the comprehensive model containing all eight predictors. Variance Inflation Factors (VIFs) ranged from 1.74 to 7.28, all below the commonly used threshold of 10 [29], indicating that severe multicollinearity was not present. The Shapiro–Wilk test further showed that the model residuals did not deviate significantly from a normal distribution (W = 0.967, p = 0.638), thereby satisfying the normality assumption of linear regression.

Four regression models were subsequently fitted—syntactic complexity, lexical complexity, morphological complexity, and comprehensive linguistic complexity—to compare their ability to explain variation in aphasia severity as measured by QAB Overall. The results are presented in Table 4.

Among the three single-domain models, the morphological complexity model showed the greatest explanatory power (adjusted R^2 = 0.523, $F(2,19)$ = 12.53, p < 0.001), followed by the syntactic complexity model (adjusted R^2 = 0.494, $F(3,18)$ = 7.84, p = 0.001). By contrast, the lexical complexity model was not statistically significant (adjusted R^2 = −0.150, $F(3,18)$ = 0.09, p = 0.966), indicating that the lexical complexity measures did not significantly explain variation in QAB Overall scores.

The model-fit indices further showed that the comprehensive linguistic complexity model had the greatest explanatory power (adjusted $R^2 = 0.805$, $F(8,13) = 11.82$, $p < 0.001$), as well as the lowest AIC (64.82) and BIC (75.73) values. The morphological complexity model (AIC = 80.81, BIC = 85.17) and the syntactic complexity model (AIC = 82.93, BIC = 88.38) provided comparatively poorer fits. The lexical complexity model had the highest AIC (100.99) and BIC (106.45) values and therefore showed the poorest fit.

Overall, the comprehensive linguistic complexity model provided the strongest explanatory power and the best fit among the four models, whereas the morphological complexity model performed best among the three single-domain models.

Table 4. Comparison of the explanatory power of different linguistic complexity models for aphasia severity (QAB Overall)

Model	Predictors Included	R^2	Adjusted R^2	F	p	AIC	BIC
Syntactic Complexity Model	MLU, Clause Density, Embedded Clause Ratio	0.566	0.494	7.84	0.001	82.93	88.38
Lexical Complexity Model	TTR, Verb Diversity, Content Word Ratio	0.014	-0.150	0.09	0.966	100.99	106.45
Morphological Complexity Model	Verb Marking, Noun Marking	0.569	0.523	12.53	< 0.001	80.81	85.17
Full Model	All eight language measures	0.879	0.805	11.82	< 0.001	64.82	75.73

Note. R^2 = coefficient of determination; adjusted R^2 = adjusted coefficient of determination. F and p refer to the overall significance test for each model. AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion. Lower AIC and BIC values indicate better model fit.

4. Discussion

Using semi-spontaneous speech samples from 22 individuals with aphasia, this study systematically examined the explanatory power and predictive potential of eight measures spanning three dimensions of linguistic complexity—syntax, lexis, and morphology—in relation to aphasia severity, as indexed by QAB Overall scores. Regarding the first research question, the Pearson correlation analyses showed that only four of the eight measures were significantly and positively associated with QAB Overall: Mean Length of Utterance (MLU; $r = 0.734$, $p < 0.001$), verb marking ($r = 0.753$, $p < 0.001$), clause density ($r = 0.592$, $p = 0.004$), and embedded clause ratio ($r = 0.488$, $p = 0.021$). These positive associations indicate that better performance on these measures was related to higher QAB Overall scores and, consequently, milder aphasia. By contrast, none of the three lexical measures—Type–Token Ratio (TTR), verb diversity, and content-word ratio—nor noun marking reached statistical significance. Regarding the second research question, the morphological complexity model demonstrated the greatest explanatory power among the three single-domain models (adjusted $R^2 = 0.523$, $p < 0.001$), followed by the syntactic complexity model (adjusted $R^2 = 0.494$, $p = 0.001$). The lexical complexity model, however, was not statistically significant (adjusted $R^2 = -0.150$, $p = 0.966$). Regarding the third research question, the comprehensive linguistic complexity model, which included all eight measures across the three domains, yielded the greatest explanatory power (adjusted $R^2 = 0.805$, $p < 0.001$). It also provided a better fit than any of the three single-domain models, as indicated by its lower AIC (64.82) and BIC (75.73) values. These findings suggest that combining measures from multiple linguistic domains provides incremental explanatory value beyond that offered by any single domain alone. Taken together, the findings indicate that variation in aphasia severity cannot be attributed to impairment at a single linguistic level. Instead, it appears to reflect the combined influence of multiple levels of language processing, although syntactic, lexical, and morphological measures differ in their relative value for explaining individual differences in aphasia severity.

In the present study, MLU was strongly and positively correlated with QAB Overall scores (r = 0.734), indicating that longer utterances were associated with better overall language functioning and milder aphasia. MLU does not merely reflect utterance length; rather, it captures the integration of several levels of language processing. Producing longer utterances requires speakers to engage in more extensive planning by organizing the intended message, selecting appropriate lexical items, and formulating the syntactic structure before articulation [31]. MLU also reflects the overall capacity for grammatical organization, as longer utterances generally require the integration of more grammatical information within a single unit of speech. When the language system is broadly impaired in aphasia, utterances tend to become markedly shorter, providing a direct indication of constrained language-processing resources. This interpretation is consistent with Gordon's [11, 12] findings that MLU primarily functions as a measure of syntactic complexity and can effectively distinguish nonfluent aphasia from fluent or milder forms of aphasia. Clause density (r = 0.592) and embedded clause ratio (r = 0.488) were also significantly and positively correlated with QAB Overall scores. This finding accords with Kuvač Kraljević et al. [13], who found that clause density was significantly lower among individuals with aphasia than among neurologically healthy controls. Clause density reflects a speaker's ability to establish hierarchical relationships among multiple clauses within a single T-unit [13]. The embedded clause ratio, by comparison, captures the ability to produce embedded syntactic structures and thus serves as an indicator of higher-order syntactic complexity and syntactic integration. Cheon et al. [32] identified relative-clause formation and center embedding as two central syntactic operations involved in the production and comprehension of complex sentences, indirectly supporting the view that embedded clauses entail more advanced and demanding syntactic operations. Processing embedded clauses involves more than the linear addition of clauses. Crucially, it requires hierarchical processing through the construction of nested relationships between matrix and subordinate clauses [33]. This places considerable demands on the cognitive-linguistic processing system. Speakers must not only formulate multiple propositions but also encode the semantic relationships among them, including causality, temporal sequence, and conditionality, through appropriate syntactic structures. These operations require information to be maintained in working memory while syntactic information is integrated across clauses. Individuals with more severe aphasia may have insufficient processing resources to maintain and manipulate multiple clausal structures simultaneously. Consequently, they may produce fewer complex sentences and rely more heavily on simple clauses or coordinate structures. Of all eight measures, verb marking showed the strongest correlation with QAB Overall scores (r = 0.753), slightly exceeding that observed for MLU. Noun marking, however, was not significantly associated with QAB Overall. This pattern is consistent with Goodglass and Berko's [34] finding that nominal inflection is impaired less frequently than verbal inflection. The result may be explained by the processing demands inherent in verbal inflection. English verb inflection is not an isolated morphological operation but reflects the interaction of several components of the grammatical system. It first requires the selection of an appropriate lexical item, followed by the specification of tense and agreement features and, finally, the encoding of these abstract grammatical features in the appropriate word form. Verb marking therefore depends on coordinated processing across the lexical, morphological, and syntactic levels. When the language system is broadly compromised in aphasia, this multilevel coordination is particularly vulnerable to disruption, potentially resulting in the omission or incorrect realization of inflectional markers. Verb marking may therefore provide a sensitive indication of the integrity of the broader grammatical system. This interpretation is also consistent with Gordon's [12] factor-analytic finding that verb marking constitutes a central component of the "phrase building" factor. Taken together, the findings suggest that syntactic and morphological measures are more sensitive than the lexical measures examined here to variation in aphasia severity. One possible explanation is that syntactic and morphological processing depends more heavily on coordinated integration across multiple linguistic levels and therefore places greater demands on the overall integrity of the language system.

The linguistic domains differed markedly in their ability to explain variation in aphasia severity, and these differences may be understood in terms of the depth of cognitive-linguistic processing involved in each domain. The morphological complexity model provided the greatest explanatory power (adjusted $R^2 = 0.523$), driven primarily by verb marking. As discussed above, verbal inflection arises from coordinated processing across the lexical, morphological, and syntactic systems. It requires the speaker to select the appropriate lexical item, determine tense and agreement relations, and encode these abstract grammatical features in the corresponding word form. The strong performance of the morphological model may therefore be attributable to the fact that verb marking reflects the language system's overall capacity for coordinated integration. The syntactic complexity model showed the second-highest explanatory power (adjusted $R^2 = 0.494$). Its three measures—MLU, clause density, and embedded clause ratio—primarily assess sentence-level organization, namely, the ability to combine lexical items into sentences according to grammatical rules. Compared with verb marking, the processing demands captured by these syntactic measures are more concentrated at the level of syntactic planning and may depend less directly on morphological encoding. This may explain why the syntactic model performed slightly less well than the morphological model. The lexical complexity model, by contrast, did not significantly explain variation in QAB Overall scores (adjusted $R^2 = -0.150$, $p = 0.966$). This result should not be interpreted as evidence that lexical ability is unrelated to aphasia severity. Although word-retrieval difficulty, or anomia, is a pervasive feature of aphasia [35], the lexical measures used in this study—TTR, verb diversity, and content-word ratio—are susceptible to several sources of variation in spontaneous speech. TTR is particularly sensitive to sample length [36]. Although the duration of the speech samples was controlled by analyzing the first five minutes produced by each participant, the number of words generated within that period may still have varied according to participants' language abilities and aphasia severity, potentially reducing the sensitivity of TTR. The fixed narrative framework of the Cinderella retelling task may also have restricted the extent to which individual differences in lexical use could emerge. In addition, individuals with aphasia often employ compensatory strategies in natural discourse, such as using general terms, substituting accessible words, or producing circumlocutions when the intended word cannot be retrieved [37]. Such strategies may help sustain communication while masking underlying word-retrieval impairments, thereby reducing the sensitivity of discourse-based lexical measures. The failure of the lexical measures to account for aphasia severity in the present study is therefore more likely to reflect limitations associated with the measures and elicitation tasks than the absence of clinically meaningful lexical impairment.

The differences among the single-domain models also help explain the advantage of the comprehensive model. Language production depends on the coordinated operation of multiple subsystems rather than on the simple accumulation of separate linguistic components. In the present study, the comprehensive linguistic complexity model yielded greater explanatory power (adjusted $R^2 = 0.805$) than any single-domain model and was accompanied by substantial reductions in both AIC and BIC. This advantage likely arose because measures from different linguistic domains captured complementary rather than redundant information. The syntactic measures reflected sentence-level organization, or broader syntactic planning, whereas verb marking captured the efficiency of operations at the lexical–syntactic interface, including the local execution of morphosyntactic rules. Together, these measures assessed both whether a speaker could organize structurally complex sentences and whether the relevant grammatical rules could be implemented accurately, thereby providing a more comprehensive account of language functioning in aphasia. These findings have direct implications for clinical assessment. An evaluation restricted to a single linguistic domain—for example, naming performance or utterance length alone—may overlook other clinically relevant aspects of impairment. A more informative assessment should therefore incorporate measures from multiple domains, particularly syntactic and morphological complexity.

The present study has several limitations. First, the sample comprised only 22 individuals with aphasia, resulting in limited statistical power and potentially reducing the likelihood of detecting significant effects for the lexical measures. Second, the speech samples were obtained exclusively from native speakers of English. Whether the findings generalize to morphologically rich languages, such as German and Russian, or to isolating languages, such as Mandarin Chinese, remains to be determined. Future research should test these findings in larger independent samples representing a wider range of linguistic backgrounds.

5. Conclusion

Using naturalistic speech samples from 22 individuals with aphasia, this study systematically examined the explanatory power of eight measures of syntactic, lexical, and morphological complexity in relation to aphasia severity, as indexed by QAB Overall scores. MLU and verb marking showed the strongest associations with QAB Overall, with higher values corresponding to better overall language functioning and milder aphasia. Among the three single-domain models, the morphological complexity model provided the greatest explanatory power, whereas the lexical complexity model did not significantly explain variation in QAB Overall. The comprehensive model combining all three linguistic domains achieved the greatest explanatory power (adjusted R^2 = 0.805) and demonstrated substantial incremental explanatory value over the single-domain models. These findings support the use of a multidimensional linguistic complexity framework for assessing naturalistic discourse in aphasia and provide empirical evidence for improving clinical language assessment practices.

Acknowledgment

The author is grateful to the AphasiaBank team and all participants for making the APROCSA subcorpus publicly available.

References

[1] Goodglass, H. (1983). *The assessment of aphasia and related disorders*. https://www.las.ac.cn/front/book/detail?id=b96ac8032bdfdcb60e8876e793c8dc0d

[2] Marte, M. J., Russell-Meill, M., Carvalho Carvalho, N., & Kiran, S. (2025). *Charting the course of aphasia recovery: Factors, trajectories, and outcomes*. https://www.annualreviews.org/content/journals/10.1146/annurev-linguistics-011724-121245

[3] Stefaniak, J. D. (2020). *The neural and neurocomputational bases of recovery from post-stroke aphasia*. https://www.nature.com/articles/s41582-019-0282-1?utm_source=chatgpt.com

[4] Brady, M. C., Mills, C., Prag Øra, H., Novaes, N., Becker, F., Constantinidou, F., Flöel, A., Sunnerhagen, K. S., Isaksen, J., Jagoe, C., Jesus, L. M., Marangolo, P., Meinzer, M., van der Meulen, I., Campbell, P., Ho, L.,

Hussain, S., & Hilari, K. (2025). European Stroke Organisation (ESO) guideline on aphasia rehabilitation. *European Stroke Journal, 10*(4), 1189–1220. https://doi.org/10.1177/23969873241311025

[5] Hilari, K., Needle, J. J., & Harrison, K. L. (2012). What are the important factors in health-related quality of life for people with aphasia? A systematic review. *Archives of Physical Medicine and Rehabilitation, 93*(1, Supplement), S86–S95.e4. https://doi.org/10.1016/j.apmr.2011.05.028

[6] Prins, R., & Bastiaanse, R. (2004). Analyzing the spontaneous speech of aphasic speakers. *Aphasiology*. https://psycnet.apa.org/record/2004-21891-001

[7] Fromm, D., Katta, S., Paccione, M., Hecht, S., Greenhouse, J., MacWhinney, B., & Schnur, T. T. (2021). A comparison of manual versus automated quantitative production analysis of connected speech. *Journal of Speech, Language, and Hearing Research, 64*(4), 1271–1282. https://doi.org/10.1044/2020_JSLHR-20-00561

[8] Casilio, M., Kasdan, A. V., Bryan, K., Shibata, K., Schneck, S. M., Levy, D. F., Entrup, J. L., Onuscheck, C., de Riesthal, M., & Wilson, S. M. (2025). Four dimensions of naturalistic language production in aphasia after stroke. *Brain, 148*(1), 291–312.

[9] Beber, S., Bontempi, G., Miceli, G., & Tettamanti, M. (2025). The neurofunctional correlates of morphosyntactic and thematic impairments in aphasia: A systematic review and meta-analysis. *Neuropsychology Review, 35*(3), 483–516. https://doi.org/10.1007/s11065-024-09648-0

[10] Rashmi, J., John, S., Krishna, V., & Krishnan, G. (2025). Interventional research on morphosyntactic deficits in agrammatic aphasia: A narrative review. *Applied Neuropsychology: Adult*, 1–16. https://doi.org/10.1080/23279095.2025.2575264

[11] Gordon, J. K. (2020b). *How fluent? Part B. Underlying contributors to continuous measures of fluency in aphasia*. https://www.tandfonline.com/doi/full/10.1080/02687038.2020.1712586

[12] Gordon, J. K. (2020a). Factor analysis of spontaneous speech in aphasia. *Journal of Speech, Language, and Hearing Research, 63*(12), 4127–4147. https://doi.org/10.1044/2020_JSLHR-20-00340

[13] Kuvač Kraljević, J., Matić Škorić, A., & Lice, K. (2023). Main concepts in the spoken discourse of persons with aphasia: Analysis on a propositional and linguistic level. *Languages, 8*(2), 120. https://doi.org/10.3390/languages8020120

[14] Borovsky, A., Saygin, A. P., Bates, E., & Dronkers, N. (2007). Lesion correlates of conversational speech production deficits. *Neuropsychologia, 45*(11), 2525–2533. https://doi.org/10.1016/j.neuropsychologia.2007.03.023

[15] Holmes, D. I., & Singh, S. (1996). A stylometric analysis of conversational speech of aphasic patients. *Literary and Linguistic Computing, 11*(3), 133–140.

[16] Bastiaanse, R. (2003). Verb retrieval problems at the word and sentence level: Localisation of the functional impairments and clinical implications. *The Sciences of Aphasia* (131–148). Elsevier. https://www.sciencedirect.com/science/article/pii/B9780080440736500109

[17] Alyahya, R. S., Halai, A. D., Conroy, P., & Lambon Ralph, M. A. (2021). Content word production during discourse in aphasia: Deficits in word quantity, not lexical–semantic complexity. *Journal of Cognitive Neuroscience, 33*(12), 2494–2511.

[18] Casilio, M., Dutta, M., Bryan, K., Crouch, K., Dewall, Z., Feld, I., Kile, J., McMahan, K., Samora, J., Shibata, K., Steel, S., Kasdan, A. V., Walljasper, L., de Riesthal, M., & Wilson, S. M. (2026). A clinical training program for auditory-perceptual rating of connected speech in aphasia. *American Journal of Speech-Language Pathology, 35*(1), 52–68.

[19] Ezzes, Z., Schneck, S. M., Casilio, M., Mefford, A., de Riesthal, M., & Wilson, S. M. (2022). An open dataset of connected speech in aphasia with consensus ratings of auditory-perceptual features. *Data, 7*(11), 148.

[20] 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

[21] Armstrong, E., Ferguson, A., & Simmons-Mackie, N. (2013). Discourse and functional approaches to aphasia. *Aphasia and related neurogenic communication disorders*, 217–231.

[22] Bastiaanse, R., & Prins, R. S. (1994). Communicative speech therapy in aphasia: What does it mean, can it be effective and how should it be done? *Aphasiology, 8*(5), 482–488. https://doi.org/10.1080/02687039408248672

[23] Glosser, G., & Deser, T. (1991). Patterns of discourse production among neurological patients with fluent language disorders. *Brain and Language, 40*(1), 67–88.

[24] Brown, R. (1973). A first language: The early stages. *George Allen & Unwin*. London. https://www.kiphub.com/paper/66de622eabc76cf7a5b26024

[25] Johnson, W. (1939). *Language and speech hygiene: An application of general semantics.* Edwards Brothers.

[26] Park, H., Obermeyer, J., Kornisch, M., Hall, J., & Ontario, C. (2023). Semantic aspects of verb production in various discourse tasks in people with nonfluent aphasia. *American Journal of Speech-Language Pathology, 32*(5S), 2418–2429. https://doi.org/10.1044/2023_AJSLP-22-00293

[27] Alyahya, R. S. W., Conroy, P., Halai, A. D., & Lambon Ralph, M. A. (2022). An efficient, accurate and clinically-applicable index of content word fluency in aphasia. *Aphasiology, 36*(8), 921–939. https://doi.org/10.1080/02687038.2021.1923946

[28] 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

[29] Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective* (7th ed.). Pearson.

[30] R Core Team. (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. https://www.R-project.org/

[31] Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. The MIT Press. https://doi.org/10.7551/mitpress/6393.001.0001

[32] Cheon, K.-H., Kim, Y., Yoon, H.-D., Nam, K.-C., Lee, S.-Y., & Jeon, H.-A. (2020). Syntactic comprehension of relative clauses and center embedding using pseudowords. *Brain Sciences, 10*(4), 202.

[33] Chomsky, N. (2014). *Aspects of the theory of syntax, 50th anniversary edition*. MIT Press.

[34] Goodglass, H., & Berko, J. (1960). Agrammatism and inflectional morphology in English. *Journal of Speech and Hearing Research, 3*(3), 257–267. https://doi.org/10.1044/jshr.0303.257

[35] Papathanasiou, I., & Coppens, P. (2022). *Aphasia and related neurogenic communication disorders (6th ed.)*. https://catalog.nlm.nih.gov/discovery/fulldisplay/alma9917723983406676/01NLM_INST:01NLM_INST

[36] Cunningham, K. T., & Haley, K. L. (2020). Measuring lexical diversity for discourse analysis in aphasia: Moving-average type–token ratio and word information measure. *Journal of Speech, Language, and Hearing Research, 63*(3), 710–721. https://doi.org/10.1044/2019_JSLHR-19-00226

[37] Goodglass, H., & Wingfield, A. (1997). *Anomia: Neuroanatomical and cognitive correlates*. Academic Press.