A comparison of three discourse elicitation methods in aphasia and age-matched adults: implications for language assessment and outcome

Brielle C. Stark, PhD
Department of Speech and Hearing Sciences
Program in Neuroscience
Indiana University Bloomington
Bloomington, IN 47401

Communicating Author:

Brielle C. Stark, PhD

Office C177 200 S Jordan Avenue

Bloomington, IN 47405

770-548-7121

bcstark@iu.edu

Running head: language during discourse in aphasia and controls

Conflict of Interest: BC Stark declares no conflicts of interest.

Funding: No funding to report.

Abstract

Purpose: Discourse analysis is commonly used to assess language ability and to evaluate language change following intervention in aphasia. The purpose of this study was to identify differences in language produced during different discourse tasks in a large aphasia group and age- and education-matched control group.

Methods: Four structured discourse tasks across three discourse types (expositional, narrative and procedural) were evaluated in a group of adults with aphasia (N=90) and an age-matched control group (N=84) drawn from AphasiaBank. CLAN software was used to extract primary linguistic variables (mean length of utterance, propositional density, type-token ratio, words per minute, open-closed class word ratio, noun-verb ratio and tokens), which served as proxies for various language abilities. Using a series of repeated measures ANCOVAs, with significantly correlated demographic and descriptive variables as covariates, main effects of discourse type were evaluated.

Results: Despite an impoverished output from the aphasia group (i.e. the control group produced significantly more overall output), there was a main effect of discourse type on most primary linguistic variables in both groups, suggesting that, in adults with and without language impairments, each discourse type taxed components of the spoken language system to varying extents. Post hoc tests fleshed out these results, demonstrating that, for example, narrative discourse produced speech highest in propositional density.

Conclusion: Each discourse type taxes the language system in different ways, verifying the importance of using several discourse tasks, and selecting the most sensitive discourse tasks, when evaluating specific language abilities and outcomes.

Introduction

Although assessment of language abilities following aphasia often involves standardized testing batteries, linguistic discourse analysis provides a supplementary assessment that allows for the identification of impairments within a more naturalistic and ecologically valid domain (i.e connected speech), allowing for the detection of additional difficulties (e.g. syntactic organization, word order, cohesion and coherence) as well as compensatory strategies (e.g. circumlocution, self-cueing, retracing). Further, unlike assessment of single-word retrieval, such as object naming, assessment of discourse allows for evaluation of the independence and interaction of different linguistic processes (e.g. phonology, morphosyntax, lexical-semantics) (Prins & Bastiaanse, 2004). Indeed, a recent review by Bryant, Spencer and Ferguson (2016) noted that 86% of clinicians (of a 123-clincian sample drawn from five English-speaking countries) reported analyzing discourse within a clinical setting. Analyzing discourse at baseline and post-intervention serves an important purpose: assessing efficacy and generalization of impairment-based treatments. In a systematic review of discourse analysis in aphasia including 165 treatment studies, spoken discourse analysis was used in 87 studies: as a primary outcome measure in 36, a secondary outcome measure in 19 and as a measure of generalization in 37 (Bryant, Ferguson, & Spencer, 2016). However, the current state of research in spoken discourse analysis suffers from considerable flaws (Kintz & Wright, 2017). One flaw is the reliance on a single type of discourse elicitation method and another concerns the consistency and appropriateness of the linguistic variables measured from the piece of discourse. The current project specifically addresses these concerns.

Spoken language can be elicited using a variety of discourse methods (for a review, see Bryant et al 2016). The most common structured method is the expositional narrative, which provides subjects with a picture, or picture sequences, to describe. In the same systematic review of 165 studies, Bryant et al (2016) noted that, across studies using expositional narrative, most studies used only one type of exponential narrative to acquire a language sample. The majority of studies analyzed a single picture description, like the Cookie Theft Picture (Swinburn, Porter, & Howard, 2004). Another common method of discourse elicitation is narrative discourse, which involves the recounting of a personal story or the retelling of a well-known story, usually retold without visual aid. Narrative discourse, because of its reliance on memory and macrolinguistic structures like story grammar and coherence, may elicit more complex language compared to expositional narrative (MacWhinney, Fromm, Holland, Forbes, & Wright, 2010). However, because of narrative discourse does not typically use visual aids, this type of discourse method may be inherently more difficult, and therefore produced highly variable language, for those who rely more on visual cues for lexical-retrieval, as in severe aphasia. Finally, another common method of discourse elicitation is procedural discourse, or the describing of a procedure / task (“tell me how to make a peanut butter and jelly sandwich”). Procedural discourse is thought to elicit more action words and gestural communication (Pritchard, Dipper, Morgan, & Cocks, 2015).

However, of the studies evaluating discourse in aphasia, few analyze more than one of these types of discourse, despite suggestion from Brookshire & Nicholas (1994) that a combination of multiple structured language samples should be used because they generate a comprehensive language sample most resemblant of actual language use. Additionally, given the variability commonly shown in aphasic performance within and across sessions, analyzing language from a single discourse method may not demonstrate the breadth of an individual’s linguistic ability or be sensitive to linguistic changes post-intervention. An additional issue concerns the consistency and appropriateness of the linguistic variable(s) being measured. When analyzing spoken language from a single type of discourse method, the extent to which different parts of the language system are taxed is likely highly dependent on the discourse method. For example, as stated above, narrative discourse may produce more syntactically complex language than procedural discourse. Therefore, there is a need to identify the extent to which each discourse method taxes parts of the spoken language system. Doing so has direct implications for research and clinical settings. Identifying discourse method(s) most sensitive to commonly measured aspects of the spoken language system will improve our ability to accurately and sensitively detect impairments, which is directly relevant to measuring treatment (and generalization) outcomes. Further, we must also consider that the restraints on assessment and research time make it difficult to acquire and subsequently analyze many discourse samples. Indeed, while acquisition time is usually short, it is estimated that transcription and coding requires 6-12 minutes of time per minute of language sample (Boles, 1998). Therefore, an optimal solution is to identify which discourse method most sensitively taxes the property of spoken language one is most interested in measuring, thereby establishing which discourse method is most appropriate for sensitively evaluating abilities and impairments as well as intervention outcomes.

Here, we have selected primary linguistic variables that serve as proxies for different aspects of spoken language. We subsequently analyzed these primary linguistic variables across four different discourse tasks, comprising three different discourse types. We relied on data from a large database, AphasiaBank, to generate a sufficiently large sample size with which to compare language derived from different discourse elicitation methods across an age- and education-matched control group and an aphasia group (MacWhinney, Forbes, & Holland, 2011). The intent of the current project is to understand the extent to which each discourse task taxes components of the spoken language system.

Methods

Participants.
The AphasiaBank database was used to acquire spoken discourse data from adults with aphasia and controls. Data was extracted from persons with aphasia (<93.8 aphasia quotient on the Western Aphasia Battery [Kertesz, 2007]) who also had samples from all four discourse tasks (described in next section). There was N=90 included in this group (see Table 1 for demographics, including age, education, aphasia quotient and duration of aphasia). The cause of aphasia was largely stroke. Anomic aphasia was present in 42 participants; Broca's, 18; conduction, 18; transcortical motor, five; transcortical sensory, one; and Wernicke's, six. Clinical impression of the presence or absence of apraxia of speech was evaluated in 80 of the 90 participants and was present in 30; dysarthria was evaluated in 79 participants and was present in 14.

Spoken discourse data was extracted from an original N=185 participants in the control group (99 females; mean age=63.49±17.41, range 23-90). To closely match the control group to the aphasia group, case-control matching in SPSS 25 was employed, matching groups on the variables of sex, age and education. Controls were matched to aphasia participants with a FUZZY interval of one standard deviation of the aphasia group's mean age and education. Following this matching, the final control group consisted of 84 participants (see Table 1 for demographics).

Discourse elicitation.
Spoken language was analyzed during four discourse tasks. Tasks from three different types of discourse elicitation method were chosen: expositional, narrative and procedural. All discourse samples were prompted using AphasiaBank protocol (MacWhinney et al., 2011) (https://aphasia.talkbank.org/protocol/).

The procedural discourse involved participants telling the investigator "how to make a peanut butter and jelly sandwich" which we have here abbreviated as 'PBJ.' The narrative discourse was a story retelling, specifically the Cinderella story, which we call 'Cinderella.' Finally, the expositional discourse involved two tasks. The first was a picture sequence description, which involved describing a storyboard of four pictures in which a boy kicks a ball through a window, which we here call the Broken Window story, or 'BW.' The second was a picture description, which involved participants describing a single picture of a cat being rescued from a tree by firemen, which we here call the Cat Rescue story, or 'Cat.'

Discourse analysis.
Spoken discourse production was audio recorded (most included video) for each participant at the respective data collection site. Transcripts were coded by trained AphasiaBank raters using the CHAT software (MacWhinney, 2000). Transcripts captured many aspects of spoken language production, including utterances, dysfluencies and word- and utterance-level errors. As the CHAT transcriptions were coupled to the accompanying video, duration (in seconds) was automatically calculated.

CLAN (v23, downloaded 12/21/2018), the accompanying analysis program to CHAT, was used to extract linguistic data from the transcripts. CLAN relies on the parsing of morphological and grammatical information by an automatic process called *mor*, which tags parts of speech within each utterance. Specifically, the following commands were used:

CLAN Command	Result
mor	Tag parts of speech automatically using mor script
*eval +t*PAR +u*	Evaluate transcripts to derive primary linguistic outcome variables • eval: evaluate microlinguistic information using the *mor* tier

- +t*PAR: evaluate only the participant tier
- +u: consolidate all files to single output

The goal of this project was to extract variables that served as proxies for different components of the spoken language production system. Therefore, variables that touched on language fluency, syntactic variation and quality and diversity of output were selected. These variables are described in Table 2, and include: mean length of utterance (measured in morphemes); verbs per utterance; type-token ratio; propositional density (Fromm et al., 2016); noun-verb ratio and open-closed class word ratio. Total tokens were also extracted (i.e. total words, which did not include repetitions, retracings or paraphasias, but which did include paraphasia targets [the word being replaced by the paraphasia]).

Analysis.
We had three lines of inquiry for this study.

I. Evaluation of primary linguistic variables in control group
Goal: To determine the extent to which primary linguistic variables were different between discourse types in a group with no language impairment.

Demographic variables (age, education) and average tokens (across tasks) were correlated with primary linguistic variables of interest (also averaged across tasks) (Table 3, part A). Correlations were corrected for multiple comparisons using Benjamini-Hochberg correction ($p<.05$). Significantly correlated factors were then used as covariates in a sequence of repeated measures ANCOVA analyses. For significant main effects, post hoc analyses adjusted for multiple comparisons using Bonferroni were conducted.

II. Evaluation of primary linguistic variables in aphasia group
Goal: To determine the extent to which primary linguistic variables were different between discourse types in a group with language impairment.

Demographic variables (age, education, aphasia quotient, presence of apraxia of speech, presence of dysarthria, duration of aphasia) and average tokens (across tasks) were correlated with primary linguistic variables of interest (also averaged across tasks) (Table 3, part B). Correlations were corrected for multiple comparisons using Benjamini-Hochberg correction ($p<.05$). Significantly correlated factors were then used as covariates in a sequence of repeated measures ANCOVA analyses. For significant main effects, post hoc analyses adjusted for multiple comparisons using Bonferroni were conducted.

III. Comparison of primary linguistic variables between control and aphasia group
Goal: To compare primary linguistic variables across discourse tasks between aphasia and control groups.

Using appropriate statistical tests (parametric or non-parametric, based on Levene's Test for Homogeneity of Variances), primary linguistic variables were compared between the aphasia and control groups to identify significant differences in language ability between the groups. To compare the extent to which discourse tasks differed in their elicitation of primary linguistic variables (i.e. the results highlighted in sections I and II, described above), were compared between aphasia and control groups.

Results

Differences in demographics across groups

As expected, due to case-control matching, there was no significant difference in age (t(172)=.92, p=.36) or education (t(160.99)=.796, p=.42) between the control and aphasia groups. Pearson's Chi Square test showed no significant differences in the number of females and males between the control and aphasia group, $\chi^2(1, N = 174)=.007$, p=.933.

Main behavioral results

The repeated measures ANCOVAs were based on significant correlations between average primary linguistic variables (across discourse tasks) and demographic and descriptive variables; the significant results for each group are shown in Table 3.

I. Evaluation of primary linguistic variables in control group

Evaluation of differences in primary linguistic variables in the control group identified how spoken language elicited by adults with no language impairment varied across discourse tasks (full ANCOVA results, Table 4; post hoc and summary of results, Table 5; visualization of ANCOVA results, Figure 1). The repeated measures ANCOVAs demonstrated that there was a main effect of discourse type (expositional [BW, Cat], narrative [Cinderella], procedural [PBJ]) on most primary linguistic variables with the exception of open-closed class word ratio and WPM. Namely, there was a main effect of discourse type on MLUs, verbs per utterance, propositional density and noun-verb ratio with no significant interactions (as no covariates significantly correlated with these variables) (Table 4). There was a significant interaction of tokens with TTR, such that, alongside a main effect of discourse type, production of more tokens correlated with a reduced TTR.

Post hoc analyses showed that each discourse type (e.g. narrative) elicited language that differs across linguistic components (e.g. density, tokens). These post hoc analyses are more comprehensively explained in Table 5. Briefly, the control group results demonstrated that procedural discourse elicited the shortest MLU, highest noun-verb ratio and the fewest verbs per utterance, indicating that this type of discourse was the most syntactically impoverished. Further, post hoc analyses suggested that narrative discourse elicited the densest language and most tokens and that expositional discourse (BW, Cat) tended to elicit the most diverse language (TTR).

II. Evaluation of primary linguistic variables in aphasia group

While evaluation of differences in primary linguistic variables in the control group identified how each discourse type was sensitive to producing certain primary linguistic components, we also quantified how linguistic components differed across discourse types in adults with aphasia (full ANCOVA results, Table 4; post hoc and summary of results, Table 5; visualization of ANCOVA results, Figure 1). In addition, this analysis also leant insight into significant interactions with covaried demographic and descriptive variables.

The repeated measures ANCOVAs demonstrated that there was a main effect of discourse type (expositional [BW, Cat], narrative [Cinderella], procedural [PBJ]) on most primary linguistic variables with the exception of MLU, open-closed class word ratio and tokens (Table 4). There was a main effect of discourse type with no significant interactions for propositional density and TTR. Alongside a main effect of discourse type on verbs per utterance, there was also a significant interaction with age observed, such that greater age correlated with more verbs per utterance. Similarly, alongside a main effect of discourse type on WPM, there was a significant interaction with both tokens and age, where more WPM correlated with more tokens and greater age. The presence of apraxia of speech interacted with noun-verb ratio alongside a main effect of discourse type. Finally, despite no main effect of discourse type on number of tokens produced, there was a significant interaction of tokens with aphasia severity (AQ), where more tokens were produced in those with milder aphasia regardless of task.

These results complement the results of the control group, suggesting that each discourse type is sensitive to producing specific aspects of spoken language. Additionally, the aphasia group results highlight mediating factors (e.g. aphasia severity, age) that interact with discourse type. Post hoc analyses in the aphasia group also complement evidence from the control group, in that procedural discourse (PBJ) elicited the fewest verbs per utterance and highest noun-verb ratio (Table 5). However, post hoc analyses in the aphasia group elaborated on this finding, suggesting that these main effects are often mediated by other variables, such as age and presence of AOS, respectively. Like the control group, narrative discourse (Cinderella) was the densest type (though not significantly denser than PBJ, which was a finding unique to the aphasia group), and expositional discourse (BW in particular) tended to have the highest TTR.

III. Comparison of primary linguistic variables between control and aphasia group

The control group elicited spoken language that was significantly different from the aphasia group across nearly all primary linguistic variables, except for TTR during expositional and narrative tasks and noun-verb ratio and open-closed class word ratio across all tasks (Table 6). Therefore, the aphasia group had (expectedly) impoverished spoken output compared with the control group. One can best appreciate this comparison in Figure 2, indicating the difference between aphasia and control groups for all primary linguistic measures. Sections I and II demonstrated that discourse types elicit consistent language properties in both control and aphasia groups. This finding is particularly striking, especially as the control group produced significantly more output than the aphasia group.

As was discussed in the introduction, most studies employing discourse as an outcome measure use single picture description to evaluate language production. In this study, we employed two expositional discourse tasks (BW, Cat), and were therefore able to directly compare primary linguistic data between tasks. Using a sequence of independent t-tests corrected for multiple comparison correction (Benjamini-Hochberg), we established that the BW picture sequence and the Cat picture description elicited significantly different linguistic properties. In the control group, the Cat picture elicited longer MLU ($p=.001$, corrected $p=.002$), more tokens ($p<.001$), more verbs per utterance ($p=.001$, corrected $p=.002$) and more WPM ($p<.001$). However, the BW picture sequence elicited a higher TTR ($p<.001$), greater density ($p=.006$, corrected $p=.008$) and greater open-closed class word ratio ($p=.037$, corrected $p=.042$). There was not a significant difference in noun-verb ratio between the two expositional tasks ($p=.07$). When comparing linguistic data between BW and Cat in the aphasia group, the pattern was similar: the Cat picture elicited longer MLU ($p<.001$), more tokens ($p<.001$) and more verbs per utterance ($p<.001$) while the BW picture sequence elicited higher TTR ($p<.001$) and higher open-closed class word ratio ($p<.001$). However, unlike in the control group, BW picture sequence elicited more WPM ($p<.001$) while the Cat picture elicited language that was denser ($p<.001$) and had a higher noun-verb ratio ($p<.001$). In summary, both expositional types of discourse may be useful for evaluating spoken language, but as prior results show, neither is the most ideal for evaluating density but may be useful when evaluating diversity (e.g. TTR), and perhaps syntactic complexity (verbs per utterance, noun-verb ratio and open-closed class word ratio).

IV. Summary

Overall, results from within each group (sections I and II) and comparison across groups (section III) demonstrate that spoken language elicited by each discourse task varies across linguistic properties in both control and aphasia groups, even with an impoverished output (as in aphasia group).

Discussion

In the current study, a large population of age- and education-matched adults and persons with aphasia was leveraged to evaluate the extent to which spoken language elicited by four different discourse tasks (spanning three discourse types) varied.

As expected, the control group produced more output across discourse tasks, aligning with prior studies comparing proxies of language, such as noun-verb ratio and type-token ratio, between age-matched control and aphasia groups (Fergadiotis & Wright, 2011; Thompson et al., 2012). Further, the control group established that each discourse type elicited spoken language that differed in underlying linguistic properties. For example, narrative discourse was shown to elicit the densest speech, while procedural discourse, at least as measured here by the peanut butter and jelly story prompt, may exhibit the least syntactically complex spoken language. Strikingly, despite impoverished output across all discourse tasks, the aphasia group exhibited similar data, supporting the conjecture that narrative discourse elicits the densest language and procedural the least syntactically complex language. Evaluating the aphasia group also demonstrated that significant interactions with variables such as aphasia severity, tokens, age and presence of apraxia of speech are liable to have a mediating effect on primary linguistic variables measured from spoken language. This is, of course, not surprising given the large amount of research suggesting an interaction of aphasia severity and spoken language output (Bond, Ulatowska, Macaluso-Haynes, & May, 1983). However, considering the shared findings from the control and aphasia groups, this study's results present a compelling argument that the type of discourse used to elicit spoken language matters.

Indeed, each discourse type elicited spoken language that differed across linguistic properties. Lexical diversity, commonly measured by type-token ratio but also by related entities such as moving-average type-token ratio, has before been identified to differ across discourse types in people with aphasia (Fergadiotis & Wright, 2011; Fergadiotis, Wright, & Capilouto, 2011) and our results concur and expand on these findings. Put together, our findings and prior findings establish that, in both people with and without aphasia, each structured discourse method has its benefits for the production of certain aspects of spoken language. In summary, data suggested that propositional density was greatest in the narrative discourse; words per minute were reduced in the narrative discourse; type-token ratio was smallest for narrative discourse; and procedural discourse produced the least syntactically complex speech (highest noun-verb ratio, fewest verbs per utterance).

There are caveats to the current study. For example, a main effect of discourse elicitation method on propositional density for all groups was identified. The narrative discourse (Cinderella story) elicited the densest language, producing the most content-rich language regardless of aphasia severity or number of tokens. Narrative discourse may not prove to be the most sensitive to noun-access or object naming (indeed, narrative discourse produced the smallest percentage of nouns across discourse types) but may prove the most sensitive elicitation method for evaluating depth of vocabulary and content richness, and therefore may best evaluate impairments or changes in these language areas. Narrative discourse, unlike expositional discourse, does not rely on visual aids and relies more heavily on processes of memory (arguably, both long-term and working memory) as well as aspects of executive function, such as planning and organization. Therefore, in participants with concomitant impairments in these cognitive domains, narrative discourse may not produce the densest language. AphasiaBank does not collect cognitive scores, and we were therefore unable to explore the extent to which individuals did or did not demonstrate impairments in memory or executive function. However, given the large sample size of people with aphasia, who demonstrated that narrative discourse produced the densest language, this suggests that at least a large proportion of the greater aphasia population will likely also elicit the densest speech during this task. However, future work should evaluate the importance of measures of cognition on linguistic structures, like those explored here, as well as on macrolinguistic variables, such as story grammar (Nicholas & Brookshire, 1995; Roth & Spekman, 1986), coherence (Galetto, Kintz, West, Mrini, & Wright, 2013; Rogalski, Altmann, Plummer-D'Amato, Behrman, & Marsiske, 2010; Van Leer & Turkstra, 1999; Wright, Koutsoftas, Capilouto, & Fergadiotis, 2014) and production of main concepts (Richardson & Dalton, 2015). We did not evaluate the difference in macrolinguistic variables in the current study, but a complementary area of future research should identify the relative differences in macrolinguistic variables across discourse elicitation methods and the impact of aphasia severity on these differences. As a final caveat, the current aphasia sample consisted of 68 with fluent aphasia (conduction, transcortical sensory, Wernicke's, anomic) and 22 with non-fluent aphasia (Broca's, transcortical motor), and the present results may have been influenced by the larger proportion of people with fluent and/or milder aphasia, who have a tendency to produce more language than those with non-fluent and/or more severe aphasia.

Bryant et al (2016) discuss how most studies utilizing discourse in aphasia employ the structured, expositional discourse elicitation method (overwhelmingly, a single picture description). This study directly evaluated linguistic variables from two types of expositional discourse tasks, a picture description (Cat) and a picture sequence description (BW). The control and aphasia group both showed that the Cat picture description tended to elicit longer MLU, more tokens and more verbs per utterance and the BW picture sequence tended to elicit higher TTR and open-closed class word ratio. The aphasia group, unlike the control group, showed that the BW picture sequence elicited more WPM while the Cat picture elicited language that was denser and had a higher noun-verb ratio. The Cat and BW, while perhaps not eliciting the densest speech, may therefore be particularly helpful for those with more severe aphasia, who may rely more on visual cues for lexical retrieval (Doyles et al., 1998).

This study, which used a large dataset of age- and education-matched control and aphasia data, specifically evaluates language produced during different structured discourse types. First, the results of the current study emphasize the importance of utilizing several types of structured discourse elicitation methods for a more comprehensive evaluation of language, demonstrating that the microlinguistic properties of spoken language elicited by each type of method is different. This finding lends support for the use of several types of discourse elicitation methods to acquire a comprehensive evaluation of language (Brookshire & Nicholas, 1994) which is important for the sensitivity of assessment and intervention. Word error (i.e. paraphasia) instability across discourse samples (particularly, short samples elicited using pictures, as in expositional tasks) also demands employing multiple discourse types to achieve a more reliable profile of paraphasias (Boyle, 2015). Second, we lay the foundation for more accurately selecting appropriate discourse elicitation methods to evaluate specific aspects of language. This is particularly relevant in clinical settings, when both acquisition and analysis time are a factor. The importance of choosing the most sensitive discourse type(s), according to your language variable(s) of interest, can be easily understood by an analogy to a common medical situation: measuring blood pressure. For example, if you want to measure blood pressure, you'd acquire the most accurate and sensitive result from using the most appropriate measure (e.g. a well-fitted blood pressure cuff) than a less appropriate measure (e.g. a too-big blood pressure cuff). This is the case in language research. If you want to examine a person's syntactic system in the most comprehensive fashion, one must use a discourse type that elicits spoken language that tends to most expansively tax this system. We show here, for example, that using a procedural discourse (alone) to measure syntactic complexity may not be the most sensitive discourse type for taxing syntactic processes. Further, understanding the microlinguistic properties of spoken language produced by each type of discourse method is important for tailoring assessments to better appreciate the depth of specific language impairments and compensatory techniques, which may not be elicited to the same degree during standardized batteries or single-modality assessment (e.g. object naming). In addition, selecting the most appropriate discourse elicitation method may be especially important when identifying generalization of impairment-based training to more functional settings. For example, discourse-level improvement (or indeed, generalization of intervention) in aphasia following a syntactic impairment-based treatment may be best exemplified by assessing verbs per utterance, open-closed class word ratio and/or noun-verb ratio elicited during narrative and expositional discourse. Meanwhile, improvements to spoken language density and productivity in aphasia following response elaboration training (Kearns, 1985) or semantic feature analysis / related therapies (Lowell, Beeson, & Holland, 1995; Ylvisaker & Szekeres, 1985) may be most sensitively evaluated using narrative discourse. Third, we suggest that evaluating tokens as the primary linguistic outcome from spoken discourse in aphasia is likely not a telling or sensitive measure of linguistic ability because of its relationship with overall severity of aphasia. This suggestion is of particular relevance for the field, as Bryant et al (2016) showed that most structured discourse studies evaluated tokens as the linguistic variable of interest. We directly evaluated the difference in tokens across discourse types in aphasia and did not find a main effect of discourse task, but did find a significant interaction of tokens with aphasia severity. This result suggests that analyzing tokens as the primary linguistic variable is highly dependent on the severity of aphasia and likely not sensitive to type of discourse task. Further, evaluating tokens alone does not provide information regarding other aspects of language that may be responding to treatment or that are relatively unimpaired despite aphasia severity (e.g. syntactic ability), which undermines the primary reason for using discourse in assessment or as an outcome measure.

Despite the ecological validity of using spoken discourse to assess language ability and outcomes in aphasia, and indeed a well-used one, a variety of methodological issues have hampered its potential in research and clinical environments. To advance the field of discourse assessment, we provide compelling evidence for the use of multiple discourse task types as well as evidence for selection of the most appropriate task for analysis of specific aspects of spoken language, useful in both assessment and treatment outcome. We believe that these suggestions will improve upon some of the methodological issues in discourse assessment, particularly in improving the sensitivity of this method in identifying specific language abilities and outcomes.

Running head: language during discourse in aphasia and controls

Acknowledgments: Alexandra Basilakos, PhD CCC-SLP for her comments on this manuscript.

References

Boles, L. (1998). Conversational discourse analysis as a method for evaluating progress in aphasia: a case report. *Journal of Communication Disorders*, *31*, 261–274.

Bond, S., Ulatowska, H., Macaluso-Haynes, S., & May, E. (1983). Disourse Production in Aphasia: Relationship to Severity of Impairment. In *Proceedings of the Clinical Aphasiology Conference* (pp. 202–210).

Boyle, M. (2015). Stability of Word-Retrieval Errors With the AphasiaBank Stimuli. *American Journal of Speech-Language Pathology*, *24*(4). https://doi.org/10.1044/2015_AJSLP-14-0152

Brookshire, R., & Nicholas, L. (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.

Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics & Phonetics*, *30*(7), 489–518. https://doi.org/10.3109/02699206.2016.1145740

Doyles, P. J., Mcneil, M. R., Spencer, K. A., Goda, A. J., Cottrell, K., & Lustig, A. P. (1998). The effects of concurrent picture presentations on retelling of orally presented stories by adults with aphasia. *APHASIOLOGY*, *12*(8), 561–574. https://doi.org/10.1080/02687039808249558

Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology*, *25*(11), 1414–1430. https://doi.org/10.1080/02687038.2011.603898

Fergadiotis, G., Wright, H. H., & Capilouto, G. J. (2011). Productive vocabulary across discourse types. *Aphasiology*, *25*(10), 1261–1278. https://doi.org/10.1007/s11103-011-9767-z.Plastid

Fromm, D., Greenhouse, J., Hou, K., Russel, G., Cai, X., Forbes, M., … MacWhinney, B. (2016). Automated Proposition Density Analysis for Discourse in Aphasia. *Journal of Speech, Language and Hearing Research*, *59*, 1123–1132. https://doi.org/10.1044/2016

Galetto, V., Kintz, S., West, T., Mrini, A., & Wright, H. (2013). Measuring global coherence in aphasia. *Procedia-Social and Behavioral Sciences*, *94,* 198–199.

Kearns, K. (1985). Response elaboration training for patient initiated utterances. *Clinical Aphasiology*, 15–25. Retrieved from http://eprints-prod-05.library.pitt.edu/854/1/15-25.pdf

Kertesz, A. (2007). *Western Aphasia Battery - Revised*.

Kintz, S., & Wright, H. H. (2017). Discourse measurement in aphasia research. *Aphasiology*, *00*(00), 1–3. https://doi.org/10.1080/02687038.2017.1398807

Lowell, S., Beeson, P. M., & Holland, A. L. (1995). The Efficacy of a Semantic Cueing Procedure on Naming Performance of Adults with Aphasia. *American Journal of Speech-Language Pathology*. https://doi.org/10.1044/1058-0360.0404.109

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analysing Talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

MacWhinney, B., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for Studying Discourse. *Aphasiology*, *25*(11), 1286–1307. https://doi.org/10.1080/02687038.2011.589893.AphasiaBank

MacWhinney, B., Fromm, D., Holland, A., Forbes, M., & Wright, H. (2010). Automated analysis of the Cinderella story. *Aphasiology*, *24*(0), 856–868. https://doi.org/10.1080/02687030903452632

Nicholas, L., & Brookshire, R. (1995). Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. *Journal of Speech, Language and Hearing Research*, *38*, 145–156.

Prins, R., & Bastiaanse, R. (2004). Review: Analysing the spontaneous speech of aphasic speakers. *Aphasiology*, *18*(12), 1075–1091. https://doi.org/10.1080/02687030444000534

Pritchard, M., Dipper, L., Morgan, G., & Cocks, N. (2015). Language and iconic gesture use in procedural discourse by speakers with aphasia. *Aphasiology*, *29*(7), 37–41. https://doi.org/10.1080/02687038.2014.993912

Richardson, J. D., & Dalton, S. G. (2015). Main concepts for three different discourse tasks in a large non-clinical sample. *Aphasiology*, *7038*(June), 1–29. https://doi.org/10.1080/02687038.2015.1057891

Rogalski, Y., Altmann, L., Plummer-D'Amato, P., Behrman, A., & Marsiske, M. (2010). Discourse coherence and cognition after stroke: A dual task study. *Journal of Communication Disorders*, *43*(3), 212–224. https://doi.org/10.1016/j.jcomdis.2010.02.001.Discourse

Roth, F., & Spekman, N. (1986). Narrative discourse: Spontaneously generated stories of learning-disabled and normally achieving students. *Journal of Speech, Language and Hearing Research, 51*, 8–23.

Swinburn, K., Porter, G., & Howard, D. (2004). *Comprehensive Aphasia Test*. East Sussex: Psychology Press.

Thompson, C. K., Cho, S., Hsu, C., Wieneke, C., Weitner, B. B., Mesulam, M. M., … Weintraub, S. (2012). Dissociations between fluency and agrammatism in primary progressive aphasia, *7038*. https://doi.org/10.1080/02687038.2011.584691

Van Leer, E., & Turkstra, L. (1999). The effect of elicitation on discourse coherence and cohesion in adolescents with brain injury. *Journal of Communication Disorders, 32*, 327–349.

Wright, H. H., Koutsoftas, A. D., Capilouto, G. J., & Fergadiotis, G. (2014). Global coherence in younger and older adults: Influence of cognitive processes and discourse type. *Aging, Neuropsychology, and Cognition, 21*(2), 174–196. https://doi.org/10.1080/13825585.2013.794894

Ylvisaker, M., & Szekeres, S. (1985). Cognitive-language intervention with brain-injured adolescents and adults. In *Annual Convention of the Illinois Speech-Language-Hearing Association*.

Figure Captions

Figure 1: Primary linguistic variables compared within groups (Aphasia, Control) across discourse types. Raw means from aphasia group (left column) and control group (right column). Black lines indicate significant differences, as described in Tables 4 and 5. Error bars represent standard deviations. *MLU = mean length of utterance; TTR = type-token ratio; WPM = words per minute, divided by 10. BW = Broken Window, PBJ = Peanut Butter and Jelly.*

Figure 2: Comparison of primary linguistic variables per discourse type for aphasia and control groups. Raw means from aphasia group (blue columns) and control group (orange columns), categorized by discourse type (BW, Cat, Cinderella, PBJ). Significant differences between control and aphasia groups are shown by a star. Error bars represent standard deviations. This figure is a complement to Table 6. *MLU = mean length of utterance; TTR = type-token ratio; WPM = words per minute, divided by 10. BW = Broken Window, PBJ = Peanut Butter and Jelly.*

Table 1: Demographics of control and aphasia groups.

AQ = Aphasia Quotient from the Western Aphasia Battery (Kertesz, 2007)

Group	**Sex**	**Age (yrs)** **Mean (SD)** **Min - Max**	**Education (yrs)** **Mean (SD)** **Min - Max**	**AQ** **Mean (SD)** **Min - Max**	**Duration of aphasia (yrs)** **Mean (SD)** **Min - Max**	**Average tokens** **Mean (SD)** **Min - Max**	**Average duration (sec)** **Mean (SD)** **Min-Max**
Control N=84	36 F 48 M	66.75 (15.51) 30-88	15.91 (2.31) 12-22			209.63 (117.63) 38-680	92.21 (48.41) 22.67-284.33
Aphasia N=90	38 F 52 M	68.92 (15.58) 30-93	15.59 (2.84) 12-25	74.78 (14.86) 28.20-93.40	5.77 (4.23) .75 – 24.70	99.54 (69.24) 8.33 – 376.00	117.91 (90.48) 13.33-536.00

Table 2: Primary linguistic outcome measurements. Data derived from the CLAN software (MacWhinney et al., 2010).

Primary measure	**Definition**	**Proxy of language component**
Mean length of utterance (MLU)	Average number of morphemes per utterance	Linguistic productivity
Propositional density	Number of verbs, adjectives, adverbs, prepositions, and conjunctions divided by the total number of words	Content richness
Words per minute (WPM)	Total number of tokens divided by the duration (converted from seconds to minutes)	Fluency
Verbs per utterance	Average number of verbs per utterance; includes verbs, copulas, and auxiliaries followed by past or present participles and does not include modals	Syntactic complexity
Type-token ratio (TTR)	Number of different words (types) divided by the number of words (tokens)	Lexical diversity
Open-closed	Ratio of open class words (all nouns, all verbs excluding auxiliaries and modals, all adjectives, all adverbs) divided by closed class words (all other classes)	Syntactic complexity
Noun-verb ratio	Ratio of nouns to verbs (excluding auxiliaries and modals)	Syntactic complexity
Tokens	Total number of words produced	Word retrieval, gross output

Table 3. Correlations of demographic and descriptive variables with primary linguistic variables. Demographic and descriptive variables included age (years), AQ (aphasia severity, represented by the aphasia quotient from the Western Aphasia Battery [Kertesz, 2007], education (years), months post-stroke, presence of apraxia of speech and of dysarthria of speech and average tokens (across all discourse tasks). Primary linguistic variables have been averaged across all discourse tasks. In some cases, a data-point for a discourse task was missing; total subjects evaluated for each correlation are shown in the 'N' rows for every variable. P-values shown here have been corrected for multiple comparisons using Benjamini-Hochberg correction, and bolded p-values are those which are significant at $p<0.05$.
MLU = mean length of utterance; TTR = type-token ratio; WPM = words per minute

Table 3A. Correlations for the control group (N=84).

	Variables	**Age**	**Education**	**Average Tokens**	**Average MLU**	**Average Verbs per Utterance**	**Average Density**	**Average TTR**	**Average WPM**	**Average Noun/Verb Ratio**	**Average Open/Closed Class Word Ratio**
Variables	*Statistics*	*I*	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*	*X*
	r	1	0.039	-0.063	-0.101	-0.189	-0.034	0.028	-0.293	-0.076	-0.323
I	p-value		0.798	0.798	0.798	0.258	0.798	0.798	**0.032**	0.798	**0.027**
	N	84	84	84	84	84	84	84	84	84	84
	r		1	0.123	0.095	0.086	0.022	0.08	-0.04	-0.004	0.095
III	p-value			0.846	0.846	0.846	0.948	0.846	0.931	0.972	0.846
	N		84	84	84	84	84	84	84	84	84
	r			1	0.047	0.069	0.234	-0.718	0.157	-0.223	0.022
III	p-value				0.756	0.732	0.123	**0**	0.344	0.123	0.842
	N			84	84	84	84	84	84	84	84
	r				1	0.902	0.172	-0.186	0.089	0.206	0.129
IV	p-value					**0**	0.266	0.266	0.474	0.266	0.437
	N				84	84	84	84	84	84	84
	r					1	0.227	-0.181	0.115	-0.066	0.151
V	p-value						0.171	0.225	0.449	0.552	0.308
	N					84	84	84	84	84	84
	r						1	-0.116	0.194	-0.184	-0.001
VI	p-value							0.438	0.209	0.209	0.993
	N						84	84	84	84	84
	r							1	-0.149	0.213	0.151
VII	p-value								0.264	0.225	0.264
	N							84	84	84	84
	r								1	-0.033	-0.08
VIII	p-value									0.763	0.6
	N								84	84	84
	r									1	0.326
IX	p-value										**0.018**
	N									84	84
	r										1
X	p-value										
	N										84

Table 3B. Correlations for the aphasia group (N=90).

	Variables	Age	AQ	Education	Months post-stroke	Presence of Apraxia of Speech	Presence of Dysarthria	Average Tokens	Average MLU	Average Verbs per Utterance	Average Density	Average TTR	Average WPM	Average Noun/Verb Ratio	Average Open/Closed Class Word Ratio
Variables	*Statistics*	*I*	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*	*X*	*XI*	*XII*	*XIII*	*XIV*
	r	1	0.541	0.176	-0.159	-0.31	-0.112	0.281	0.552	0.492	0.297	-0.254	0.261	-0.095	-0.039
I	p-value		**0**	0.155	0.176	**0.013**	0.385	**0.015**	**0**	**0**	**0.013**	**0.026**	**0.024**	0.406	0.717
	N	90	90	85	90	80	79	90	90	90	90	90	90	90	90
	r		1	-0.026	-0.008	-0.165	-0.034	0.379	0.663	0.633	0.401	-0.244	0.15	-0.16	-0.11
II	p-value			0.878	0.942	0.227	0.878	**0**	**0**	**0**	**0**	**0.043**	0.227	0.227	0.395
	N		90	85	90	80	79	90	90	90	90	90	90	90	90
	r			1	-0.184	-0.048	-0.038	0.252	0.046	-0.013	-0.106	-0.278	0.012	0.037	0.038
III	p-value				0.348	0.915	0.915	0.13	0.915	0.915	0.866	0.13	0.915	0.915	0.915
	N			85	85	76	75	85	85	85	85	85	85	85	85
	r				1	0.17	0.12	0.033	-0.188	-0.173	0.039	0.006	-0.154	0.099	0.103
IV	p-value					0.321	0.507	0.894	0.321	0.321	0.894	0.954	0.321	0.507	0.507
	N				90	80	79	90	90	90	90	90	90	90	90
	r					1	0.277	-0.2	-0.415	-0.447	-0.159	0.093	-0.443	0.383	0.246
V	p-value						**0.03**	0.124	**0**	**0**	0.188	0.444	**0**	**0**	0.052
	N					80	78	80	80	80	80	80	80	80	80
	r						1	-0.206	-0.133	-0.143	-0.017	0.085	-0.056	0.105	0.001
VI	p-value							0.442	0.661	0.661	0.954	0.744	0.901	0.661	0.991
	N						79	79	79	79	79	79	79	79	79
	r							1	0.491	0.469	0.291	-0.715	0.281	-0.351	-0.31
VII	p-value								**0**	**0**	**0.009**	**0**	**0.01**	**0.003**	**0.007**
	N							90	90	90	90	90	90	90	90
	r								1	0.931	0.463	-0.34	0.449	-0.436	-0.392
VIII	p-value									**0**	**0**	**0.001**	**0**	**0**	**0**
	N								90	90	90	90	90	90	90
	r									1	0.505	-0.291	0.502	-0.54	-0.424
IX	p-value										**0**	**0.007**	**0**	**0**	**0**
	N									90	90	90	90	90	90
	r										1	-0.242	0.352	-0.444	-0.373
X	p-value											**0.039**	**0.002**	**0**	**0**
	N										90	90	90	90	90
	r											1	-0.205	0.367	0.504
XI	p-value												0.068	**0**	**0**
	N											90	90	90	90

	r												1	-0.441	-0.347
XII	p-value													**0**	**0.002**
	N												90	90	90
	r													1	0.647
XIII	p-value														**0**
	N													90	90
	r														1
XIV	p-value														
	N														90

Table 4: Analysis of the main effect of discourse type and interactions. First column highlights primary linguistic variables. Second column demonstrates the results of the repeated measured ANCOVA for the main effect of discourse type. Far right column includes any significant interactions with covariates (described in first column). ** = significant at p<0.01; * = significant a p<0.05.

MLU = mean length per utterance; WPM = words per minute; TTR = type-token ratio.

Primary linguistic variables	**Significant correlations and covariates included in ANCOVA**		**Main Effect of Discourse Type** ^ = *does not meet sphericity assumption; Greenhouse-Geiser reported*		**Significant Interactions**	
	Aphasia	**Control**	**Aphasia**	**Control**	**Aphasia**	**Control**
MLU	Age, p=.018 AQ, p<.001 AOS, p<.001 Tokens, p<.001	None	^F(1.62,121.49)=.342, p=.665	^F(2.58,195.88)=16.18, p<.001**	Age, p=.049 Tokens, p<.001	
Verbs per Utterance	Age, p<.001 AQ, p<.001 AOS, p<.001 Tokens, p<.001	None	F(3,225)=3.924, p=.009**	^F(2.40,182.33)=38.11, p<.001**	Age, p=.012	
Propositional density	Age, p=.018 AQ, p<.001 Tokens, p=.009	None	^F(2.61,224.28)=9.92, p<.001**	^F(2.55,193.96)=11.90, p<.001**	None	
TTR	Age, p=.026 AQ, p=.043 Tokens, p<.001	Tokens, p<.001	^F(2.67,229.85)=3.66, p=.017*	^F(2.50,187.18)=.002**	None	Tokens, p<.001
WPM	Age, p=.024 AOS, p<.001 Tokens, p=.01	Age, p=.032	^F(2.74,235.32)=6.77, p<.001**	^F(2.78,236.14)=2.04, p=.110	Tokens, p=.001 AQ, p<.001	Age, p=.022
Noun-verb Ratio	AOS, p<.001 Tokens, p=.003	None	^F(2.45,158.99)=5.20, p=.004**	^F(1.64,124.78)=64.27, p<.001**	AOS, p=.012	
Open-closed Class Word Ratio	Tokens, p=.007	Age, p=.027	^F(2.35,199.62)=2.27, p=.097	^F(2.38,201.90)=2.48, p=.076	None	None
Tokens	Age, p=.015 AQ, p<.001	None	^(F1.15,100.01)=.55, p=.484	^F(1.07,81.26)=115.53, p<.001	AQ, p=.01	

Table 5: Summary of the results of the repeated measures ANCOVA main effects and post hoc analyses.
Cind = Cinderella; BW = Broken Window; PBJ = Peanut Butter and Jelly
MLU = mean length of utterance; Prop. Density = propositional density; TTR = type-token ratio; WPM = words per minute

Table 5A: Post hoc tests of significant main effects. First column summarizes if there was a main effect found (for full ANCOVA results, see Table 4). The significant post hoc tests are then described in the right column; p-values have been corrected using Bonferroni correction for multiple comparisons.

Primary linguistic variables	**Main Effect of Discourse Type**		**Significant post hoc tests**	
	Aphasia	**Control**	**Aphasia**	**Control**
MLU	No	Yes		BW>PBJ, p<.001 Cat>PBJ, p<.001 Cind>PBJ, p<.001
Verbs per Utterance	Yes	Yes	BW>PBJ, p<.001 Cat>PBJ, p<.001 Cind>PBJ, p<.001	BW>PBJ, p<.001 Cat>PBJ, p<.001 Cind>PBJ, p<.001
Propositional density	Yes	Yes	Cind>BW, p<.001 Cind>Cat, p<.001 PBJ>BW, p=.03	Cind>BW, p<.001 Cind>Cat, p<.001 Cind>PBJ, p<.001
TTR	Yes	Yes	BW>Cind, p<.001 BW>Cat, p=.011 Cat>Cind, p<.001 PBJ>Cind, p<.001	BW>Cat, p<.001 BW>Cind, p<.001 BW>PBJ, p<.001 Cat>Cind, p<.001 PBJ>Cind, p<.001
WPM	Yes	No	BW>Cind, p=.003 Cat>Cind, p=.046 PBJ>Cind, p=.002	
Noun-verb Ratio	Yes	Yes	PBJ>BW, p<.001 PBJ>Cat, p<.001 PBJ>Cind, p<.001	PBJ>BW, p<.001 PBJ>Cat, p<.001 PBJ>Cind, p<.001
Open-closed Class Word Ratio	No	No		
Tokens	No	Yes		Cind>BW, p<.001 Cind>Cat, p<.001 Cind>PBJ, p<.001 Cat>BW, p<.001 Cat>PBJ, p=.004

Table 5B: A layman's summary of repeated-measures ANCOVA results and comparison between groups. This table highlights main results, interactions and a comparison of the aphasia and control groups in layman's terms.

Primary linguistic variable	**Language proxy**	**Aphasia**	**Control**	**Qualitative Comparison of Control with Aphasia**
MLU	Linguistic productivity	• No main effect of discourse type, but interaction with age and tokens ○ Regardless of discourse type: longer MLU correlates with greater age and more tokens produced	• Main effect of discourse type, no interactions • Procedural discourse (PBJ) had shortest MLUs of all types • No significant difference in MLU length between expositional (BW, Cat) or narrative (Cinderella)	• MLU was more mediated by age and total tokens produced in aphasia group but not in control group • An effect of discourse type of MLU was only observed for control group
Verbs per Utterance	Syntactic complexity	• Main effect of discourse type, interaction with age ○ Regardless of discourse type: more verbs per utterance correlated with greater age • Procedural discourse (PBJ) had fewest verbs per utterance of all types • No significant difference in verbs per utterance between expositional (BW, Cat) or narrative (Cinderella)	• Main effect of discourse type, no interactions • Procedural discourse (PBJ) had fewest verbs per utterance of all types • No significant difference in verbs per utterance between expositional (BW, Cat) and narrative (Cinderella)	• Procedural discourse had the fewest verbs per utterance of all types • No interaction with age in control group, but an interaction with age in aphasia group
Propositional density	Content richness	• Main effect of discourse type, no interactions • Narrative discourse (Cinderella) was the densest type, though not significantly more dense than procedural (PBJ) • Expositional (BW) was the least dense type	• Main effect of discourse type, no interactions • Narrative discourse (Cinderella) was the densest type • No significant differences in density between expositional (BW, Cat) and procedural (PBJ)	• Narrative discourse (Cinderella) was the densest type • For aphasia, expositional (BW) was the least dense; no significant difference for control group between expositional and procedural
TTR	Vocabulary / lexical diversity	• Main effect of discourse type, no interactions • Narrative (Cinderella) was the least diverse • Expositional (BW) was the most diverse, but not significantly different from procedural (PBJ)	• Main effect of discourse type, interactions with tokens ○ Higher TTR correlated with fewer tokens produced • Expositional (BW) had the most diverse speech (highest TTR), thus the greatest different words to total words ratio • Narrative (Cinderella) was the least diverse	• Expositional (BW) tended to have the most diverse speech with few exceptions • Narrative (Cinderella) was the least diverse

WPM	Efficiency, fluency	• Main effect of discourse type, interaction with tokens and aphasia severity ○ More WPM correlated with more tokens and milder aphasia • Narrative (Cinderella) had the lowest WPM • No significant differences in WPM between expositional (BW, Cat) and procedural (PBJ)	• No main effect of discourse type, but interaction with age ○ Regardless of discourse type: more WPM correlates with greater age	• Main effect only in aphasia group, and this main effect was mediated by aphasia severity and total tokens produced • No main effect of discourse type, but interaction with age in control group
Noun-verb Ratio	Syntactic complexity	• Main effect of discourse type, interaction with presence of AOS ○ This was a small relationship (R2=.15), and not linear; there was a tendency for greater noun-verb ratio to correlate with presence of AOS • Procedural (PBJ) had the highest noun-verb ratio, suggesting it was the least complex discourse type • No significant differences in noun-verb ratio between expositional (BW, Cat) and narrative (Cinderella)	• Main effect of discourse type, no interactions • Procedural (PBJ) had the highest noun-verb ratio, suggesting it was the least complex discourse type • No significant differences in noun-verb ratio between expositional (BW, Cat) and narrative (Cinderella)	• Procedural (PBJ) had the highest noun-verb ratio, suggesting it was the least complex discourse type • No significant differences in noun-verb ratio between expositional (BW, Cat) and narrative (Cinderella)
Open-closed Class Word Ratio	Syntactic complexity	• No main effect of discourse type, no interactions	• No main effect of discourse type, no interactions	• No main effect of discourse type, no interactions
Tokens	Word retrieval	• No main effect of discourse type, interaction with aphasia severity ○ Regardless of discourse type: milder aphasia correlates with more tokens produced	• Main effect of discourse type, no interactions • Narrative (Cinderella) produced the most tokens of all types • Expositional (Cat) produced more tokens than both expositional (BW) and procedural (PBJ) • No significant difference between expositional (BW) and procedural (PBJ)	• Main effect of discourse type only for control group, where narrative produced the most tokens • Tokens was highly mediated by aphasia severity in the aphasia group, but no main effect was found

Table 6: Primary linguistic variables summarized for control and aphasia groups. Statistically compared using, where appropriate, parametric Independent Samples T-Test or non-parametric Mann-Whitney U Test. Use of parametric versus non-parametric comparison of means was based on the significance of Levene's Test of Homogeneity of Variances based on mean (p>0.05 indicates use of parametric test; p<.05 indicates use of non-parametric test). Raw means are shown in Figure 2. Original p-values are listed, with a * indicating if this p-value was statistically significant after Benjamini-Hochberg correction (p<0.05).

BW = Broken Window, expositional discourse; Cat, expositional discourse; Cind = Cinderella, narrative discourse; PBJ = Peanut Butter and Jelly, procedural discourse.

MLU = mean length per utterance; TTR = type-token ratio; WPM = words per minute.

Primary linguistic variable	**Discourse Type**	**Aphasia Mean (SD) [N]**	**Control Mean (SD) [N]**	**Difference in Means** *^Independent samples t-test* *#Mann-Whitney U Test*
MLU	Expositional (BW)	6.45 (2.85) [90]	11.34 (3.62) [84]	^t(172)=9.92, p<.001*
	Expositional (Cat)	7.24 (3.01) [90]	11.43 (3.53) [78]	^t(166)=8.30, p<.001*
	Narrative (Cind)	6.65 (2.45) [90]	11.26 (2.79) [84]	^t(172)=11.61, p<.001*
	Procedural (PBJ)	5.77 (2.15) [90]	9.14 (2.70) [82]	^t(170)=9.08, p<.001*
Verbs per utterance	Expositional (BW)	.90 (.52) [90]	1.62 (.54) [84]	^t(172)=9.02, p<.001*
	Expositional (Cat)	1.03 (.56) [90]	1.74 (.58) [78]	^t(166)=8.12, p<.001*
	Narrative (Cind)	.93 (.48) [90]	1.61 (.39) [84]	^ t(172)=10.23, p<.001*
	Procedural (PBJ)	.69 (.41) [90]	1.13 (.26) [82]	#U=6192.00, p<.001*
TTR	Expositional (BW)	.59 (.13) [90]	.61 (.07) [84]	#U=4314.50, p=0.11
	Expositional (Cat)	.55 (.13) [90]	.53 (.07) [78]	#U=3195.00, p=.32
	Narrative (Cind)	.42 (.14) [90]	.39 (.11) [84]	^ t(172)=1.89, p=.06
	Procedural (PBJ)	.60 (.17) [90]	.51 (.11) [82]	#U=2487.00, p<.001*
WPM	Expositional (BW)	64.28 (38.33) [90]	140.75 (31.54) [83]	^t(171)=14.26, p<.001*
	Expositional (Cat)	60.30 (38.35) [90]	151.56 (32.20) [78]	^t(166)=16.56, p<.001*
	Narrative (Cind)	54.89 (30.75) [90]	130.92 (29.21) [84]	^t(172)=16.70, p<.001*
	Procedural (PBJ)	63.41 (33.73) [90]	152.64 (34.20) [82]	^t(170)=17.21, p<.001*
Propositional density	Expositional (BW)	.38 (.11) [90]	.46 (.05) [84]	#U=5934.00, p<.001*
	Expositional (Cat)	.40 (.10) [90]	.45 (.04) [78]	#U=4991.50, p<.001*
	Narrative (Cind)	.44 (.09) [90]	.49 (.03) [84]	#U=5546.50, p<.001*
	Procedural (PBJ)	.41 (.11) [90]	.46 (.05) [82]	#U=4793.00, p<.001*
Noun-verb ratio	Expositional (BW)	1.58 (1.49) [86]	1.21 (.35) [84]	#U=3188.00, p=.19
	Expositional (Cat)	1.67 (1.72) [86]	1.20 (.35) [78]	#U=3177.50, p=.56
	Narrative (Cind)	1.33 (1.03) [88]	1.12 (.26) [84]	#U=4090.00, p=.23
	Procedural (PBJ)	2.63 (2.16) [83]	2.00 (.78) [82]	#U=3238.00, p=.59
Open-closed class word ratio	Expositional (BW)	.89 (.60) [88]	.78 (.13) [84]	#U=3964.00, p=.41
	Expositional (Cat)	.79 (.53) [89]	.74 (.11) [78]	#U=4020.50, p=.08
	Narrative (Cind)	.80 (.85) [90]	.69 (.07) [84]	#U=4541.50, p=.02

	Procedural (PBJ)	.82 (.59) [89]	.74 (.16) [82]	#U=4032.00, p=.24
Tokens	Expositional (BW)	50.91 (33.18) [90]	83.86 (37.21) [84]	^t(172)=6.17, p<.001*
	Expositional (Cat)	73.62 (46.47) [90]	106.91 (43.48) [78]	^t(166)=4.77, p<.001*
	Narrative (Cind)	204.89 (160.69) [90]	458.87 (307.97) [84]	#U=5973.50, p<.001*
	Procedural (PBJ)	42.81 (35.60) [90]	85.76 (50.63) [82]	#U=6000.50, p<.001*

Figure 1

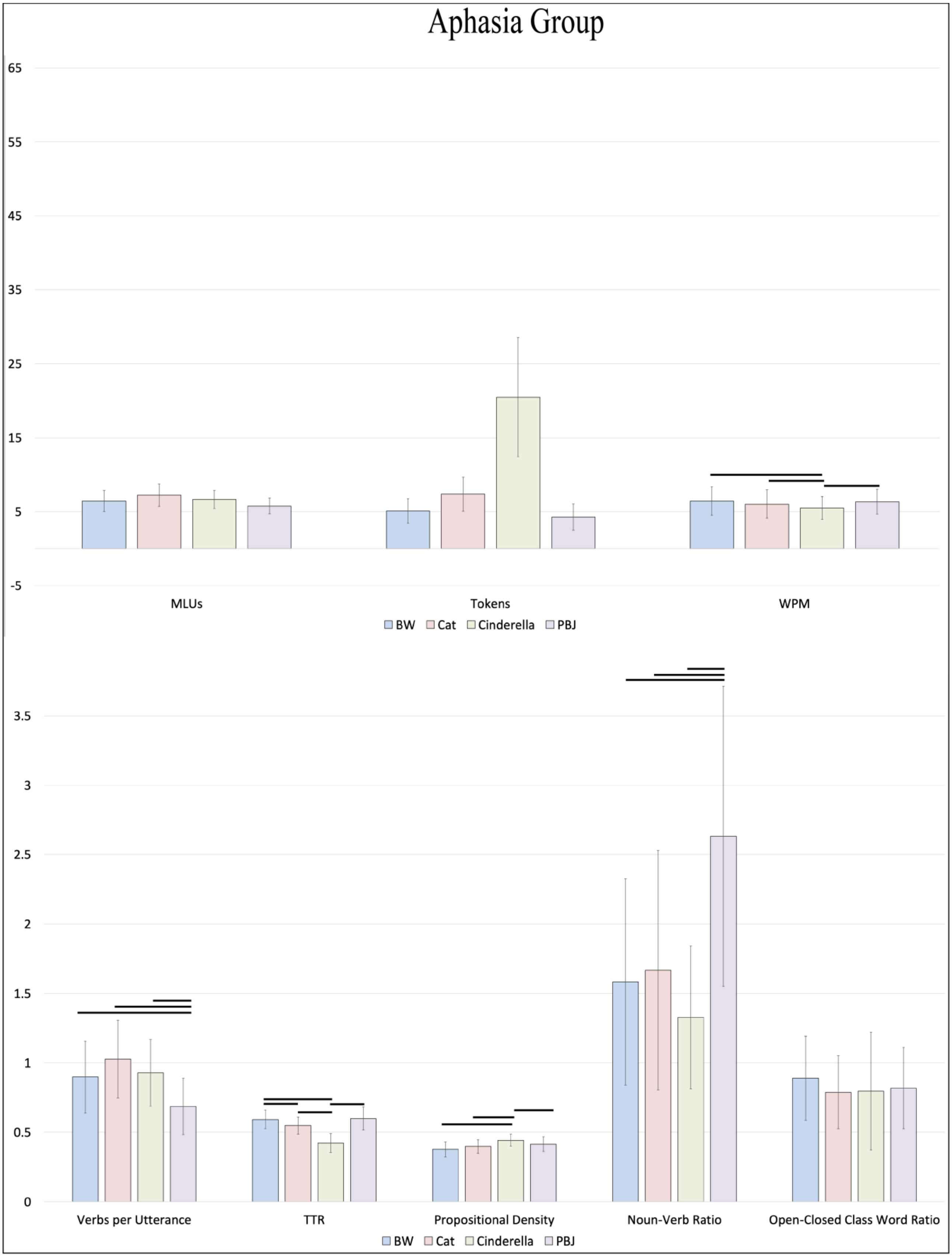

Aphasia Group
65
55
45
35
25
15
5
-5
MLUs
Tokens
WPM
BW Cat Cinderella PBJ
3.5
3
2.5
2
1.5
1
0.5
0
Verbs per Utterance
TTR
Propositional Density
Noun-Verb Ratio
Open-Closed Class Word Ratio
BW Cat Cinderella PBJ

Figure 1

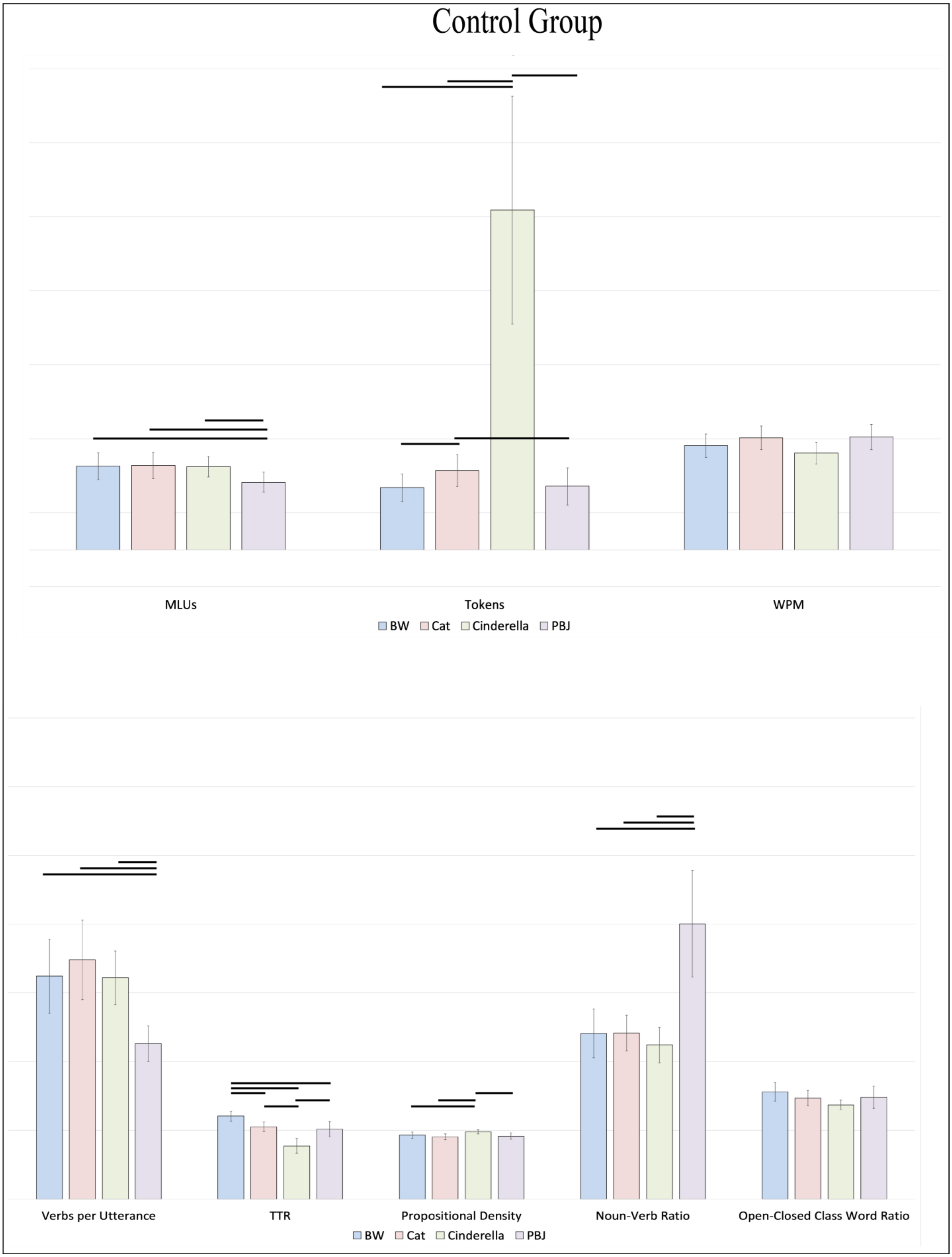

Control Group
MLUs
Tokens
WPM
BW
Cat
Cinderella
PBJ
Verbs per Utterance
TTR
Propositional Density
Noun-Verb Ratio
Open-Closed Class Word Ratio
BW
Cat
Cinderella
PBJ

Figure 2

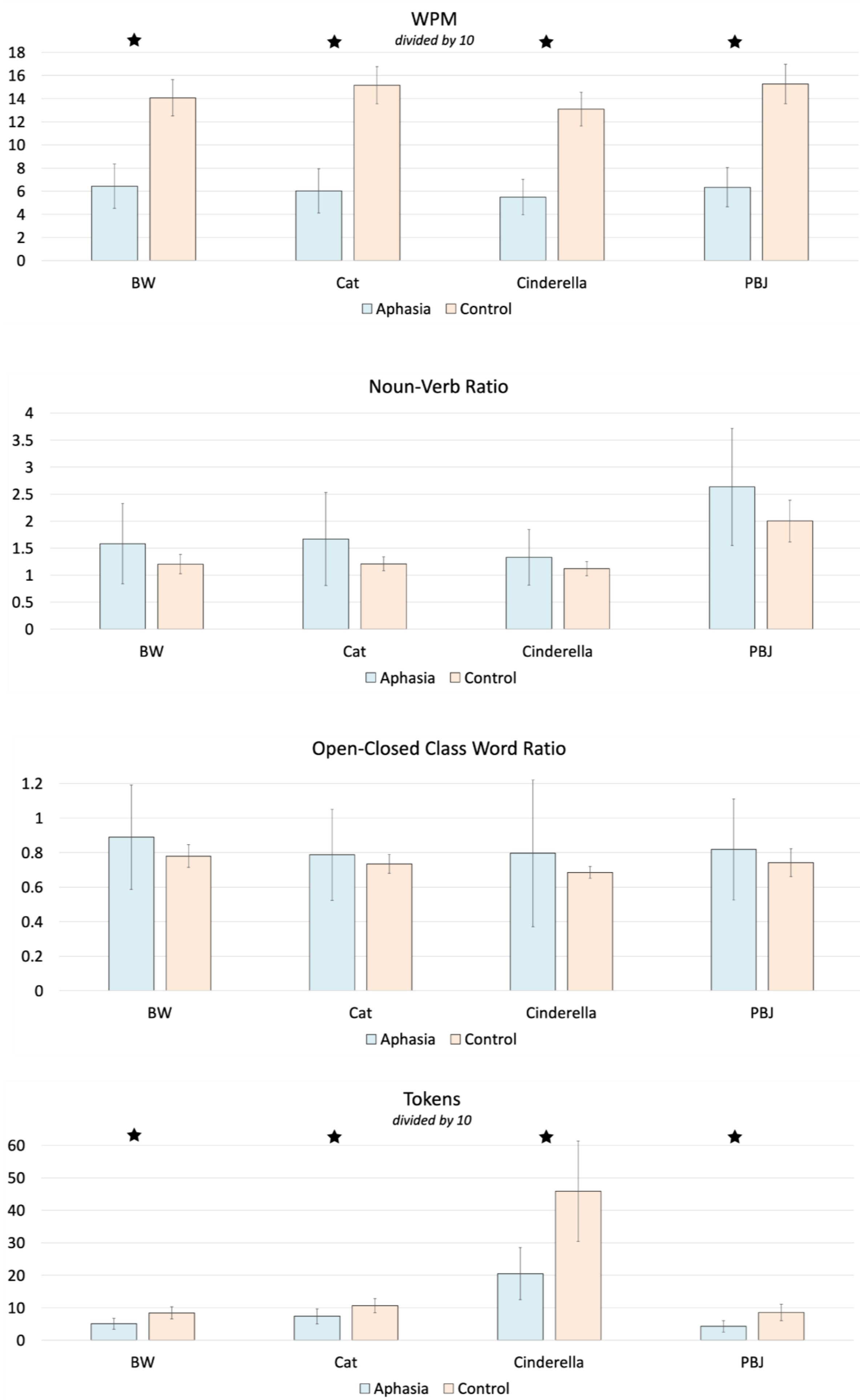
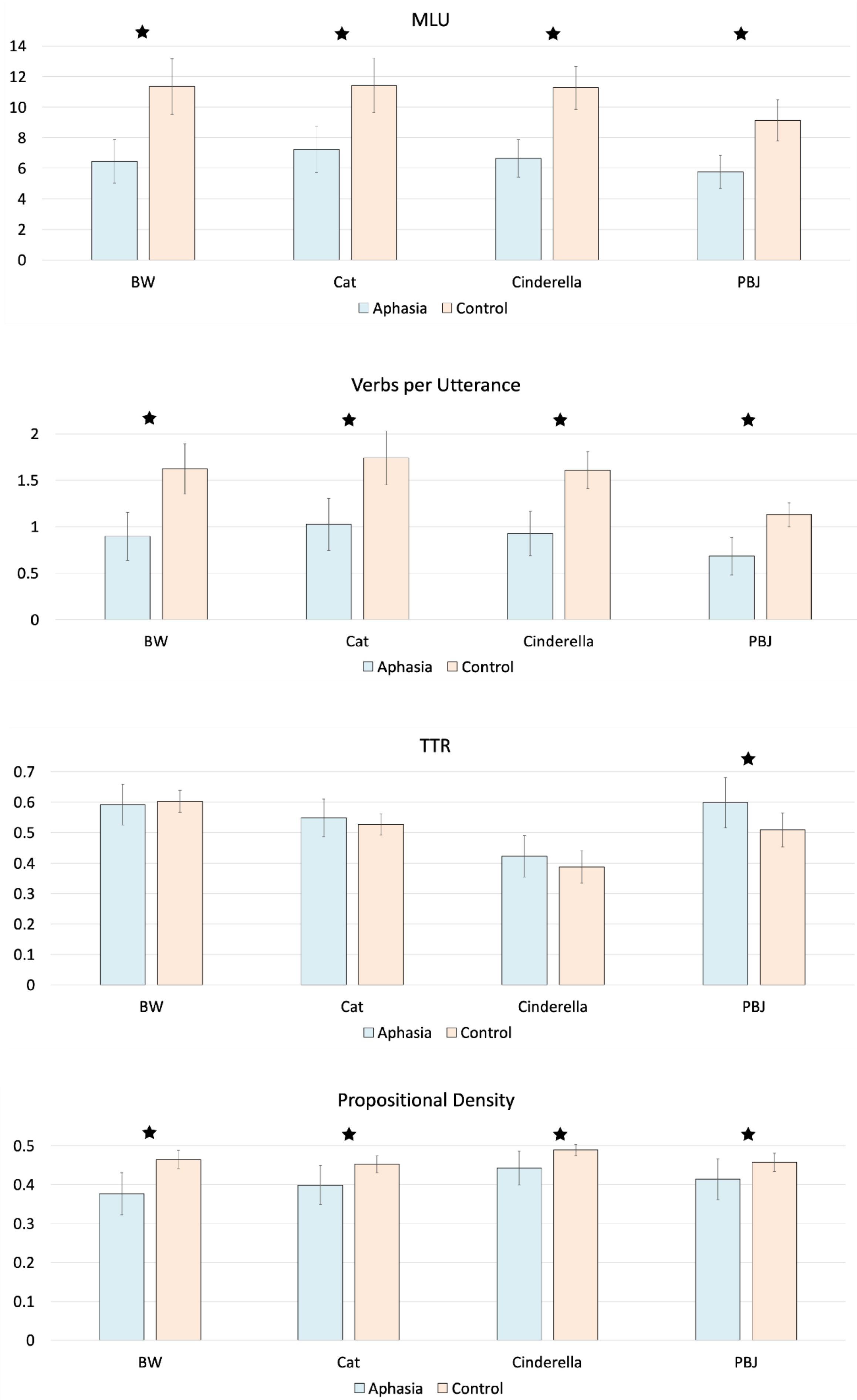

WPM
divided by 10
18
16
14
12
10
8
6
4
2
0
BW
Cat
Cinderella
PBJ
Aphasia
Control
Noun-Verb Ratio
4
3.5
3
2.5
2
1.5
1
0.5
0
BW
Cat
Cinderella
PBJ
Aphasia
Control
Open-Closed Class Word Ratio
1.2
1
0.8
0.6
0.4
0.2
0
BW
Cat
Cinderella
PBJ
Aphasia
Control
Tokens
divided by 10
60
50
40
30
20
10
0
BW
Cat
Cinderella
PBJ
Aphasia
Control

Figure 2


MLU
14
12
10
8
6
4
2
0
BW
Cat
Cinderella
PBJ
Aphasia
Control
Verbs per Utterance
2
1.5
1
0.5
0
BW
Cat
Cinderella
PBJ
Aphasia
Control
TTR
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
BW
Cat
Cinderella
PBJ
Aphasia
Control
Propositional Density
0.5
0.4
0.3
0.2
0.1
0
BW
Cat
Cinderella
PBJ
Aphasia
Control