



## Leveraging big data to understand the interaction of task and language during monologic spoken discourse in speakers with and without aphasia

Brielle C. Stark & Julia Fukuyama

To cite this article: Brielle C. Stark & Julia Fukuyama (2020): Leveraging big data to understand the interaction of task and language during monologic spoken discourse in speakers with and without aphasia, *Language, Cognition and Neuroscience*, DOI: [10.1080/23273798.2020.1862258](https://doi.org/10.1080/23273798.2020.1862258)

To link to this article: <https://doi.org/10.1080/23273798.2020.1862258>

 [View supplementary material](#) 

 Published online: 21 Dec 2020.

 [Submit your article to this journal](#) 

 Article views: 29

 [View related articles](#) 

 [View Crossmark data](#) 

# Leveraging big data to understand the interaction of task and language during monologic spoken discourse in speakers with and without aphasia

Brielle C. Stark<sup>a,b,c,\*</sup> and Julia Fukuyama<sup>d,\*</sup>

<sup>a</sup>Department of Speech and Hearing Sciences, Indiana University Bloomington, Bloomington, IN, USA; <sup>b</sup>Program in Neuroscience, Indiana University Bloomington, Bloomington, IN, USA; <sup>c</sup>Cognitive Science Program, Indiana University Bloomington, Bloomington, IN, USA; <sup>d</sup>Department of Statistics, Indiana University Bloomington, Bloomington, IN, USA

## ABSTRACT

Monologic spoken discourse allows us to evaluate every day speech while retaining some experimental constraint. It also has clinical relevance, providing cognitive-linguistic information not measured on typical standardised tests. Here, we leverage big behavioural data (AphasiaBank) to understand how discourse genres (narrative, procedural, expository), and unique tasks within those genres, influence microstructural elements of discourse. We compare task × microstructure interaction across speakers with and without aphasia and evaluate the influence of aphasia type and overall aphasia severity on this interaction. Using multivariate statistical methods, we find that, for both speaker groups, discourse microstructure is most similar for tasks within the same discourse genre and that microstructure is largely dissociable across discourse genres. The aphasia group had more speaker variance per task, which was partially explained by aphasia type and overall aphasia severity. Our results provide necessary information for usage and interpretation of monologic discourse in research and clinical contexts.

## ARTICLE HISTORY

Received 15 April 2020  
Accepted 20 November 2020

## KEYWORDS

Aphasia; “connected speech”; discourse; language; stroke

## 1. Introduction

When we produce spoken language, we do not simply name objects or produce single words (Dell, 1988; Levelt, 1989). Instead, we produce meaningfully organised output consisting of many words, constantly suppressing alternatives at multiple linguistic levels (e.g. semantic, lexical, phonological, syntactic) in discourse. Discourse breaks down in a variety of disorders, like aphasia – a language disorder which most commonly occurs after acquired brain damage to the dominant [left, in most cases] cerebral hemisphere. It isn't surprising that discourse is affected in aphasia, given that successful discourse requires an intact linguistic system (i.e. access to semantic, lexical, phonological and morphosyntactic information), which is the primary impairment in aphasia. As such, discourse elicitation in aphasia is thought to be an ecologically valid and comprehensive means to understand the language system's affected and unaffected processes (Bryant et al., 2016), lending information about complex cognitive-linguistic behaviours and cognitive-communicative deficits that are not otherwise measured by standardised assessments. It follows that discourse is commonly employed in clinical and research settings to assess language and

language-related cognitive processes (e.g. organisation) in persons with aphasia (Bryant et al., 2016; Linnik et al., 2016), and is used as a primary outcome of speech-language intervention (Brady et al., 2016).

In linguistics, the term *discourse* is typically defined as a language unit longer than a single sentence, but in clinical contexts, is more narrowly defined as language beyond a single simple clause, which is used for a specific purpose or function (Dipper & Pritchard, 2017). Discourse can be either monologic or dialogic, dependent on the number of interlocutors. Discourse analysis comprises three components (as described by Frederiksen et al., 1990; Sherratt, 2007): (1) how language is used (lexical and grammatical resources), typically termed its “microstructural” level, (2) what information is included, and (3) how the information is structured. Levels two and three typically are thought of as “macrostructural” or “functional” levels, i.e. pertaining to information beyond linguistic structure. For the purposes of a concise introduction, we will focus on the microstructural level, as this is the crux of the analysis we provide in our results section, but we will make note of evidence from macrostructural levels where possible. It should also be noted that the evidence we provide, unless noted otherwise, is from English speakers.

**CONTACT** Brielle C. Stark  bcstark@iu.edu  Department of Speech and Hearing Sciences, Indiana University Bloomington, 200 S Jordan Avenue, Bloomington, IN 47405, USA

\*These authors contributed equally to this work.

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/23273798.2020.1862258>.

© 2020 Informa UK Limited, trading as Taylor & Francis Group

Here, we will restrict discussion to monologic discourse. There are several elicitation genres of monologic discourse (Bryant et al., 2016): (1) *Expositional* discourse (or expository discourse; sometimes called “descriptive” discourse) is elicited by providing subjects with a set of structured instructions and, often, a picture or picture sequence to describe (e.g. “tell me what’s happening in this picture”); (2) *Narrative* discourse is elicited by having the subject recount a personal story (“personal narrative”) or retell a well-known story (“story retelling”). Typically, narrative discourse does not employ visual aids, and the instructions are semi-structured (e.g. “tell me what you know about...”); and (3) *Procedural* discourse is elicited by having subjects describe a set of actions or a procedure (e.g. “tell me how you might make a ...”), typically without pictorial cues. Within each elicitation genre, a variety of tasks exist. For example, within expositional discourse, the most commonly used tasks are single picture descriptions and picture sequence descriptions (Bryant et al., 2016). Within the narrative context, there are autobiographical tasks, as well as tasks that rely more on semantic memory, such as fictional story retell (“Cinderella”).

When evaluating spoken discourse in aphasia, most speech-language pathologists employ a single discourse elicitation method (Bryant et al., 2017). Indeed, the most commonly used elicitation method for adult clinical populations (acquired aphasia; primary progressive aphasia; other dementias) is the single picture description task, a type of expositional elicitation method (Bryant et al., 2016; Faroqi-Shah et al., 2020; K. C. Fraser et al., 2015). Administering discourse samples is not time consuming, but subsequent analysis can be; for that reason, doing so is uncommon clinically for a variety of reasons, the predominant being time constraint (Bryant et al., 2017; Cruice et al., 2020). This is not surprising, given that it is estimated to take a trained transcriber anywhere from 6 to 12 min to transcribe a single minute of a discourse sample from a person with aphasia (Boles, 1998) and additional time to code and analyse that sample. Therefore, time is a non-negligible restraint, which may directly impair comprehensive assessment of discourse. Despite this, there is mounting evidence that language sampling using multiple, varied discourse elicitation methods will provide the most comprehensive language assessment.

Firstly, there is evidence that language must be sampled in sufficient quantity to obtain reliable estimates. For example, test-retest reliability of correct information units in speakers with aphasia, a commonly employed metric of language efficiency and productivity, is low unless samples contain at least 300–

400 words (Brookshire & Nicholas, 1994). Notably, this amount of word production is not typically acquired from a single discourse sample in most speakers with aphasia. Indeed, an analysis of the data from Aphasia-Bank (MacWhinney et al., 2011) demonstrates that, during a single picture description task (e.g. Cat Rescue), persons with aphasia produce on average 70 tokens (lemmas) (Stark, 2019). Indeed, even speakers without aphasia (non-brain-damaged control group matched on age and education) produced, on average, only a few more tokens (~90) on this task (Stark, 2019). A cursory examination of the Pitt Corpus from Dementia-Bank (Becker et al., 1994) lends a similar result, with the average tokens produced on the Cookie Theft picture (Goodglass & Kaplan, 1972) by persons with probable Alzheimer’s Disease at their first assessment being ~81 ( $N = 196$ ) (*this analysis run for example purposes by the authors*). This is also the case for RHDBank (MacWhinney et al., 2011), a database of persons with right hemisphere stroke, when retelling the Cat Rescue story. Indeed, they produced on average ~186 tokens ( $N = 16$ ) (*this analysis run for example purposes by the authors*). Therefore, despite a single picture description being the most prevalent means of acquiring a discourse sample in clinical practice, it does not align with the evidence on sampling reliability. There is a need to acquire multiple discourse samples, to maximise speaking amount and enhance reliability. It should be noted, though, that some studies that have acquired multiple samples still do not hit the 300–400 word recommended level (Boyle, 2014; Cameron et al., 2010), emphasising that more work must be done to ascertain which discourse-extracted outcome measures are most stable and reliable.

Second, there is considerable evidence that elicitation method interacts with language produced in the discourse; that is, simply put, the task constrains the language produced. This evidence makes perfect sense in the context of several influential discourse models: the first stage of discourse production is the generation or retrieval of a frame structure, which reflects the different genres of discourse elicited (e.g. narrative, procedure) (Frederiksen et al., 1990; Sherratt, 2007). That is, each elicitation method requires a unique frame structure. It follows that each task, and each genre, will have a considerable impact on micro- and macrostructural levels.

Wright and Capilouto (2009) found that, in 24 healthy older adults, task instructions influenced linguistic information: if subjects were told to “talk about what is going on in the picture” (“describe”), subjects tended to produce fewer past-tense verbs than if they were told to “look at the picture(s) and tell a story that has a

beginning, middle and end" ("storytelling") (Wright & Capilouto, 2009). In a later study, Fergadiotis et al. (2011) explored differences in lexical diversity (a measure of the proportion of lexical items [i.e. nouns, verbs, adjectives and some adverbs]) across discourse genres (expositional, narrative and procedural) in older adults ( $N=27$ ) and persons with aphasia ( $N=25$ ), finding that lexical diversity was greatest during narrative discourse (and, in the case of aphasia, in picture sequences) (Fergadiotis & Wright, 2011), demonstrating genre-specific microstructural information. In a larger sample size ( $N=86$ ) of persons without aphasia, lexical diversity, a microlinguistic measure, was demonstrated to be influenced by discourse type and age (Fergadiotis et al., 2011). Not surprisingly, elicitation methods have also been found to influence micro- and macrostructural processes of language in speakers with aphasia (a non-exhaustive list: Brenneise-Sarshad et al., 1991; Brookshire & Nicholas, 1994; Correia et al., 1989, 1990; Doyle et al., 1998; Easterbrook et al., 1982; Glosser et al., 1988; Olness, 2006; Roberts & Wertz, 1989; Sahraoui & Nespoulous, 2012; Stark, 2019). For example, people with aphasia were found to produce more present tense verbs during a picture description (expositional discourse) and more past tense verbs during a story retelling (narrative discourse) ( $N=12$ ) (Olness, 2006). Reduced syntactic complexity during procedural discourse, compared with narrative discourse, has also been shown in aphasia (Ulatowska et al., 1981). A study in French-speaking persons with agrammatic speech and aphasia likewise revealed across-task variability in morphosyntax during spoken language (contrasting different discourse elicitation methods as well as constrained sentence production), suggesting that participants tended to adjust the morphosyntactic aspects of their speech according to task-dependent factors (Sahraoui & Nespoulous, 2012).

While some of these early studies in aphasia included small sample sizes and did not directly compare performance of speakers with aphasia to speakers without aphasia, there has been considerable methodological improvement, with studies contrasting language in speakers with and without aphasia across a variety of discourse genres (a non-exhaustive list including S. G. H. Dalton & Richardson, 2019; Richardson & Dalton, 2016; Stark, 2019), replicating early findings. Indeed, recent work by our group evaluated a handful of microstructural variables (e.g. mean length of utterance, propositional density) in a large group of speakers with aphasia ( $N=90$ ) and without aphasia ( $N=84$ ; age and education matched) across three discourse genres (narrative, expositional, procedural) (Stark, 2019). Despite the speakers with aphasia producing nearly

50% fewer words than the control group (Stark, 2019), there were genre- and task-specific microstructural patterns evidenced in both the aphasia and control groups. For example, all speakers produced the least syntactically complex language during the procedural narrative, while all speakers produced the densest language during the story retelling narrative (Stark, 2019). It is beyond the scope of this introduction to examine the interaction between elicitation method and macrostructural elements of spoken language in speakers with and without aphasia, but note that genre- and task-specific effects have been shown in this literature as well (S. G. Dalton & Richardson, 2015; S. G. H. Dalton & Richardson, 2019; Glosser & Deser, 1992; Van Leer & Turkstra, 1999). Additionally, we see task-specificity for micro- and macrostructural components of discourse in children as young as 5 years old (Berman, 2009), further emphasising that the frame generation stage molds the later stages of discourse production (i.e. insertion and integration of semantic information, selection and topicalization of information, generation of propositions and linguistic formulation) (Frederiksen et al., 1990; Sherratt, 2007) from an early age.

Altogether, the aforementioned studies demonstrate that use of multiple and varied elicitation methods is the gold standard for comprehensive analysis of discourse. However, there is a considerable research-practice gap (that is, it is uncommon to acquire multiple and varied samples clinically), due to non-negligible constraints on clinical assessment (e.g. time). For this reason, an increased focus on implementable recommendations for spoken discourse elicitation in aphasia is necessary.

When time or resources are constrained, one such option is to prioritise an elicitation method that best elicits discourse with particular properties of interest. Evidence from a variety of sources have demonstrated that understanding task-language interactions can inform selection of appropriate elicitation methods. For example, if the goal is to comprehensively understand a person's morphosyntactic system or to identify changes in morphosyntax as a result of intervention, it may be best to sample language using a narrative elicitation method rather than a procedural elicitation method (Stark, 2019; Ulatowska et al., 1981), as narratives typically associate with denser and more syntactically complex language. Pritchard and Dipper offer another hypothetical situation to the same effect: that if a client has difficulty with cohesion (a macrostructural variable), the clinician should select a discourse elicitation method that is likely to use cohesion, such as a narrative discourse with multiple characters, in order to test this hypothesis (Dipper & Pritchard, 2017). While understanding task-specific interactions with language is ideal,

doing so does not necessarily argue against collection of multiple samples (indeed, in the case of both examples, it would argue for the collection of multiple *narrative* samples). What is suggested is that there may be cases where employing a variety of elicitation methods is not fruitful, i.e. does not provide more information than a single sample might. To our knowledge, there has been limited research evaluating this conjecture. To introduce this concept we will provide some theoretical examples, below.

It is reasonable that aphasia characteristics (overall aphasia severity, type) mediate the relationship between elicitation method and language output. Unfortunately, the majority of research evaluating task-language interaction in aphasia has grouped all speakers with aphasia into a single group, despite aphasia groups typically displaying high intra-group heterogeneity in language ability. Grouping all persons with aphasia into a single group does not tease out the extent to which aphasia characteristics modulate task-language interaction.

To demonstrate this point, consider the following theoretical examples which illustrate how aphasia type and overall aphasia severity might reduce the need for use of multiple, varied elicitation methods. Some types of aphasia (e.g. Broca's aphasia) are thought to preferentially produce content words with limited use of functor words. But, it remains unclear the extent to which the ratio of content-to-functor words, for example, is influenced by elicitation method. If the primary purpose for evaluating discourse in a person with Broca's aphasia is to examine content-to-functor word ratio, and the content-to-functor word ratio is not a variable that is heavily influenced by elicitation method, this may be an opportunity to use clinical linguistic knowledge to pick the most sensitive elicitation method (e.g. to measure propositional density, choosing a narrative task rather than a procedural task, Fromm et al., 2016) and collect multiple samples using that same elicitation method, rather than employing varied methods.

Another illustrative example that is especially pertinent to the analyses we present in this project, evaluates the impact of elicitation method on language as a result of aphasia factors (type *and* overall aphasia severity). Wernicke's aphasia, a more severe fluent-type of aphasia, is characterised by production of indefinite words, paragrammatic speech, and paraphasias (word errors). A speaker with Wernicke's aphasia might produce many semantically-related, phonemically-related, neologistic, and/or unrelated paraphasias, suggesting impaired, inconsistent access to their lexical-semantic-phonological system. Characteristic of Wernicke's aphasia, there are few self-corrections and

revisions. Conduction aphasia, typically presenting as a less severe fluent-type of aphasia, is likewise characterised by production of indefinite words, paraphasias, and some paragrammatic speech, but also by *conduite d'approche*, or the successive approximations at producing target words. Nadeau (2001) and others suggest that *conduite d'approche* reflects self-correcting mechanism underpinned by intact lexical-semantic information (Nadeau, 2001). From this information, we may make the assumption that communication between the lexical-semantic-phonological system is less impaired in a speaker with Conduction aphasia than a speaker with Wernicke's aphasia. Therefore, microstructural information may not differ by discourse elicitation method in persons with Wernicke's aphasia because these speakers are, in general, unable to reliably access linguistic information, meaning that they are liable to produce unreliable linguistic information regardless of the elicitation method. That is, these individuals may have a deficit at the top of the discourse implementation hierarchy (frame selection, thought to be heavily reliant on semantic access and memory), which we discussed earlier. An inability to frame resultant language (and macrostructural) information may result in discourse appearing very similar (at micro- and macrostructural levels) across all elicitation methods. In a similar vein, speakers with severe aphasia, regardless of aphasia type, may have a limited linguistic repertoire available to them, and as such, elicitation method may not influence discourse microstructure. Despite an ability to correctly choose the discourse frame, their downstream planning (i.e. selection of linguistic information) is inherently limited. Therefore, despite probable accurate frame selection, they may produce the same, limited linguistic building blocks for all elicitation methods.

New evidence using main concept analysis (a measure evaluating the extent to which speakers produce context-relevant lexical items during tasks) supports the supposition that aphasia type and overall aphasia severity may interact with task (S. G. H. Dalton & Richardson, 2019). During most tasks (a single picture description, a picture sequence description, a procedural narrative, and a fictional narrative), persons with Wernicke's aphasia ( $N=19$ ) and Broca's aphasia ( $N=61$ ) tended to produce the same number of main concept units per task. On the other hand, persons with Anomic aphasia ( $N=86$ ) produced varying numbers of main concept units dependent on the task (ranging from 8 to 25 units), and a similar pattern was noted for the non-brain-damaged control group ( $N=145$ ) (ranging from 17 to 55 units) (S. G. H. Dalton & Richardson, 2019). What we cannot glean from these

quantitative results is whether the lexical content was similar across tasks (i.e. were they accessing similar numbers of verbs and nouns?), but we can surmise that a similar amount of macrostructural information was being conveyed by persons with Broca's and Wernicke's aphasia per task, whereas a task-dependent amount of macrostructural information was being conveyed by persons with Anomic aphasia and persons without brain damage. Research at the microstructural level, evaluating propositional density, has also shown distinct interactions of discourse task and aphasia type (Fromm et al., 2016).

Such literature raises the possibility that, in the case of certain aphasia types or severities, the use of multiple, varied discourse elicitation methods may not be required for comprehensive assessment of language. The primary objective of this project is to understand the extent to which *elicitation method* interacts with discourse (at the microstructural level) in speakers with and without aphasia. Importantly, we will examine the interaction of *elicitation method* within *speaker group*, contrasting speakers with varying aphasia types and overall aphasia severity.

Understanding the extent to which type and overall aphasia severity influence task-dependent language would inform clinical practice, bridging the aforementioned research-practice gap in spoken discourse acquisition and analysis. Further, this work has translational importance for understanding the purpose and utility of language sampling in clinical contexts, not only limited to neurogenic communication disorders (e.g. aphasia) but also other developmental, progressive and psychiatric disorders that influence language (e.g. autism, depression, dementia).

Additionally, this research has significant theoretical ramifications. Notably, each discourse task is thought to draw upon unique cognitive components; for example, fictional story retelling likely draws more on semantic memory whereas narrative story retelling (autobiographical) draws more on episodic memory (as emphasised in development by Bliss & McCabe, 2006). A distinction in microstructure produced by each task may further emphasise the unique cognitive contributions required for the task. Another important contribution to theoretical understanding of discourse is the extent to which discourse microstructure is consistent within similar groups of people (e.g. those without brain damage). If there is intra-group homogeneity in microstructure for each task, this suggests that the task being employed taps into shared cognitive and language resources across people, thus supporting the validity for acquiring monologic discourse to examine more naturalistic language processes.

Therefore, this study has both clinical and theoretical implications.

## 2. Materials and methods

### 2.1. Database

This project draws from a large database of behavioural data, AphasiaBank (MacWhinney et al., 2011) (aphasia-talkbank.org). This database comprises more than 300 speakers with a brain injury (overwhelmingly, chronic stroke) and more than 200 non-brain-damaged (typical) speakers. Detailed demographic data is provided for all speakers. For speakers with brain injury, data for a neuropsychological battery of tests is provided in addition to data from a detailed monologic discourse elicitation protocol (described in Section 2.3).

The AphasiaBank database includes some individuals who were seen multiple times. For this study, we included each speaker only once, analysing each speaker's first administration of the discourse protocol.

### 2.2. Subjects

#### 2.2.1. Classifying by overall aphasia severity and type

We acquired all participants from AphasiaBank.

For the aphasia group, the Western Aphasia Battery - Revised (WAB) (Kertesz, 2007) was used to qualify presence, overall aphasia severity, and type of aphasia. To create an overall aphasia severity metric (the Aphasia Quotient [AQ] of the WAB), four sub-component, composite scores are acquired: auditory verbal comprehension, naming, spontaneous speech and repetition. Auditory verbal comprehension composite score contains scores from yes/no questions, auditory word recognition, and sequential commands subtests. Naming composite score contains scores from object naming, word fluency, sentence completion and responsive speech subtests. Spontaneous speech composite score contains scores about the informational content and fluency of a picture description subtest and a short conversational section. Finally, repetition composite score contains scores from a single subtest comprising single-word and sentence-level stimuli. These four sections allow us to derive the AQ, which is a score from 0 to 100. According to the WAB manual, an AQ < 93.8 indicates presence of aphasia. The WAB also provides guidelines for assigning overall aphasia severity based on the AQ score. An AQ score of <50 is considered severe; a score of 50–74 is considered moderate; and a score of >75 and <93.8 is considered mild. It should be noted that the overall aphasia severity score is not necessarily reflective of discourse

ability – it is a composite, standardised score meant to reflect overall aphasia severity having taken into account repetition, auditory comprehension, naming, and spontaneous speech information.

Additionally, composite scores of auditory verbal comprehension, naming, and repetition, and the fluency subtest score from the spontaneous speech composite score, are used to identify each person's aphasia type. Aphasia types identifiable using the WAB are: Global, Broca's, Conduction, Wernicke's, Transcortical Sensory, Transcortical Motor, Isolation (Mixed Transcortical) and Anomic. Additionally, people having an AQ of >93.8 are considered to be "not aphasic by WAB" (in some cases, called "latent aphasia") but, given the presence of a chronic brain injury, often present with residual language problems (i.e. most do not score a perfect 100 on the assessment). Indeed, all persons in the AphasiaBank database, regardless of WAB score, were identified as having aphasia by a clinician and/or seeking aphasia services in a university clinic or aphasia centre. For the purposes of this study, we did not include data from people with Transcortical motor aphasia, Transcortical sensory aphasia, Mixed transcortical aphasia or Global aphasia because there were too few cases in the database. In the case of mixed transcortical aphasia, also called isolation aphasia, there were no cases in the database. We did include individuals who were not considered aphasic by WAB standards (AQ > 93.8) but also who did not score a perfect 100 on the exam; we consider this our "Not Aphasic by WAB [NABW]" group. This group was included because (1) prior work has shown mild language impairments in these individuals (Fromm et al., 2017) and (2) this group serves as a within-group control because they have had a left hemisphere brain injury and clinically present with aphasia but they test in the normal range on a formal, standardised test of aphasia.

Below is a brief description of language production characteristics associated with aphasia types included in the current study (Davis; Goodglass & Kaplan, 1972; Goodglass et al., 2000):

- *Broca's aphasia*: Broca's aphasia is typified by halting and effortful speech. Speech tends to retain content words, may contain paraphasias, and may contain impaired grammar (typically, omission). Typically, this type presents with poor repetition of words and phrases;
- *Conduction aphasia*: Conduction aphasia is characterised by relatively fluent speech (typically, can produce connected speech) with intact sentence structure. Speech may lack meaning due to word finding difficulties and also include paraphasias. Typically,

this type presents with acute difficulty in repeating phrases;

- *Wernicke's aphasia*: Wernicke's aphasia is characterised by speech that is fluent but may be paragrammatic, empty, and/or full of paraphasias (sometimes, jargon). Repetition of words and phrases is impaired;
- *Anomic aphasia*: Anomic aphasia is characterised by relatively fluent speech, though fluency may be disrupted by word finding difficulty, culminating in speech containing generic fillers (e.g. "thing") and circumlocution. Repetition of words/phrases is generally intact.
- *Not aphasic by WAB [NABW]*: Latent aphasia, or "Not Aphasic by WAB" here, is characterised by fluent speech, with some word finding difficulty and characteristics of Anomic aphasia (e.g. generic speech, circumlocution, slower speech rate). Typically, repetition is likewise intact. This type of aphasia tests in the normal range on formal, standardised batteries, but persons with this type of aphasia may self-identify as having aphasia, or be identified as having aphasia by clinicians.

Aphasia types do not necessarily present with a single overall aphasia severity, e.g. individuals with Wernicke's and Broca's aphasia typically score as severe (an AQ of <50) or moderate (AQ of 50–74), whilst individuals with Conduction and Anomic aphasias typically score as moderate or mild (AQ of >75) (Kertesz, 2007). That is, there are a range of overall aphasia severity scores (here, represented as AQ) represented within each aphasia type.

### 2.2.2. All speaker characteristics

Because we were interested in modelling microstructure (specifically, lexical variables) and needed sufficient data to do so, we excluded all speakers (including controls) who produced fewer than 20 tokens (lemmas) per task. The monologic discourse elicitation protocol from AphasiaBank was not completed in its entirety for every participant, and therefore there was a different number of speakers (control and aphasia) for each task. Therefore, it is best to refer to Table 1, which describes, in detail, demographic information for included participants. Table 1 shows the number of speakers that had data for each discourse task, separated by group and aphasia type; years of education; relevant standardised testing scores; age at testing; and sex. This table also provides an overview for each discourse task by showing total number of tokens, mean length of utterance (in words) and total utterances.

For plots demonstrating demographic and some linguistic variables by task and aphasia type, see Figure S1.

**Table 1.** Demographics and basic task-related language information for controls (top) and aphasia group (bottom), broken down by task.

Data	Control participants					Window
	Cat	Cinderella	Important event	Sandwich	Stroke	
Total <i>n</i>	224	237	74	215	238	238
Education (years)	15.2 (2.3)	15.2 (2.3)	15.3 (2.8)	15.2 (2.3)	15.3 (2.3)	15.3 (2.3)
MMSE (adjusted)	29 (1.3)	29.1 (1.2)	28.9 (1.3)	29.1 (1.3)	29.1 (1.2)	29.1 (1.2)
Age at test (years)	56.1 (23.1)	55.5 (22.9)	76.1 (7)	55.5 (23.3)	55.6 (22.8)	55.6 (22.8)
Females	0.58	0.59	0.42	0.57	0.58	0.58
Mean length of utterance (in words)	10.4 (3.2)	9.8 (2.1)	9.5 (2.3)	8.7 (2.3)	10.2 (4.9)	10.2 (4.9)
Tokens	109 (46.7)	447.2 (265.8)	338 (327.7)	95 (57.1)	85 (46.2)	85 (46.2)
Utterances	10.9 (5.6)	46.2 (31.8)	37.9 (42.7)	10.8 (5.9)	8.9 (6.6)	8.9 (6.6)

Data	Aphasia participants					Window
	Cat	Cinderella	Important event	Sandwich	Stroke	
Total <i>n</i>	247	243	259	162	273	238
Education (years)	15.7 (2.8)	15.6 (2.7)	15.6 (2.8)	15.9 (2.7)	15.5 (2.8)	15.6 (2.7)
Age at test (years)	62.2 (12.3)	61.5 (12.4)	61.7 (12.5)	61.9 (12.4)	61.9 (12.5)	62.1 (12.3)
Females	0.43	0.46	0.42	0.50	0.42	0.43
Aphasia quotient (AQ) [max=100; aphasia cut-off=93.8]	75 (17)	75 (16.4)	73.5 (17.9)	80.1 (14.6)	72 (18.8)	75.4 (17)
Aphasia duration (years)	5.2 (4.9)	5.3 (4.9)	5.2 (4.8)	5 (4.4)	5.2 (4.8)	4.9 (4.6)
Anomic aphasia	0.40	0.40	0.37	0.45	0.36	0.41
Broca's aphasia	0.17	0.20	0.22	0.10	0.26	0.17
Wernicke's aphasia	0.10	0.09	0.09	0.07	0.09	0.10
Conduction aphasia	0.21	0.20	0.20	0.20	0.20	0.21
Not aphasic by WAB	0.12	0.12	0.11	0.18	0.10	0.12
Has apraxia of speech	0.30	0.32	0.33	0.25	0.34	0.31
Has dysarthria	0.08	0.08	0.08	0.07	0.08	0.08
Cause of aphasia: not stroke	0.03	0.04	0.03	0.02	0.04	0.03
Cause of aphasia: stroke	0.92	0.92	0.92	0.94	0.92	0.92
Mean length of utterance (in words)	7 (2.9)	6.4 (2.4)	5.1 (2.1)	6.5 (2.1)	5.2 (2.3)	6.6 (2.7)
Tokens	79.9 (45)	200.7 (148.3)	149.9 (147.4)	53.5 (31.7)	299.8 (340.4)	57.3 (34.7)
Utterances	11.8 (6.3)	31.2 (20.3)	27.3 (21.3)	8.5 (4.6)	50.8 (47.4)	9.2 (5)

Notes: Wherever possible, mean and standard deviation (parentheses) are listed; others are proportions. AQ, Western Aphasia Battery Revised, Aphasia Quotient; MMSE, Mini Mental State Exam.

### 2.3. Monologic discourse elicitation protocol

Six discourse tasks were analysed in the aphasia group and five in the control group. Both groups completed the same five tasks: Broken Window, Cat Rescue, Cinderella, Important Event and Sandwich. Only speakers in the aphasia group had speaking data for the Stroke story. The following sections detail the instructions for each task. If no spontaneous response was elicited to the instructions below, troubleshooting questions were given. See AphasiaBank protocol for more detail ([aphasia.talkbank.org](http://aphasia.talkbank.org)).

#### 2.3.1. Expository genre

##### Broken Window ("Window") - Picture sequence description

Examiner: "Now I'm going to show you these pictures." *Present picture series.* "Take a little time to look at these pictures. They tell a story. Take a look at all of them, and then I'll ask you to tell me the story with a beginning, a middle, and an end. You can look at the pictures as you tell the story." *If no response in 10 s, give second prompt:* "Take a look at this picture (point to first picture) and tell me what you think is happening." *If needed, point to each picture sequentially, giving the prompt:* "And what happens here?" *For each panel, if*

*no response, provide the prompt:* "Can you tell me anything about this picture?"

##### Cat Rescue ("Cat") - Single picture description

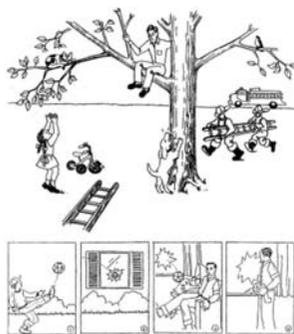
*Present picture.* "Here is another picture. Look at everything that's happening and then tell me a story about what you see. Tell me the story with a beginning, a middle, and an end." *If no response in 10 s, give second prompt:* "Take a look (point to picture) and tell me any part of the story." *If fewer than 2 utterances, give third prompt:* "Anything else you can tell me about the story?"

See Figure 1 for Window and Cat picture stimuli.

#### 2.3.2. Narrative genre

##### Cinderella - Fictional story retell

*Present picture book.* "I'm going to ask you to tell a story. Have you ever heard the story of Cinderella? Do you remember much about it? These pictures might remind you of how it goes. Take a look at the pictures and then I'll put the book away, and ask you to tell me the story in your own words." *Allow participant to look through book (assist with page turning if needed).* *Remove book. If necessary, prompt:* "Now tell me as much of the story of Cinderella as you can. You can use any details you know about the story, as well as



**Figure 1.** Visual information for Cat (bottom) and for Window (top) tasks. Stimuli openly available at [aphasia.talkbank.org](http://aphasia.talkbank.org).

the pictures you just looked at." *If participant gives a response of fewer than three utterances, or seems to falter, allow 10 s, then prompt: "What happened next?" or "Go on." Continue until participant concludes story or it is clear s/he has finished.*

### Important Event - Personal narrative

Examiner: "Thinking back, can you tell me a story about something important that happened to you in your life? It could be happy or sad or from any time – from when you were a kid or more recently." *If no response in approximately 10 s, prompt: "For instance, you could tell me about a trip you took or something about your family or your work – anything."*

### Stroke Story - Personal narrative

The Stroke story was collected only for members of the aphasia group.

"Do you remember when you had your stroke?" *If yes, "Please tell me about it." If no, "Well, how about your first memories after the stroke. What can you tell me about that?" If no response in approximately 10 s, prompt: "Try to tell me about the day you had your stroke."*

### 2.3.3. Procedural genre

#### Sandwich - Procedural narrative

"Let's move on to something a little different. Tell me how you would make a peanut butter and jelly sandwich." *If no response in 10 s, give second prompt: "If you were feeling hungry for a peanut butter and jelly sandwich, how would you make it?"* Note that a picture of a peanut butter and jelly sandwich was presented only if participant could not give a verbal response.

In cases of particularly non-fluent aphasia, a picture of a peanut-butter and jelly sandwich was provided to the participants. 69 speakers with aphasia received a picture (out of 162 speakers) (43%).

### 2.4. Microstructural (linguistic) information derived from discourse

Language samples were transcribed in CHAT format at the collection site or by AphasiaBank personnel. In all cases, transcripts were checked by a second experienced AphasiaBank transcriber (MacWhinney et al., 2011).

Prior research has predominantly evaluated the influence of elicitation method on composite measures of language (e.g. noun-verb ratio, mean length of utterance, lexical diversity), analysed independently of one another. Variable independence at the linguistic level is not the rule (e.g. number of nouns produced tends to highly relate to number of verbs produced; number of verbs produced is highly related to number of inflections produced). Here, we wanted to model many (inter-related) linguistic variables using multivariate methodology (described in Section 2.5). To do so, we derived the following variables from each speaker's discourse samples: word class (i.e. open-class, closed-class), part of speech (i.e. noun, verb) and morphosyntactic form (i.e. past tense, present participle). We first used CLAN's *gem* function to separate each speaker's transcript by task (e.g. Cat, Cinderella) (MacWhinney, 2000). Then, morphological and grammatical dependency information was automatically tagged using the *mor* command (overall, the accuracy of *mor* tagging for AphasiaBank transcripts is above 98% MacWhinney et al., 2010). The *mor* command creates a tier with a one-to-one correspondence between speaker words and morphological and syntactic tags. Then, the *eval* command was used to extract selected linguistic variables (specifically, we only extracted information from the persons with aphasia and not the experimenter/ other interlocutor). An example of a coded utterance from our sample, of describing the Window picture:

```
*PAR: &-um there was a young boy.
%mor: pro:exist|there cop|be&PAST&13S
det:art|a adj|young n|boy.
```

In this example, the participant's utterance is in the first line (\*PAR). Then, the automatic morphological (%mor) assignment is shown. Here, we can see that "young" is assigned adjective, "boy" a noun, and "was" is defined as a copula of be in the past tense (with the same form for first and third person).

Words that are repeated or revised on the main speaker tier are excluded from the %mor tier (as is typical for the CLAN program's default analysis). In instances where a paraphasia had a known target (e.g. "bik" for "book"), data for the target was used (e.g. "noun" for "book"). We did not exclude morphological errors (e.g. saying "birds" instead of "bird", the incorrect

form “birds” was included). In general, morphological errors (e.g. agreement errors) were rare in the aphasia group: there were on average <0.57 errors made per task (range: 0.05 [Window]–0.57 [Cinderella]). We did not exclude agrammatic or paragrammatic utterances.

In sum, we extracted the following variables, each of which was calculated as a proportion of total words produced:

- nouns;
- plurals;
- verbs (includes those tagged as a participle, copula and modal);
- auxiliaries;
- third person singular (e.g. verb suffix “s” and “es”);
- identical forms for first and third person (e.g. I was, he was);
- past tense (e.g. verb suffix “ed, d”);
- past participle (e.g. verb suffix “en”);
- present participle (e.g. verb suffix “ing”);
- prepositions;
- adverbs;
- adjectives;
- conjunctions;
- determiners (includes articles, demonstratives, interrogatives, numbers and possessives used as determiners);
- pronouns;
- open class words;
- closed class words.

The distribution of these variables by group (aphasia, control) and task is shown in [Table 2](#).

## 2.5. Statistical analysis

### 2.5.1. A $p$ -dimensional space for linguistic variables

Our main goal was to describe and contrast language produced in response to different elicitation methods across speaker groups (control, aphasia) and by aphasia type and/or aphasia AQ subscores, using all of the linguistic variables simultaneously instead of looking for differences one variable at a time. We have two kinds of analysis that need to be described separately: first, describing language structure in terms of a set of categorical variables (e.g. task, aphasia type), and second, describing language in terms of continuous variables (e.g. aphasia AQ subscore). In both cases we can use multivariate regression to model the linguistic variables, but the methods to describe and visualise the resulting model differ.

### 2.5.2. Between-class analysis to describe the relationship between task and language microstructure

To analyse the relationship between categorical variables and language microstructure, we compute the mean of the linguistic variables corresponding to each category. Recall that each mean is a vector in  $p$ -dimensional space, and note that these means are equal to the fitted values in a multivariate regression model.

To visualise the relationships between the categories and the category means, we turn to between-class analysis (BCA). In BCA, the category means are projected down from the initial  $p$ -dimensional space to a lower-dimensional space in such a way that the distances between the category means in the lower-dimensional spaces reflect as well as possible the original distances between the categories in the high-dimensional space. We use a two-dimensional space as the lower-dimensional space so that we can plot the representation of the category means. Since the distances between category means in the two-dimensional BCA plot are a good approximation of the true distances between the category means, categories that have similar language composition are plotted near to each other, while categories that have dissimilar language composition are positioned farther away.

In addition to the representation of the category means in a lower-dimensional space, BCA also provides information about the relationships between the variables (e.g. language data) and the axes in the BCA plot. If a category has mean  $x = (x_1, \dots, x_p)$ , then its position on the  $j$ th BCA axis is given by  $\sum_{i=1}^p x_i v_{ji}$ . This gives us insight about the reasons for differences in the position of the category means on the BCA axes. For instance, if two category means have similar values for all but variable  $i$  and  $v_{1i} > 0$ , the category with the larger value of variable  $i$  will have a larger score on the first BCA axis. For variable  $i$ , we can make a point whose position on the first BCA axis is  $v_{1i}$  and whose position on the second BCA axis is  $v_{2i}$ . This representation of the variables gives us some insight into the variables that are likely responsible for differences between the categories.

In addition to BCA for visualisation, we can summarise the relationships between categories by computing the average distance between category means. In the  $p$ -dimensional space, we can compute means within categories (for instance, task type, aphasia type, combinations of the two) and we can then use the relative distances between the category means as a measure of the relative similarities of language produced by different groups and/or in response to different tasks.

**Table 2.** Summary of linguistic variables in mean(SD) by group (aphasia, control) and task (Important Event, Stroke, Cinderella, Cat, Window, Sandwich), as expressed by proportion of total words produced.

Task	Controls					
	Important event	Cinderella	Cat	Window	Sandwich	
Plurals	0.03 (0.02)	0.03 (0.01)	0.01 (0.01)	0.01 (0.01)	0.02 (0.02)	
Nouns	0.17 (0.04)	0.19 (0.03)	0.21 (0.04)	0.21 (0.04)	0.26 (0.05)	
Verbs	0.18 (0.04)	0.19 (0.03)	0.19 (0.03)	0.19 (0.03)	0.16 (0.04)	
Auxiliary	0.02 (0.01)	0.02 (0.01)	0.04 (0.02)	0.03 (0.02)	0.00 (0.01)	
Third-person singular	0.01 (0.01)	0.04 (0.04)	0.05 (0.04)	0.08 (0.06)	0.01 (0.01)	
1st/3rd person identical forms	0.03 (0.02)	0.02 (0.01)	0.01 (0.01)	0.01 (0.02)	0.00 (0.00)	
Past tense	0.10 (0.03)	0.07 (0.05)	0.06 (0.04)	0.05 (0.05)	0.00 (0.01)	
Past participle	0.02 (0.01)	0.01 (0.01)	0.02 (0.02)	0.01 (0.01)	0.00 (0.01)	
Present participle	0.02 (0.01)	0.02 (0.01)	0.03 (0.02)	0.04 (0.02)	0.00 (0.01)	
Prepositions	0.09 (0.02)	0.07 (0.02)	0.07 (0.02)	0.09 (0.03)	0.11 (0.03)	
Adjectives	0.04 (0.02)	0.03 (0.01)	0.02 (0.01)	0.03 (0.02)	0.02 (0.02)	
Adverbs	0.07 (0.03)	0.07 (0.02)	0.08 (0.03)	0.07 (0.03)	0.07 (0.03)	
Conjunctions	0.02 (0.01)	0.02 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	
Determiners	0.06 (0.02)	0.10 (0.02)	0.15 (0.03)	0.12 (0.04)	0.11 (0.04)	
Pronouns	0.16 (0.04)	0.14 (0.03)	0.08 (0.03)	0.11 (0.04)	0.11 (0.05)	
Open-class words	0.38 (0.07)	0.40 (0.04)	0.41 (0.05)	0.42 (0.06)	0.42 (0.06)	
Closed-class words	0.57 (0.10)	0.58 (0.06)	0.57 (0.06)	0.56 (0.07)	0.57 (0.07)	

Task	Aphasia					
	Important event	Stroke	Cinderella	Cat	Window	Sandwich
Plurals	0.03 (0.03)	0.02 (0.02)	0.03 (0.02)	0.02 (0.02)	0.00 (0.01)	0.02 (0.02)
Nouns	0.15 (0.07)	0.13 (0.07)	0.18 (0.07)	0.20 (0.08)	0.20 (0.08)	0.25 (0.09)
Verbs	0.18 (0.06)	0.19 (0.06)	0.18 (0.05)	0.18 (0.05)	0.18 (0.05)	0.15 (0.05)
Auxiliary	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.04 (0.03)	0.04 (0.03)	0.01 (0.02)
Third-person singular	0.03 (0.03)	0.02 (0.02)	0.04 (0.04)	0.06 (0.05)	0.06 (0.05)	0.02 (0.03)
1st/3rd person identical forms	0.02 (0.03)	0.02 (0.02)	0.02 (0.02)	0.01 (0.03)	0.01 (0.03)	0.00 (0.01)
Past tense	0.06 (0.04)	0.06 (0.04)	0.07 (0.05)	0.04 (0.04)	0.06 (0.05)	0.01 (0.02)
Past participle	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.02)	0.00 (0.01)
Present participle	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.04 (0.03)	0.04 (0.04)	0.00 (0.01)
Prepositions	0.05 (0.03)	0.05 (0.03)	0.05 (0.03)	0.05 (0.03)	0.06 (0.04)	0.07 (0.04)
Adjectives	0.04 (0.03)	0.03 (0.03)	0.04 (0.03)	0.02 (0.02)	0.02 (0.03)	0.01 (0.02)
Adverbs	0.06 (0.04)	0.07 (0.04)	0.07 (0.04)	0.08 (0.05)	0.06 (0.05)	0.08 (0.05)
Conjunctions	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.01 (0.02)	0.01 (0.01)
Determiners	0.04 (0.03)	0.04 (0.03)	0.09 (0.04)	0.14 (0.07)	0.14 (0.07)	0.09 (0.05)
Pronouns	0.17 (0.06)	0.18 (0.06)	0.15 (0.07)	0.11 (0.07)	0.13 (0.07)	0.14 (0.06)
Open-class words	0.35 (0.07)	0.33 (0.07)	0.38 (0.08)	0.38 (0.08)	0.40 (0.08)	0.40 (0.08)
Closed-class words	0.53 (0.12)	0.56 (0.10)	0.58 (0.09)	0.58 (0.09)	0.56 (0.09)	0.59 (0.08)

These distances can then be used in a hypothesis-testing framework (e.g. PERMANOVA) to assess significance of the overall distances between the categories, or we can create bootstrap confidence intervals for the between-category distances.

### 2.5.3. Regression and PCA-IV to describe the relationship between aphasia AQ and language microstructure

While BCA is appropriate for modelling and visualising the relationship between language microstructure and a categorical variable, it is not applicable to describing the relationship between language microstructure and a continuous variable such as aphasia AQ. To do so, we used a combination of multivariate regression of the linguistic variables on aphasia AQ subscore and the other relevant variables and used PCA with respect to instrumental variables (PCA-IV) to visualise the resulting regression model.

The problem that PCA-IV aims to solve is as follows: In a multivariate regression model, we have  $\hat{X} = ZB$ , where

$\hat{X}$  are the fitted values,  $Z$  are the predictor variables, and  $B$  are the fitted coefficients. In a simple linear regression model with a one-dimensional response and a one-dimensional predictor, we can plot the response on the vertical axis and the predictor on the horizontal axis. When there are  $p > 1$  response variables and  $q > 1$  predictor variables, we are no longer able to make such a plot. What we can do, however, is project the fitted values onto a two-dimensional subspace that explains the maximal amount of variance in the fitted values. Having done so, we can describe the relationship between

- (1) That space and the  $p$  response variables, and
- (2) That space and the  $q$  predictor variables

If  $P$  is the matrix giving the projection of  $\hat{X}$  onto the space that explains the most variance in  $\hat{X}$ , we plot the rows of  $\hat{X}P$  (for the samples), the rows of  $P$  (for the response variables), and the rows of  $BP$  (for the predictor variables).

The plot containing the rows of  $\hat{X}P$  gives us a description of the fitted values.

The plot containing the rows of  $P$  tells us about the relationship between the full fitted value space and the low-dimensional approximation of that space. Each variable point represents where a fitted value would fall in the space if it had a value one unit above the mean for that variable and exactly at the mean for the other variables.

The plot containing the rows of  $BP$  tells us about the relationship between the predictor variables and the space with the low-dimensional approximation of the fitted values. Each variable point in this plot again tells us where the fitted value for a sample would fall if it had a value for that predictor variable one unit above the mean and at the mean for the remaining variables.

Therefore, we can use the PCA-IV plot to visualise the fitted values from our model, as well as the relationship between the fitted values and the predictor variables (Table 3).

#### 2.5.4. Quantifying within-class variance

The final set of analyses we performed have to do with the variance of microstructural (linguistic) information within each class. To quantify the overall spread of the samples within each class in the multi-dimensional space, we used  $\text{tr}(\hat{\Sigma}_i)$ , where  $\hat{\Sigma}$  is a  $p \times p$  giving the estimated covariance within class  $i$ . In one dimension, this would simply be the estimated within-class variance. In more than one dimension, as we have here, this quantity is an aggregate measure of variance across all of the variables. The other statistic we used to quantify the within-class variance was the correlation between the scores on the first and second BCA axes within each class. Again, if  $\hat{\Sigma}_i$  is the estimated covariance matrix within class  $i$  and  $v_1$  and  $v_2$  are the loadings of the variables on the first and second PCA-IV axes, our statistic is

$v_1^T \hat{\Sigma}_i v_2 / \sqrt{v_1^T \hat{\Sigma}_i v_1 \cdot v_2^T \hat{\Sigma}_i v_2}$ . This measure is complementary to  $\text{tr}(\hat{\Sigma}_i)$  in that it tells us about the shape of the distribution within class  $i$ , while  $\text{tr}(\hat{\Sigma}_i)$  tells us about the overall size of the distribution within class  $i$ .

#### 2.5.5. Data considerations

Before applying any of the downstream methods, the data were transformed to remove the effect of certain demographic variables and to bring the data more in line with the normality assumptions implicit in several of the methods. A started log transformation was applied to each language variable. Note that a log transformation would allow us to read off ratios between variables as differences between the log-transformed variables (for instance, the log ratio of plurals to nouns would be the difference between log plurals and log nouns), and the started log transformation approximately retains this characteristic while allowing us to use variables that take a value of zero. In addition, age, years of education, and log-transformed number of words were regressed out of the transformed microstructure matrix. BCA was then performed using the `ade4` package in R. The confidence intervals provided are bootstrap intervals, with the resampling done by subject to preserve the hierarchical structure of the data.

#### 2.5.6. Data and code availability

AphasiaBank data is supported by NIH-NIDCD grant R01-DC008524 for 2007–2022 (MacWhinney et al., 2011). Access is granted after researchers read the ground rules (<https://talkbank.org/share/>). AphasiaBank is password protected for participant identity protection. Code used in this paper is available by contacting either author.

### 3. Results and discussion

#### Hypotheses

Our specific hypotheses were as follows (as discussed between the authors in August 2019):

- (1) Within-genre tasks would produce the most comparable microstructure. We also hypothesised, given prior work by our group (Stark, 2019), that we would observe this pattern in both the control group and the aphasia group, despite the aphasia group producing fewer words overall.
- (2) The interaction of task with discourse microstructure would differ by aphasia type. We hypothesised that overall aphasia severity would play some role, in that the interaction of task and linguistic variables in

**Table 3.** Summary statistics for the Western Aphasia Battery - Revised Aphasia Quotient (AQ) subcomponent scores for the aphasia group.

Variable	Mean	SD	Maximum score allowed
SpontSp	15.01	3.66	20
AudVbl	8.44	1.48	10
Rep	6.85	2.47	10
Naming	7.18	2.35	10

Notes: These four categories represent four language categories, and the score from each is used in the calculation of overall aphasia severity, AQ. SpontSp, Spontaneous speech, comprising scores relating to informational content and fluency of a spontaneous picture description and short conversation; AudVbl, Auditory Comprehension, comprising yes/no questions, auditory word recognition, sequential commands subtests; Rep, Repetition, comprising scores relating to word- and sentence-level repetition abilities; and Naming, comprising scores relating to object naming, word fluency, sentence completion and responsive speech subtests.

speakers without aphasia by WAB (“not aphasic by WAB” group) and speakers with Anomic aphasia would demonstrate discourse microstructure most comparable to the control group. In other words, these groups would show a similar clustering of linguistic variables by task and similar within-group covariance structure.

- (3) Finally, we anticipated that aphasia types associated with more severe language impairments (e.g. Broca’s aphasia, Wernicke’s aphasia) would not demonstrate as strong of a task  $\times$  microstructure interaction, in that both within-genre tasks and across-genre tasks (e.g. Cinderella compared with Window) would demonstrate considerable overlap in linguistic variables. This hypothesis speaks to our theory that linguistic information is inherently limited in this population (i.e. severely impoverished language due to aphasia), and may not change with task demands.

### 3.1. Group comparison

For each task and group (aphasia, control), we statistically compared demographic and basic language information using  $t$ -tests with Bonferroni correction. For years of education, there was not a significant difference for subjects included in the Important Event task ( $p = .48$ ), Cinderella ( $p = .17$ ) or Window ( $p = .11$ ); there was a significant difference for Cat ( $p = .046$ ) and Sandwich ( $p = .01$ ). For age (years) at testing, there was a significant difference for all tasks ( $p < .05$ ), as the aphasia group had a tendency to be older. For sex, there was not a significant difference for Important Event ( $p = .96$ ) or Sandwich ( $p = .20$ ) but there was a significant difference for Cinderella ( $p < .01$ ), Cat ( $p < .01$ ) and Window ( $p < .01$ ). Finally, for total tokens produced, there was a significant difference for all tasks ( $p < .001$ ), where speakers with aphasia produced fewer total tokens.

For these reasons, we controlled for these demographics and language information in downstream analyses.

### 3.2. Discourse microstructure is specific to genre and task

#### 3.2.1. Evidence from the control group

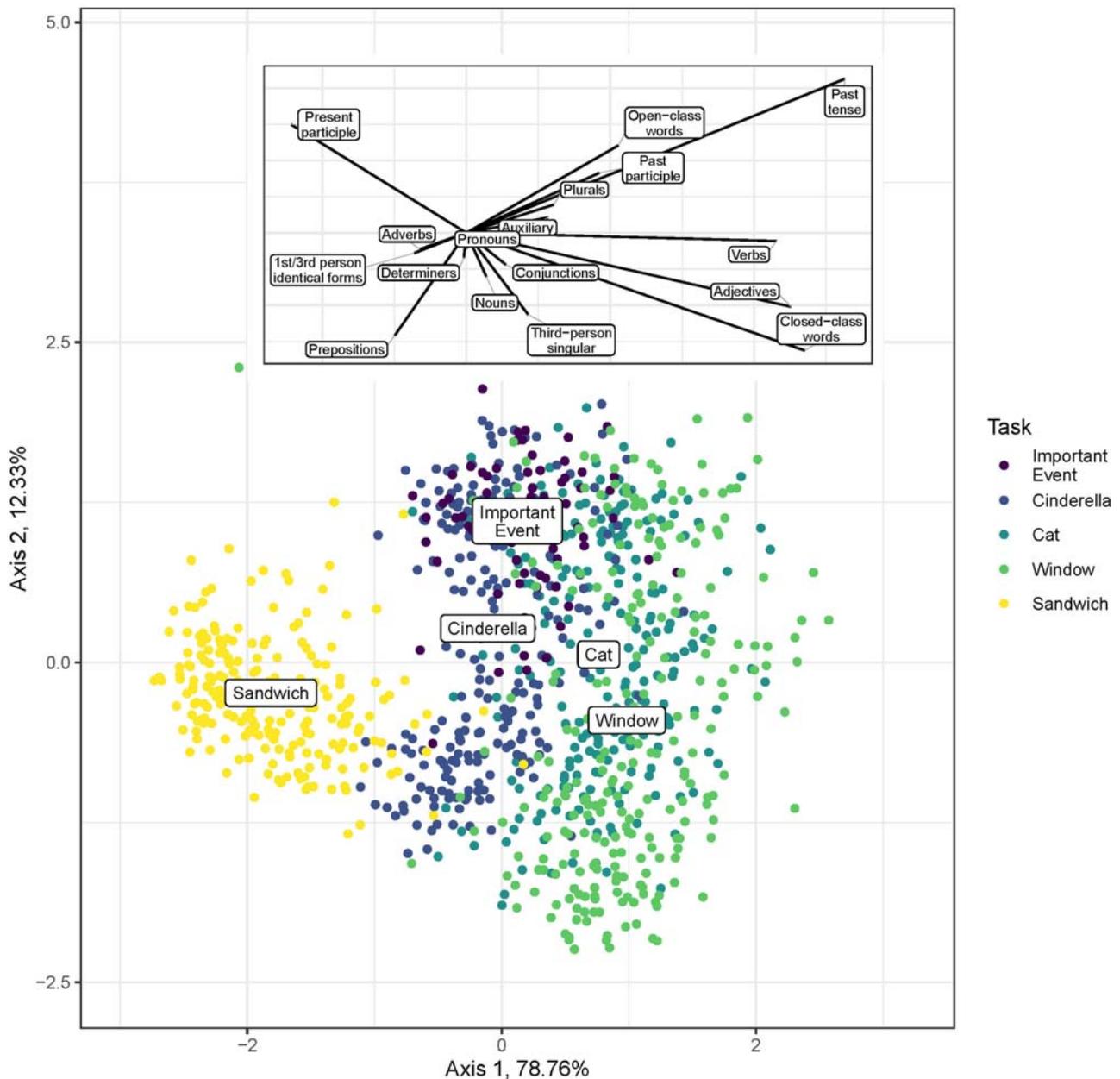
Task explained 26.8% (25.4%, 28.7%) of the variance of discourse microstructure in the controls. The loadings of the linguistic variables in BCA space are shown in the sub-box in Figure 2, where the two axes explain a total of 90.84% of the variance. This reduced lower-dimensional space robustly reflects the linguistic

variables of interest. Note that the  $y$ -axis tends to reflect word class, with open class loading higher and closed class loading lower, whilst the  $x$ -axis tends to reflect part of speech and grammar, with verbs, adjectives and past tense loading to the right and nouns, present participle, and determiners to the left. It should be noted that CLAN MOR program will tag semi-auxiliary constructions (e.g. “are going to make”, “are going to need”) as present participles (top left corner). One can directly map this sub-box to the larger plot, showing, for example, that the Sandwich task loads left of centre on the  $x$ -axis, suggesting present participle, preposition and noun usage.

First, we note that the narrative tasks occupy a similar space. The Important Event and Cinderella tasks cluster together in the BCA space (as demonstrated in Table 4, there is considerable overlap in their class mean confidence intervals), suggesting a similar linguistic profile. Similar to what was observed with the two narrative tasks, the two expository tasks cluster together (Cat, Window) in space (Table 4). Notably, the procedural genre (Sandwich task) was positioned significantly further from all other tasks (Table 4). The means and confidence intervals for the Sandwich task do not overlap with those of any other task, signifying a true dissociation in linguistic microstructure from both narrative and expository genres.

We next evaluated within-genre dissociation of tasks – that is, despite being a similar genre (e.g. narrative), does each task have unique microstructure? While discourse microstructure is indeed most similar within-genre, the tasks within the narrative genre dissociate. Table 6 clarifies that Important Event and Cinderella (both narrative) have significantly different positions on each axis. On Axis II, the Cinderella task loads somewhat negatively ( $M = -0.20$  [ $-0.27, -0.14$ ]) while the Important Event loads somewhat positively ( $M = 0.36$  [ $0.29, 0.45$ ]). On Axis I, the Cinderella task loads somewhat negatively ( $M = -0.13$  [ $-0.18, -0.07$ ]) while the Important Event loads near zero ( $M = 0.03$  [ $-0.06, 0.12$ ]). Unlike the Cinderella and Important Event tasks, the Cat and Window tasks are not significantly differently positioned in either Axis I or Axis II. Both tasks similarly (overlapping confidence intervals) load positively on Axis I (Cat,  $M = 0.87$  [ $0.63, 1.11$ ]; Window,  $M = 1.01$  [ $0.96, 1.09$ ]) and negatively on Axis II (Cat,  $M = -0.47$  [ $-0.59, -0.25$ ]; Window,  $M = -0.45$  [ $-0.54, -0.37$ ]). The Sandwich task is positioned negatively along Axis I ( $M = -1.74$  [ $-1.80, -1.67$ ]) and slightly negatively along Axis II ( $M = -1.03$  [ $-1.10, -0.97$ ]).

Notably, the control group produces relatively homogeneous, task-specific linguistic information. We can



**Figure 2.** Between-class analysis on the control population. The main panel shows projections of task means (labeled boxes) and individual samples (points) onto the BCA axes. The inserted image shows the loadings of the linguistic variables on the BCA axes. The projection of the task means suggests similarities between tasks from the same genre. The projections of the samples (each subject is a point) indicate that while there is within-task variance, the language produced in response to the different tasks are quite distinct and differentiable by task genre.

appreciate the covariance structure of the control group in Figure 2 and Table 7. That is, the linguistic information produced by each member of the group tends to cluster quite closely together for every task. This is first affirmed by our small confidence interval range in Table 6. This is further confirmed by examining the trace of the empirical covariance matrix shown in Table 7, where the control group has the smallest covariance trace in every task.

### 3.2.2. Evidence from the aphasia group

We were interested in modelling language usage by a combination of task and either aphasia type or AQ sub-component score. To this end, we performed multivariate regressions using the four AQ subcomponent scores (naming, repetition, spontaneous speech, comprehension), type of aphasia, and task, either on their own or in combination with the others. Task alone explained a smaller fraction of the variance among the

**Table 4.** Distances between tasks for the controls, measured in the full space. 95% bootstrap confidence intervals are reported in parentheses for each distance.

	Important event	Cinderella	Cat	Window
Cinderella	1.19 (0.98, 1.43)			
Cat	1.70 (1.59, 1.87)	1.46 (1.37, 1.56)		
Window	2.11 (1.95, 2.30)	1.83 (1.74, 1.93)	0.93 (0.84, 1.07)	
Sandwich	2.46 (2.34, 2.59)	1.90 (1.81, 1.99)	2.61 (2.52, 2.70)	2.83 (2.74, 2.94)

Notes: This table demonstrates the clustering of task genres, where closer distances demonstrate similar linguistic information produced during the task (i.e. Important Event/Cinderella [narrative], Cat/Window [expositional], Sandwich [procedural]).

aphasia group: 10.4% (9.6%, 11.8%). Aphasia type alone explained a smaller fraction of the variance than task: 4.9% (4.2%, 7.5%). Task and aphasia type together explained 18.1% (18.5%, 22.8%) of the variance. Although the distances between classes are on average smaller in the aphasia groups, the smaller proportion of variance explained by task in the aphasia group than the control group is at least in part due to a large within-group variance by task among the aphasia groups, which we will highlight below. Additive models with task and AQ scores or task and type of aphasia explained similar fractions of the variance in the linguistic variables. Models with an interaction between AQ scores and task or between aphasia type and task did not explain substantially more of the variance in the linguistic variables than did the corresponding additive models. As AQ score is a linear combination of the four individual AQ subcomponents, we could also test whether a model with task and the four AQ subcomponents explained substantially more of the variance than a model with task and the overall AQ score: we found that it did ( $p = .02$ ). This led us to the model with additive terms for task and the four AQ subcomponents as the best model of language usage.

As shown in Figure 3 and Table 5, the relationships between the tasks within the aphasia group mirror the relationships between the tasks in the control group.

First, on the bottom left of Figure 3, we can appreciate the loading of discourse genre and task by aphasia type, just as we did for controls in Figure 2. Note that narrative tasks (three in the case of aphasia: Important Event, Stroke, Cinderella) occupy a space in the top left-to-centre quadrant of the BCA space. The two expositional tasks (Cat, Window) occupy a far-right quadrant

**Table 6.** Average distance between tasks by aphasia type and control group. Stroke story was excluded from the computations so as to enable comparison with the controls.

	Average distance between tasks
Control	1.93 (1.87, 2.00)
Not aphasic by WAB	1.93 (1.85, 2.13)
Anomic	1.87 (1.80, 1.99)
Wernicke's	1.57 (1.53, 1.96)
Conduction	1.55 (1.45, 1.77)
Broca's	1.25 (1.17, 1.55)

in the BCA space. One can immediately appreciate that the distances between the tasks is not as large for any of the aphasia types as for the control group (Table 6). Notably, and similar again to the control group, the Sandwich task occupies the lower left-hand quadrant of the BCA space, and, for every aphasia type group, the Sandwich task did not overlap with any other tasks.

Like in the control group's narrative genre, Cinderella task was significantly different from both the Important Event and the Stroke tasks in all aphasia type groups (Table 6). Notably, for most aphasia type groups, Cinderella's mean on Axis I and Axis II did not overlap with the confidence intervals of Important Event or Stroke on either axis.

We have also projected this information into a visualisation divided by task, which is located in the top left of Figure 3. This allows us to appreciate the relative location of the task in the BCA space, and the general overlap of task-specific linguistic information produced by each aphasia type. As a whole group, the aphasia group does not produce the same extent of task-specific linguistic homogeneity that we found in the control group. This covariance structure seems to be modulated both by aphasia type (e.g. Broca's) and overall aphasia severity (e.g. severe) (Table 7). In fact, overall aphasia severity consistently shows a negative slope between the two BCA axes (Figure 4). No aphasia type shows such a consistent, negative relationship between the two axes, suggesting that overall aphasia severity may explain a large portion of the covariance shown in the aphasia group. That is, those individuals with the most severe aphasia demonstrate less distinction between tasks (i.e. produce similar linguistic information in dissimilar tasks, as is shown in Figure 3 and Table 6). It is also helpful to appreciate the shape of the groups in the bottom left of Figure 3, where there is variation within the NABW and the Anomic aphasia group along the Axis I but there is variation along both axes for the Conduction, Wernicke's and Broca's aphasia groups. The NABW and Anomic group covariance structure most resemble the control group's (Table 7).

See Table S1 for difference in distances between tasks, separated by each aphasia type.

**Table 7.** Within-class dispersion, as measured by the trace of the empirical covariance matrix for the class, and bootstrap confidence intervals.

Group	Cat	Cinderella	Important Event	Sandwich	Stroke	Window
Control	4.0 (3.7, 4.2)	2.8 (2.6, 3.0)	2.2 (1.9, 2.6)	2.6 (2.4, 2.8)	NA	4.8 (4.5, 5.1)
NABW	4.5 (3.4, 5.1)	3.0 (2.0, 3.8)	3.1 (2.3, 3.8)	2.8 (2.1, 3.2)	1.6 (1.0, 2.0)	5.8 (4.1, 6.9)
Anomic	6.5 (5.5, 7.4)	4.9 (4.1, 5.7)	4.9 (4.2, 5.6)	4.7 (3.7, 5.8)	3.6 (3.0, 4.2)	7.0 (6.1, 7.8)
Conduction	6.4 (5.0, 7.7)	4.6 (3.5, 5.7)	5.3 (4.2, 6.2)	5.2 (3.8, 6.2)	3.9 (2.8, 4.8)	7.5 (6.2, 8.5)
Wernicke's	7.5 (4.9, 9.4)	5.1 (3.2, 6.4)	5.8 (4.1, 7.0)	5.7 (3.1, 7.2)	3.3 (2.3, 4.2)	7.5 (5.4, 8.9)
Broca's	10.9 (8.8, 12.7)	9.3 (7.7, 10.6)	10.2 (8.8, 11.2)	8.3 (6.0, 9.2)	8.9 (7.7, 9.9)	12.2 (10.0, 14.0)

Notes: This table demonstrates the large variance within the Broca's aphasia group compared with most other groups for each task. It also demonstrates the relatively small variance for the control and NABW groups. NABW, not aphasic by WAB.

Together, these results support our hypothesis that individuals with severe overall aphasia likely have limited linguistic information available to them, limiting the language breadth able to be produced. There is nuance to this interpretation, however, which we discuss in the discussion. Of importance, is that the procedural discourse is always clearly defined in our BCA space (bottom left of Figure 3), demonstrating that even the more severe individuals don't produce linguistic information during procedural discourse that resembles linguistic information during expository or narrative genres.

### 3.2.3. Linguistic information is elicitation method-specific

We created a two-dimensional space representative of the full space of linguistic variables for our control group and for our aphasia group, separately. Importantly, these two language spaces (as shown in Figures 2 and 3) resemble one-another. For instance, the x-axis (Axis I) appears to relate to morphosyntax and the y-axis (Axis II), lexical information. This is, of course, an over-simplification for the purposes of understanding the multidimensional space represented by our linguistic variables, but it does aid in interpretation. Additionally, in the aphasia group, we took into account variables drawn from a neuropsychological battery (the WAB), because doing so better described the relationship of task with linguistic variables (bottom right of Figure 3). Altogether, we can use the BCA space to understand the relative difference in microstructure for each discourse genre and task. For regression coefficients of linguistic variables and neuropsychological variables on task in the aphasia group, refer to Table 8.

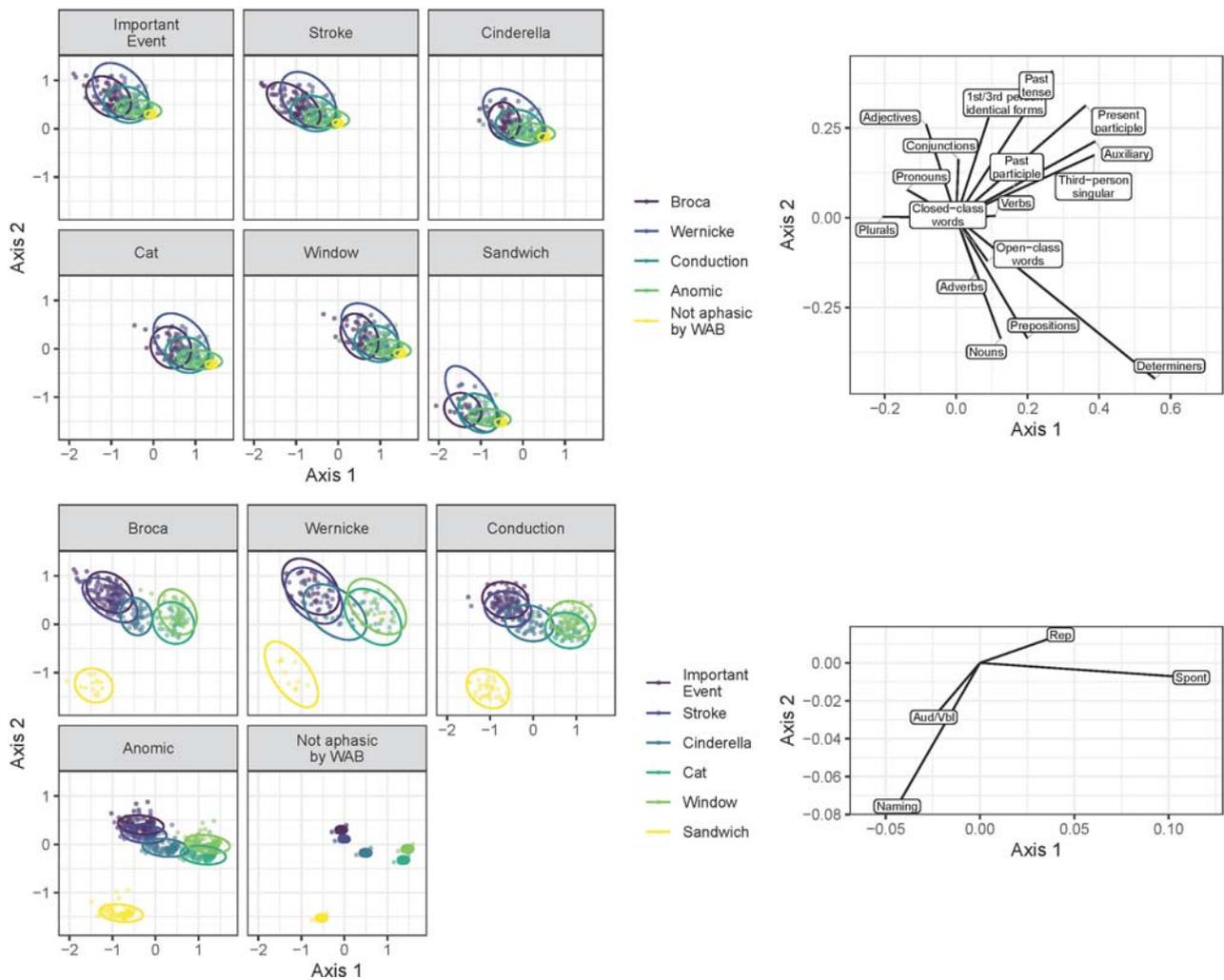
The procedural discourse (Sandwich) loads in the bottom left quadrant of BCA space for both subject groups. As such, we can surmise that its morphosyntax is mostly present tense, and that its lexical information is predominantly nouns and prepositions. As noted in Table 8, which demonstrates the regression coefficients for the aphasia group, we can appreciate the extent to which nouns, prepositions and conjunctions load most positively, and the high negative loadings of past

tense, participles (present and past), and auxiliaries. This loading jives with prior research by our group suggesting that the Sandwich task, representative of the procedural genre, produces morphosyntactically simple phrasing with a reliance on nouns (Stark, 2019).

The narrative discourse genre (Important Event, Cinderella, and Stroke [aphasia group only]) loads in the upper half of the y-axis (Axis II) and relatively central on the x-axis (Axis I) in BCA space. We interpret this to mean a mix of past and present tense usage and, in terms of lexical information, fewer nouns produced. Again referring to Table 8 for loadings of linguistic variables onto narrative genre tasks for the aphasia group, we see positive loadings for plurals (including first person plural), past tense, adjectives, conjunctions and pronouns for the Important Event and the Stroke tasks. We see less reliance on nouns, determiners and adverbs for these tasks. In contrast, linguistic variables loading most positively for the Cinderella task include plurals, past tense and adjectives, as the other narrative genres did, but also nouns and determiners and pronouns. Pronouns are understandably higher in this task, likely due to subject-related local coherence (e.g. talking about Cinderella often, resulting in referencing "her" often).

The expository discourse genre (Cat, Window) loads to the right on the x-axis (Axis I) and relatively central on the y-axis (Axis II) in BCA space. Positive loadings for linguistic variables suggest that both Cat and Window tasks are marked by nouns, verbs, auxiliaries, third person plurals, both past and present participles, and determiners (Table 8) in the aphasia group. Negative loadings suggest relatively little use of pronouns and adjectives.

Recall that, in the aphasia group, task performance was best explained by an interaction with linguistic variables as well as neuropsychological variables (i.e. language subtests from the WAB, including auditory verbal comprehension, spontaneous speech, naming and repetition). Therefore, we are also able to evaluate the extent to which underlying language processes (tapped by these four subtests) explain microstructure by genre and by task. The bottom right hand



**Figure 3.** PCA-IV corresponding to multiple regression of linguistic variables AQ subcomponent scores. For each sample, we have both the predictor variables (AQ subcomponent scores) and the response variables (language). For each sample, the multiple regression model gives us  $p = 17$  fitted values, one for each variable. These fitted values are projected down into a lower-dimensional space that maximises the variance in the fitted values (left panels, top and bottom). In the left panels, each point represents the fitted values for one sample. Those fitted values can be described as linear combinations of either the response/linguistic variables or the predictor/AQ subcomponent variables. The right-hand panels give the weights for either the linguistic variables (top) or AQ subcomponent variables (bottom) that describe the fitted values for the samples. We see that although a substantial amount of the variance in the fits can be explained by task, aphasia type also influences language production in a systematic way. In addition to the composition being different by aphasia type, the variance within a group also changes with aphasia type, with the less severe types of aphasias exhibiting lower variance.

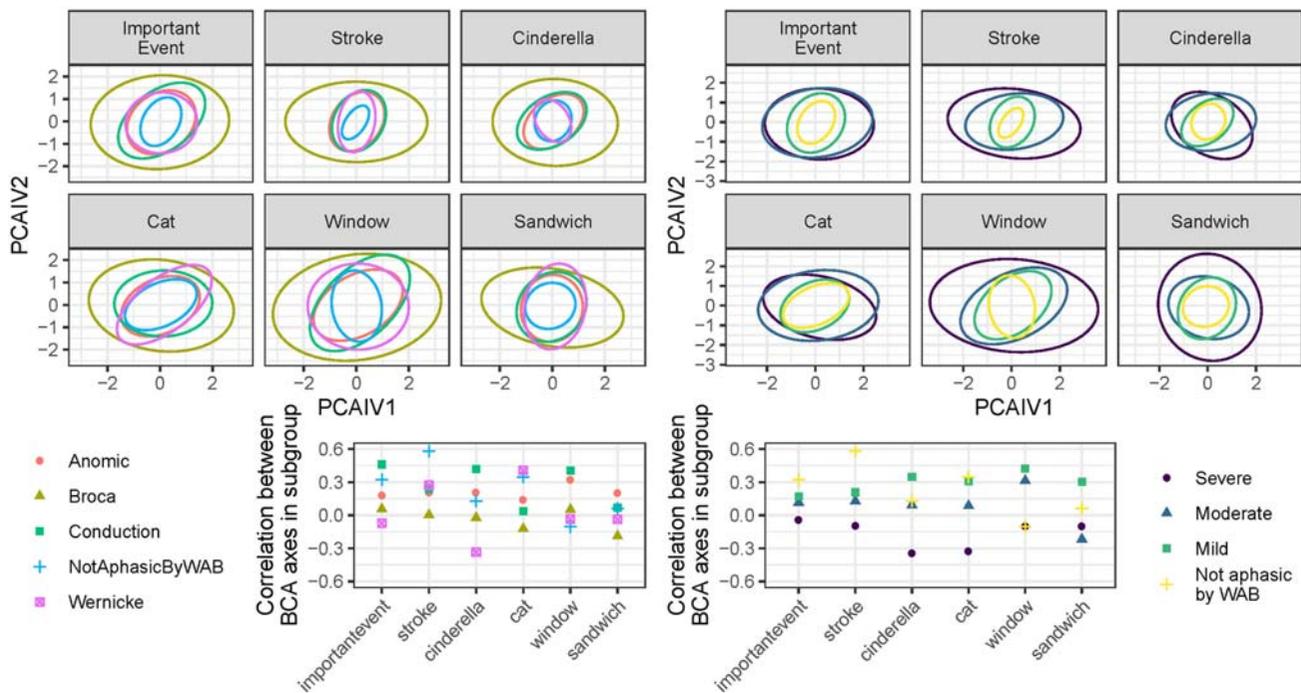
corner of Figure 3 shows the relative loading space of these variables, similar to what we have been looking at for language variable space (top right of Figure 3). If we look at the bottom left of Figure 3, where the tasks

are plotted by aphasia type, we are able to appreciate the extent to which each task is explained by the neuropsychological variables in tandem with the linguistic variables. We can surmise that the procedural task,

**Table 5.** Distances between tasks for the aphasia group, measured in the full space.

	Important event	Stroke	Cinderella	Cat	Window
Stroke	0.57 (0.48, 0.66)				
Cinderella	0.92 (0.82, 1.02)	0.94 (0.84, 1.04)			
Cat	1.85 (1.73, 1.99)	1.69 (1.56, 1.82)	1.16 (1.03, 1.29)		
Window	1.78 (1.65, 1.91)	1.66 (1.52, 1.80)	1.17 (1.03, 1.30)	0.83 (0.72, 0.96)	
Sandwich	1.99 (1.84, 2.13)	1.80 (1.67, 1.94)	1.80 (1.66, 1.94)	2.27 (2.12, 2.42)	2.40 (2.24, 2.58)

Notes: 95% bootstrap confidence intervals are reported in parentheses for each distance. This table demonstrates the clustering of task genres, where closer distances demonstrate more similar linguistic information produced during the task (i.e. Important Event/Cinderella [narrative], Cat/Window [expositional], Sandwich [procedural]).



**Figure 4.** Covariances (top, visualised as ellipses) and correlations (bottom) for each combination of task and type (left) or overall aphasia severity (right). Overall aphasia severity has a more systematic relationship to the covariance structure than aphasia type. The covariance overall is the largest for the severe group, and decreases as overall aphasia severity decreases. The severe group is the only one that has a consistently negative slope. Within each task, the overall variance decreases and the major axis of the ellipse rotates around. “Severe” severity tends to have a slightly negative slope.

which loads in the far bottom left quadrant of neuropsychological and language variable space, likely relies on an individual’s naming ability. Naming is reflective of lexical-semantic production capacities, and together with prior evidence that the procedural task is most reliant on production of content words and relatively simple morphosyntax (e.g. Stark, 2019; Ulatowska et al., 1981) suggests that procedural discourse may be a task that engages lexical-semantic ability but not morphosyntactic ability. It also suggests that those performing better on the naming subtest of the WAB (and those who, by proxy, have relatively accessible lexical-semantic processes), may be those driving the linguistic pattern we see in both aphasia and control groups, i.e. the production of content words. There is also a sanity check built into this analysis, which is that the Spontaneous Speech subtest from the WAB (“Spont” in the BCA space) should be logically most associated with the expository discourse genre tasks (Cat, Window). This is because the Spontaneous Speech score is acquired by rating informational content and fluency of the subject’s speaking during a different picture description task (the Picnic Scene), as well as during a basic conversational interview (Kertesz, 2007). Therefore, there should be similarity between one single picture description expository task (Picnic Scene) and others (Cat) if indeed there is some generalisability across

testing procedures and time, which is what we see here. Auditory Verbal Comprehension and Repetition scores do not appear to be particularly associated with microstructure on any of the tasks.

### 3.3. Results summary

We demonstrate clear between-genre dissociation and considerable within-genre shared microstructure in both aphasia and control groups. In both aphasia and control groups, there is a noticeable deviation between narrative tasks (e.g. Cinderella from Important Event and/or Stroke [in aphasia group]), suggesting that the Cinderella task may rely on different cognitive components than its narrative neighbours. Meanwhile, both the Cat and the Window expository tasks do not have significantly different locations on either axis for either subject group (aphasia, control), suggesting that expository tasks like these share considerable similarity in frame selection and, therefore, shared downstream linguistic selection. The procedural task (Sandwich) boasts microstructure unlike any of the others.

In relation to homogeneity of within-group speakers, task-specific linguistic information produced by the control group is relatively homogeneous. Therefore, while there is some variance among control speakers (as demonstrated across many prior studies) (S. G. H.

**Table 8.** Regression coefficients for task and Western Aphasia Battery (WAB) language subtest scores as predictors of linguistic variables in the aphasia group.

Language variable	SpontSp	AudVbl	Rep	Naming	Imp. event	Stroke	Cinderella	Cat	Window	Sandwich
Plurals	-0.019	0.020	-0.004	0.027	0.252	0.020	0.215	-0.055	-0.510	0.079
Nouns	0.000	-0.004	-0.005	0.101	-0.143	-0.202	0.036	0.052	0.063	0.194
Verbs	0.042	0.005	0.014	-0.051	-0.011	0.004	-0.037	0.045	0.086	-0.087
Aux	0.044	-0.041	0.021	-0.042	-0.093	-0.108	-0.002	0.489	0.344	-0.630
3S	0.018	-0.047	-0.020	0.005	-0.197	-0.235	0.015	0.516	0.463	-0.561
1S.3S	0.029	0.032	0.037	-0.061	0.243	0.115	0.077	-0.038	0.075	-0.471
PAST	0.050	0.067	0.038	-0.070	0.252	0.096	0.201	-0.066	0.367	-0.850
PASTP	0.007	-0.011	0.016	0.000	0.039	-0.157	0.098	0.116	0.184	-0.279
PRESP	0.034	-0.049	0.014	-0.019	-0.089	0.073	-0.083	0.412	0.425	-0.738
prep.	0.072	0.077	0.038	-0.081	-0.098	-0.193	-0.101	0.015	0.120	0.258
adj.	-0.006	-0.056	0.009	0.012	0.309	0.030	0.151	-0.211	-0.020	-0.259
adv.	0.049	-0.035	0.000	-0.051	-0.128	-0.033	-0.115	0.179	-0.110	0.207
conj.	0.009	-0.014	0.006	-0.038	0.152	-0.002	-0.016	0.130	-0.079	-0.185
det.	0.063	-0.007	0.022	0.010	-0.579	-0.481	0.107	0.491	0.471	-0.009
pro.	0.042	0.016	0.003	-0.108	0.183	0.091	-0.080	-0.353	-0.010	0.168
open class	0.014	-0.005	0.002	0.029	-0.043	-0.088	0.026	0.016	0.062	0.026
closed class	0.025	0.012	0.003	-0.039	-0.088	-0.038	-0.016	0.044	0.024	0.074

Notes: SpontSp, spontaneous speech subtest from the WAB; AudVbl, auditory verbal comprehension subtest from the WAB; Rep, repetition subtest from the WAB; Naming, naming subtest from the WAB; Imp. event, important event discourse task.

Dalton & Richardson, 2019; Fergadiotis & Wright, 2011), neurotypical speakers produced relatively similar, task-specific microstructure. This is an important finding because it provides a normative comparison for the aphasia group; confirms the need for consistency of task instructions and demands within a study; and demonstrates that linguistic information logically and reliably shifts according to task demands. On the other hand, there was large covariance within each aphasia type, showing differential task-specific microstructure according to aphasia type. Overall aphasia severity was the most influential of the aphasia characteristics, mediating the relationship between microstructure and task.

To summarise the findings in reference to our original hypotheses: (i) Each genre (narrative, expository, procedural) elicits relatively unique linguistic information (with some overlap, i.e. expository and narrative) in both the control group and the aphasia group. Broadly, tasks within a genre demonstrate comparable microstructure in both the control group and the aphasia group; and (ii) The interaction of task with linguistic information is at least partially related to aphasia characteristics (type and overall aphasia severity). The task  $\times$  aphasia (type, overall aphasia severity) interaction likely drives the heterogeneity in the aphasia group as a whole and the within-type covariance. Further, individuals whose aphasia types tend to be associated with greater overall aphasia severity (e.g. Broca's, Wernicke's aphasia) did not demonstrate task dissociation as strongly. There is some nuance to interpreting the influence of overall aphasia severity on microlinguistic information extracted from spoken discourse. Overall aphasia severity is representative of more than just spoken language ability (indeed, it incorporates scores from auditory comprehension,

naming, repetition, and spontaneous speech), and the result that persons with severe aphasia tend to demonstrate microlinguistic information that is not as strongly dissociable by task must be interpreted through the lens of aphasia type and characteristics. For example, this finding may be driven in one way by individuals with severe Broca's aphasia who have difficulty producing a variety of linguistic information, or by individuals with severe Wernicke's aphasia who have difficulty producing meaningful, well-structured (grammatical) information. These two things may therefore be occurring, regardless of task, thus resulting in a similarity of microlinguistic structure across tasks for individuals with overall severe aphasia. It should be noted that nearly half of our Wernicke's sample, and roughly 35% of our Broca's sample had overall severe aphasia, meaning that the rest of the individuals with Broca's and Wernicke's aphasia presented with overall moderate aphasia. Note, too, that we did not include individuals with Global aphasia in this analysis, as there were very few in the database. Future work with a larger sample of individuals with Global aphasia and severe Broca's and Wernicke's aphasia could test our hypothesis, that overall aphasia severity results in microlinguistic information that is similar across discourse elicitation methods. As it stands, the interaction of overall aphasia severity and type must be taken into account.

#### 4. Implications, future directions and limitations

We provided evidence that microstructure is similar within-genre and relatively dissociable across-genre, in speakers with and without aphasia. We demonstrate

that overall aphasia severity has some influence, though relatively small, on mediating the relationship between task and microstructure (i.e. greater overall aphasia severity reflects increasingly similar microstructure across tasks). Below, we discuss basic and translational implications for this work.

#### **4.1. Clinical implications**

Perhaps the most innovative portion of the study, we evaluated the extent to which aphasia characteristics, like type and overall aphasia severity, mediated task- and genre-specific microstructure. The goal of this was to explore any situations where acquiring multiple, varied discourse samples may not be altogether more helpful than acquiring a single elicitation method. In summary, we determined that there are genre- and task-specific microstructural patterns regardless of aphasia type and/or overall aphasia severity. These results are indicative of a few things. While severe aphasia does seem to blur elicitation method boundaries (i.e. those with more severe aphasia tend to produce similar microstructure during all tasks), overall aphasia severity does not completely blur these task boundaries. Therefore, even in speakers with severe Broca's or Wernicke's aphasia, we would expect, at the very least, different microstructural information to be produced during procedural discourse compared with other discourse genres. But, these results beg the question of how language sampling strategies might differ in those with severe aphasia, as we showed that linguistic information was not particularly different for narrative or expository tasks for these individuals. One might argue that, for those with severe aphasia, acquiring a sample from several discourse genres may not be necessary to comprehensively evaluate that individual's language abilities. But, we would hesitate to argue as such because each discourse genre (and task) likely produces informative, additional language and communicative information beyond microstructural information. Indeed, recall that linguistic information (microstructure) is one of three hypothesised levels of analysis that can and should be done on discourse (Frederiksen et al., 1990; Sherratt, 2007), with the other levels assessing more macrostructural/functional information. For instance, research has shown that certain discourse genres, like expository and narrative, produce unique gesture types and gesturing loads (i.e. number of linguistic-accompanying gestures, or standalone gestures) (Sekine & Rose, 2013). Additionally, individuals with severe aphasia may still be able to produce task-relevant words (e.g. components of core lexicon and/or main concepts), even if the linguistic structure is the same

(S. G. Dalton & Richardson, 2015; S. G. H. Dalton & Richardson, 2019). Therefore, collecting language samples across genres provides information about microstructure as well as information that is very useful for understanding communicative ability as a whole.

Our results further highlight wide intra-group (or, within-group) variability. That is, speakers with a shared aphasia type (e.g. Broca's aphasia) produced variable linguistic information during a task. This observation has also been appreciated for language function. Main concept production is one means of evaluating language function by evaluating the ability of the speaker to cover the main topics that should logically be produced during the task (e.g. in the Cinderella story, speakers should discuss the glass slipper, the fairy godmother, coming home at midnight, etc., Richardson and Dalton (2016). In a recent large study evaluating main concept production across several discourse elicitation methods from AphasiaBank in speakers with and without aphasia, there was considerable within-group heterogeneity (S. G. H. Dalton & Richardson, 2019). Like with the microstructural within-group variation across discourse genres shown here, the within-group heterogeneity for main concept production was much larger for aphasia types that typically include both moderate and severe aphasia presentation (e.g. Broca's, Wernicke's) (S. G. H. Dalton & Richardson, 2019). Altogether, our work in microstructure and other groups' work in functional measures of language demonstrate the complicated interaction of aphasia type and overall aphasia severity with language.

Altogether, we would argue that prioritising collection of several discourse genres is as important in those with severe aphasia. While this is indeed time consuming, there has been considerable improvements in transcription and analysis, like automatic speech to text [in populations with communication disorders] (K. Fraser et al., 2013; Le et al., 2018; Le & Provost, 2016) as well as checklists to be used during the live task (S. G. H. Dalton et al., 2020).

#### **4.2. Transdiagnostic implications**

We expect that our findings (i.e. that within-genre microstructure is similar and across-genre microstructure is dissimilar) would hold when investigated in populations of developmentally disparate speakers, due to frame selection being at the earliest level of discourse hierarchy, and thus influencing downstream (micro- and macrostructural) processes (Frederiksen et al., 1990; Sherratt, 2007). Bliss and McCabe (2006) discuss the notion that personal narratives (like our "Important

Event" here), when used to elicit language samples in children, tend to also be characterised by use of the past tense and first and/or third person (Bliss & McCabe, 2006). Children reciting a fictional narrative (like a story retell, akin to our "Cinderella" story) may elicit more utterances and be viewed as an overall "easier" task because of the given, structured information (e.g. the story components in a picture book). But, this fictional narrative may elicit more indefinite speech because of the shared referents (e.g. both child and experimenter have seen the story book), which is very similar to the greater proportion of pronoun and indefinite words used in the Cinderella task (compared with the other narrative tasks) in our results. Together, these findings further clarify different aspects of frame selection, the first step in this kind of discourse. On the one hand, there are task instructions, which we know can influence frame selection even when the same stimuli is presented (Wright & Capilouto, 2009), and further, the shared referents between interlocutor and recipient/addressee further refine the frame. Therefore, our findings regarding discourse microstructure and task interaction replicate ideas about language sampling across development, and emphasise the importance of collecting language samples using multiple, varied discourse genres (which, in themselves, have variable instructions and referents) to comprehensively examine microstructural elements. It is not far-fetched to make the conjecture that it is equally important to acquire multiple discourse genres to assess other aspects of language, communication and cognition (e.g. different memory, attention and executive function systems at work in tandem with language).

In the field of aphasiology - and indeed, across many populations where behavioural intervention is considered best practice (e.g. speech-language therapy) - we are tasked with demonstrating that behavioural therapy generalises to more natural language processes. Speech-language therapy has a tendency to focus at the level of isolated processes, like noun/verb retrieval. There has been a considerable issue with demonstrating generalisability of therapy success to discourse in populations such as aphasia (Boyle, 2011; Rose et al., 2015), in that very few behavioural interventions have actually been shown to generalise to discourse. There are a few explanations for this. A contributing factor is the extraction of insensitive outcome measures from the discourse, such as choosing an outcome (e.g. mean length per utterance) that is not linearly related to the isolated process being trained (e.g. noun or verb retrieval) (Rose et al., 2015). There is a considerable amount of work to do to understand how isolated processes being trained in therapy are best represented by

variables extracted from discourse. One such way of doing so is to evaluate patterns of discourse (e.g. microstructural variables in a high-dimensional space) rather than single measures of discourse, as we have done here. There have been efforts to model multi-level discourse outcomes (Marini et al., 2011) but each level has comprised single outcome measures (e.g. mean length of utterance) and not variables modelled co-dependently. We emphasise moving away from (arbitrary) discourse outcome selection toward comprehensive modelling of all available variables. While we focused on examining microstructure, the analyses we describe here are possible for any group of variables that share some common space. For example, our statistical analysis could easily extend to examining syntax (e.g. dependency grammar). It would not be far fetched that other, more macrostructural variables, such as coherence and cohesion (Galletto et al., 2013), main concept (S. G. H. Dalton & Richardson, 2019) and/or core lexicon (S. G. Dalton & Richardson, 2015) analysis could be modelled in such a fashion, allowing us to appreciate how each of these variables describes a part of the functional language being used during each task.

An ultimate goal would be to use this statistical modelling to marry structure and function (or, micro- and macrostructural properties), thus providing a layered, high dimensional system that allows us to understand the relation between, and contribution of unique linguistic variables during each task. Further, analyses such as this are appropriate for transdiagnostic populations (Husain, 2017). For instance, language is affected in a variety of neurogenic and neuropsychiatric disorders, such as post-stroke aphasia, post-traumatic brain injury language impairments, major depressive disorder and primary progressive aphasia. Analysing spoken discourse in each of these populations, and modelling this information in a high-dimensional space, would allow us to sensitively explore shared (and disparate) patterns of language, as well as the interaction between language severity, disorder, and discourse elicitation method. This is a key future direction that we are presently exploring.

### **4.3. Theoretical implications**

Another explanation for the lack of generalisation from isolated processes being trained in therapy to discourse-level improvements is that we do not yet understand the cognitive-linguistic processes contributing to discourse tasks. For instance, we discussed how narrative discourse relies on processes of long-term memory, and dependent on the task, the memory system being relied

upon may be autobiographical or semantic. Further, multiword environments like discourse come with processes of language that may not be necessary for the processes being trained during therapy. For example, suppression of word alternatives (whether they are semantic, lexical, or phonological) co-occurs with morphosyntactic processes in multiword environments, whereas suppression of word alternatives in isolated tasks does not. While we have yet to understand how the cognitive-linguistic complexity of discourse best relates to isolated tasks, what we have shown here is that each discourse task/genre seems to preferentially tap into a part of the language system. For instance, we showed that speakers tend to preferentially produce more content words and simplest syntax during procedural discourse compared with narrative and expository discourse. This suggests that procedural discourse may be able to pick up on changes in content word retrieval and production, if indeed that is being trained in therapy. As it stands, researchers and clinicians may be choosing insensitive or inappropriate discourse tasks from which to extract language measures, but we hope that the present study, and other work highlighted in the introduction, are helpful for careful selection of discourse tasks.

Another important point that our results emphasise is the importance of task-specific interpretation in papers utilising monologic discourse. Just as we would not (and should not) assume that single words produced during one isolated task (e.g. confrontation naming) reflect the same cognitive-neural resources as single words produced during another, similar isolated task (e.g. verbal fluency), we cannot and should not make the assumption that linguistic information produced during one discourse task (or within one discourse genre) closely resembles linguistic information produced during another type of discourse task/genre. This has important implications for interpretation of studies employing monologic discourse. For example, we should not generalise the findings from a study employing fictional story retelling (e.g. Cinderella) to findings from a study employing autobiographical story telling (e.g. Important Event). Indeed, the emphasis on different memory systems may be a primary factor explaining the dissociation in linguistic information produced across narrative tasks, as well as other non-negligible factors like salience, motivation, and familiarity. This result, in tandem with the result that there is genre-specific microstructure, is an important finding for fields that acquire discourse as a means of understanding a person's linguistic (and indeed, cognitive) abilities in a more naturalistic setting than provided by standardised testing. Further, language can be

understood also at a broader (not just deeper) level by using multiple discourse genres. We believe that the push by others Brookshire and Nicholas (1994), S. G. Dalton and Richardson (2015), S. G. H. Dalton et al. (2020), Fergadiotis and Wright (2011), Nicholas and Brookshire (1995), Olness (2006), Richardson and Dalton (2016), Ulatowska et al. (1981) and ourselves Stark (2019) for the acquisition of multiple, varied discourse genres when sampling discourse gains more traction here, specific to evaluating microstructure.

Our results highlight that task instructions and, in following, experimental fidelity is critical. Fidelity is, of course, hyper-critical for all experiments, but it becomes increasingly challenging to ensure fidelity across studies where instructions are semi-structured and open to interpretation, like during monologic discourse and other naturalistic tasks. Our work and others Stark (2019) and Wright and Capilouto (2009) provide evidence for the establishment of fidelity during all stages of language sampling, from acquisition to transcription to coding to analysis. A recently founded working group, FOQUSaphasia (FOstering QUality of Spoken discourse research in Aphasia; [www.foqusaphasia.com](http://www.foqusaphasia.com); Stark et al., 2020) has endeavoured to create initiatives to foster fidelity and standardisation for discourse in aphasia. Without having these standards, it is difficult to reproduce or replicate prior findings in the field, and to draw conclusions (e.g. meta-analyses, scoping reviews) about spoken discourse across a variety of settings (e.g. assessment, therapy).

Finally, this study demonstrates that monologic discourse retains consistent microstructure within subject groups. Recall that members of the control group, as well as members of the aphasia group with mildest language impairment (i.e. Not Aphasic by WAB and mild anomic groups) produced relatively homogeneous genre- and task-specific linguistic information (i.e. smaller within-group covariance). This finding suggests that the tasks being employed here produce consistent linguistic elements when given to people with no language impairment or mild language impairment, and also demonstrates consistency across subjects of varying education, age, literacy, sex and presence of brain injury. This gives considerable credence to the use of these monologic discourse in experimental settings, allowing us to bridge the gap between every day communication (e.g. conversation) and isolated language tasks (e.g. confrontation naming). In general, standardised norms for discourse-level language are lacking in the field of aphasiology (and in development Applebee, 1978; Peterson & McCabe, 2013). But, the accumulation of evidence, including results from this study, suggests that we may be able to create

standardised norms for aspects of discourse. This will improve our understanding of discourse change (i.e. in development, after brain injury, over the course of neurodegenerative disease).

#### 4.4. Limitations and future directions

While we drew from a large database of speakers with and without aphasia (AphasiaBank), we did exclude individuals who produced fewer than 20 tokens during a task. Doing so eliminated some of the speakers with more severe aphasia (e.g. very severe Broca's aphasia, global aphasia, etc), but also a few speakers who were simply not verbose. Therefore, some comparisons had small group sizes (e.g. sixteen speakers with Broca's aphasia in the Sandwich task). Further, few of the speakers were bi- or multilingual and most speakers had relatively similar levels of education (ranging from 12-20 years, typically). Therefore, our sample may not generalise to populations that are traditionally bi- or multilingual, or to samples with lower socioeconomic status and education. A priority should be on replication of our findings in diverse samples.

Our analysis relied on semi-automatic morphological tagging done by the CLAN program (MacWhinney, 2000). While the morphological tagging has been validated in AphasiaBank prior research, it may still be the case that there is incorrect word-level tagging due to inherent ambiguity in linguistic information when few tokens were produced. We would argue that this would also be the case if done by the "gold standard" (e.g. human raters). Replication of these results in a new sample will speak to the quality and reproducibility of task x microstructure interaction.

Finally, this analysis was restricted to only a single level of discourse (microstructure, or the language used in the discourse), and did not evaluate further levels, such as macrostructure/functional information (Frederiksen et al., 1990; Sherratt, 2007). We intend to conduct future work linking together micro- and macrostructural levels, as has been suggested by others Dipper and Pritchard (2017), Marini et al. (2011) and Sherratt (2007).

#### Disclosure statement

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

#### References

Applebee, A. N. (1978). *The child's concept of story: Ages two to seventeen*. University Of Chicago Press.

- Becker, J. T., Lopez, O. L., Saxton, J., & Mcgonigle, K. L. (1994). The natural history of Alzheimer's disease description of study cohort and accuracy of diagnosis. *Archives of Neurology*, 51(6), 585–594. <https://doi.org/10.1001/archneur.1994.00540180063015>
- Berman, R. A. (2009). Trends in research on narrative development. In S. Foster-Cohen (Ed.), *Language acquisition* (pp. 294–318). Palgrave Macmillan.
- Bliss, L. S., & McCabe, A. (2006). Comparison of discourse genres: Clinical implications. *Contemporary Issues in Communication Sciences and Disorders*, 33, 126–167.
- Boles, L. (1998). Conversational discourse analysis as a method for evaluating progress in aphasia: A case report. *Journal of Communication Disorders*, 31(3), 261–274. [https://doi.org/10.1016/S0021-9924\(98\)00005-7](https://doi.org/10.1016/S0021-9924(98)00005-7)
- Boyle, M. (2011). Discourse treatment for word retrieval impairment in aphasia: The story so far. *Aphasiology*, 25(11), 1308–1326. <https://doi.org/10.1080/02687038.2011.596185>
- Boyle, M. (2014). Test-retest stability of word retrieval in aphasic discourse. *Journal of Speech, Language, and Hearing Research*, 57(3), 966–978. [https://doi.org/10.1044/2014\\_JSLHR-L-13-0171](https://doi.org/10.1044/2014_JSLHR-L-13-0171)
- Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews* (6), CD000425. <https://doi.org/10.1002/14651858.CD000425.pub4>
- Brenneise-Sarshad, R., Nicholas, L. E., & Brookshire, R. H. (1991). Effects of apparent listener knowledge and picture stimuli on aphasic and non-brain-damaged speakers' narrative discourse. *Journal of Speech, Language, and Hearing Research*, 34(1), 168–176. <https://doi.org/10.1044/jshr.3401.168>
- Brookshire, R. H., & Nicholas, L. E. (1994). Speech sample-size and test-retest stability of connected speech measures for adults with aphasia. *Journal of Speech, Language, and Hearing Research*, 37(2), 399–407. <https://doi.org/10.1044/jshr.3702.399>
- 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>
- Bryant, L., Spencer, E., & Ferguson, A. (2017). Clinical use of linguistic discourse analysis for the assessment of language in aphasia. *Aphasiology*, 31(10), 1105–1126. <https://doi.org/10.1080/02687038.2016.1239013>
- Cameron, R. M., Wambaugh, J. L., & Mauszycki, S. C. (2010). Individual variability on discourse measures over repeated sampling times in persons with aphasia. *Aphasiology*, 24(6–8), 671–684. <https://doi.org/10.1080/02687030903443813>
- Correia, L., Brookshire, R. H., & Nicholas, L. E. (1989). The effects of picture content on descriptions by aphasic and non-brain damaged speakers. *Clinical Aphasiology*, 18, 447–462.
- Correia, L., Brookshire, R. H., & Nicholas, L. E. (1990). Aphasic and non-brain-damaged adults' descriptions of aphasia test pictures and gender-biased pictures. *Journal of Speech and Hearing Disorders*, 55(4), 713–720. <https://doi.org/10.1044/jshd.5504.713>
- Cruise, M., Botting, N., Marshall, J., Boyle, M., Hersh, D., Pritchard, M., & Dipper, L. (2020). UK speech and language therapists' views and reported practices of discourse analysis in aphasia rehabilitation. *International Journal of Language and Communication Disorders*, 55(3), 417–442. <https://doi.org/10.1111/jlcd.v55.3>

- Dalton, S. G. H., Kim, H., Richardson, J. D., & Wright, H. H. (2020). A compendium of core lexicon checklists. *Seminars in Speech and Language*, 41(1), 45–60. <https://doi.org/10.1055/s-0039-3400972>
- Dalton, S. G., & Richardson, J. D. (2015). Core-Lexicon and main-concept production during picture-sequence description in adults without brain damage and adults with aphasia. *American Journal of Speech-Language Pathology*, 24(4), S923–S938. [https://doi.org/10.1044/2015\\_AJSLP-14-0161](https://doi.org/10.1044/2015_AJSLP-14-0161)
- Dalton, S. G. H., & Richardson, J. D. (2019). A large-scale comparison of main concept production between persons with aphasia and persons without brain injury. *American Journal of Speech-Language Pathology*, 28, 15293–320. [https://doi.org/10.1044/2018\\_AJSLP-17-0166](https://doi.org/10.1044/2018_AJSLP-17-0166)
- Davis, G. A. (2007). *Aphasiology: Disorders and clinical practice* (2nd ed.). Pearson.
- Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, 27(2), 124–142. [https://doi.org/10.1016/0749-596X\(88\)90070-8](https://doi.org/10.1016/0749-596X(88)90070-8)
- Dipper, L. T., & Pritchard, M. (2017). Discourse: Assessment and therapy. In *Advances in speech-language pathology*. InTech. <https://doi.org/10.5772/intechopen.69894>
- Doyle, 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(7–8), 561–574. <https://doi.org/10.1080/02687039808249558>
- Easterbrook, A., Brown, B. B., & Perera, K. (1982). A comparison of the speech of adult aphasic subjects in spontaneous and structured interactions. *International Journal of Language & Communication Disorders*, 17(3), 93–107. <https://doi.org/10.3109/13682828209012223>
- Faroqi-Shah, Y., Treanor, A., Ratner, N. B., Ficek, B., Webster, K., & Tsapkini, K. (2020). Using narratives in differential diagnosis of neurodegenerative syndromes. *Journal of Communication Disorders*, 85, 105994. <https://doi.org/10.1016/j.jcomdis.2020.105994>
- 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/10.1080/02687038.2011.606974>
- Fraser, K. C., Meltzer, J. A., Rudzicz, F., & Garrard, P. (2015). Linguistic features identify Alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease*, 49(2), 407–422. <https://doi.org/10.3233/JAD-150520>
- Fraser, K., Rudzicz, F., Graham, N., & Rochon, E. (2013). Automatic speech recognition in the diagnosis of primary progressive aphasia. In *Proceedings of the Fourth Workshop on Speech and Language Processing for Assistive Technologies, 4th Workshop on Speech and Language Processing for Assistive Technologies* (pp. 47–54).
- Frederiksen, C. H., Bracewell, R. J., Breuleux, A., & Renaud, A. (1990). The cognitive representation and processing of discourse: Function and dysfunction. In *Discourse ability and brain damage* (pp. 69–110). Springer. [https://doi.org/10.1007/978-1-4612-3262-9\\_4](https://doi.org/10.1007/978-1-4612-3262-9_4)
- Fromm, D., Forbes, M., Holland, A., & Dalton, G. (2017). Discourse characteristics in aphasia beyond the Western Aphasia Battery cutoff. *American Journal of Speech-Language Pathology*, 26(3), 762–768. [https://doi.org/10.1044/2016\\_AJSLP-16-0071](https://doi.org/10.1044/2016_AJSLP-16-0071)
- Fromm, D., Greenhouse, J., Hou, K., Russell, G. A., Cai, X., Forbes, M., Holland, A., & 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\\_JSLHR-L-15-0401](https://doi.org/10.1044/2016_JSLHR-L-15-0401)
- Galetto, V., Kintz, S., West, T., Marini, A., Wright, H. H., & Fergadiotis, G. (2013). Measuring global coherence in aphasia. *Procedia – Social and Behavioral Sciences*, 94, 198–199. <https://doi.org/10.1016/j.sbspro.2013.09.097>
- Glosser, G., & Deser, T. (1992). A comparison of changes in macrolinguistic and microlinguistic aspects of discourse production in normal aging. *Journal of Gerontology*, 47(4), P266–P272. <https://doi.org/10.1093/geronj/47.4.P266>
- Glosser, G., Wiener, M., & Kaplan, E. (1988). Variations in aphasic language behaviors. *Journal of Speech and Hearing Disorders*, 53(2), 115–124. <https://doi.org/10.1044/jshd.5302.115>
- Goodglass, H., & Kaplan, E. (1972). *Boston diagnostic aphasia examination*. Lea & Febiger.
- Goodglass, H., Kaplan, E., & Barresi, B. (2000). *Boston diagnostic aphasia examination* (3rd ed.). Pearson.
- Husain, M. (2017). Transdiagnostic neurology: Neuropsychiatric symptoms in neurodegenerative diseases. *Brain*, 140(6), 1535–1536. <https://doi.org/10.1093/brain/awx115>
- Kertesz, A. (2007). *Western aphasia battery - revised*. Psychological Corporation.
- Le, D., Licata, K., & Provost, E. M. (2018). Automatic quantitative analysis of spontaneous aphasic speech. *Speech Communication*, 100, 1–12. <https://doi.org/10.1016/j.specom.2018.04.001>
- Le, D., & Provost, E. M. (2016, September 8–12). Improving automatic recognition of aphasic speech with AphasiaBank. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (pp. 2681–2685). <https://doi.org/10.21437/Interspeech.2016-213>
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.
- Linnik, A., Bastiaanse, R., & Höhle, B. (2016). Discourse production in aphasia: A current review of theoretical and methodological challenges. *Aphasiology* 30(7). <https://doi.org/10.1080/02687038.2015.1113489>
- MacWhinney, B. (2000). *The CHILDES project: Tools for analysing talk* (3rd ed.). Lawrence Erlbaum.
- MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, 25(11), 1286–1307. <https://doi.org/10.1080/02687038.2011.589893>
- MacWhinney, B., Fromm, D., Holland, A., Forbes, M., & Wright, H. (2010). Automated analysis of the Cinderella story. *Aphasiology*, 24(6–8), 856–868. <https://doi.org/10.1080/02687030903452632>
- Marini, A., Andreetta, S., del Tin, S., & Carlomagno, S. (2011). A multi-level approach to the analysis of narrative language in aphasia. *Aphasiology*, 25(11), 1372–1392. <https://doi.org/10.1080/02687038.2011.584690>
- Nadeau, S. E. (2001). Phonology: A review and proposals from a connectionist perspective. *Brain and Language*, 79(3), 511–579. <https://doi.org/10.1006/brln.2001.2566>

- Nicholas, L. E., & Brookshire, R. H. (1995). Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. *Journal of Speech, Language, and Hearing Research*, 38(1), 145–156. <https://doi.org/10.1044/jshr.3801.145>
- Olness, G. S. (2006). Genre, verb, and coherence in picture-elicited discourse of adults with aphasia. *Aphasiology*, 20(2–4), 175–187. <https://doi.org/10.1080/02687030500472710>
- Peterson, C., & McCabe, A. (2013). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. Springer.
- Richardson, J. D., & Dalton, S. G. (2016). Main concepts for three different discourse tasks in a large non-clinical sample. *Aphasiology*, 30(1), 45–73. <https://doi.org/10.1080/02687038.2015.1057891>
- Roberts, J. A., & Wertz, R. T. (1989). Comparison of spontaneous and elicited oral-expressive language in aphasia. *Clinical Aphasiology*, 18, 479–488.
- Rose, M. L., Mok, Z., Carragher, M., Katthagen, S., & Attard, M. (2015). Comparing multi-modality and constraint-induced treatment for aphasia: A preliminary investigation of generalisation to discourse. *Aphasiology*, 30(6), 678–698. <https://doi.org/10.1080/02687038.2015.1100706>
- Sahraoui, H., & Nespoulous, J.-L. (2012). Across-task variability in agrammatic performance. *Aphasiology*, 26(6), 785–810. <https://doi.org/10.1080/02687038.2011.650625>
- Sekine, K., & Rose, M. L. (2013). The relationship of aphasia type and gesture production in people with aphasia. *American Journal of Speech-Language Pathology*, 22(4), 662–672. [https://doi.org/10.1044/1058-0360\(2013\)12-0030](https://doi.org/10.1044/1058-0360(2013)12-0030)
- Sherratt, S. (2007). Multi-level discourse analysis: A feasible approach. *Aphasiology*, 21(3–4), 375–393. <https://doi.org/10.1080/02687030600911435>
- Stark, B. C. (2019). A comparison of three discourse elicitation methods in aphasia and age-matched adults: Implications for language assessment and outcome. *American Journal of Speech-Language Pathology*, 28(3), 1067–1083. [https://doi.org/10.1044/2019\\_AJSLP-18-0265](https://doi.org/10.1044/2019_AJSLP-18-0265)
- Stark, B. C., Dutta, M., Murray, L. L., Bryant, L., Fromm, D., MacWhinney, B., Ramage, A. E., Roberts, A., den Ouden, D. B., Brock, K., McKinney-Bock, K., Jin Paek, E., Harmon, T. G., Yoon, S. O., Themistocleous, C., Yoo, H., Aveni, K., Gutierrez, S., & Sharma, S. (2020). Standardizing assessment of spoken discourse in aphasia: A working group with deliverables. *American Journal of Speech-Language Pathology*. June 2020, 25. [https://doi.org/10.1044/2020\\_AJSLP-19-00093](https://doi.org/10.1044/2020_AJSLP-19-00093)
- Ulatowska, H. K., North, A. J., & Macaluso-Haynes, S. (1981). Production of narrative and procedural discourse in aphasia. *Brain and Language*, 13(2), 345–371. [https://doi.org/10.1016/0093-934X\(81\)90100-0](https://doi.org/10.1016/0093-934X(81)90100-0)
- Van Leer, E., & Turkstra, L. (1999). The effect of elicitation task on discourse coherence and cohesion in adolescents with brain injury. *Journal of Communication Disorders*, 32(5), 327–349. [https://doi.org/10.1016/S0021-9924\(99\)00008-8](https://doi.org/10.1016/S0021-9924(99)00008-8)
- Wright, H. H., & Capilouto, G. J. (2009). Manipulating task instructions to change narrative discourse performance. *Aphasiology*, 23(10), 1295–1308. <https://doi.org/10.1080/02687030902826844>