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

Brielle C. Stark, PhD\*<sup>†‡</sup> Julia Fukuyama, PhD<sup>§</sup>

November 18, 2020

BCS, JF have equal author contribution

Corresponding author: BC Stark, bcstark@iu.edu

Address: 200 S Jordan Avenue, Department of Speech and Hearing Sciences, Indiana University Bloomington 47405 USA

The authors have no competing interests to disclose.

The authors have no funding to disclose.

<sup>\*</sup>Indiana University Bloomington, Department of Speech and Hearing Sciences

<sup>&</sup>lt;sup>†</sup>Indiana University Bloomington, Program in Neuroscience

<sup>&</sup>lt;sup>‡</sup>Indiana University Bloomington, Cognitive Science Program

<sup>&</sup>lt;sup>§</sup>Indiana University Bloomington, Department of Statistics

### 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 standardized tests. Here, we leverage big behavioral data (AphasiaBank) to understand how discourse genres (narrative, procedural, expositional), and unique tasks within those genres, influence microstructural elements of discourse (specifically, linguistic forms including part of speech, lexical type [open, closed] and morphological tense). We compare task x 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 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.

### **Keywords**

Aphasia, "connected speech", discourse, language, stroke

### 1 Introduction

When we produce spoken language, we do not simply name objects or produce single words 1 [Levelt, 1989, Dell, 1988]. Instead, we produce meaningfully organized output consisting of 2 many words, constantly suppressing alternatives at multiple linguistic levels (e.g., semantic, 3 lexical, phonological, syntactic) in discourse. Discourse breaks down in a variety of disorders, 4 like aphasia—a language disorder which most commonly occurs after acquired brain damage 5 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., 7 access to semantic, lexical, phonological and morphosyntactic information), which is the pri-8 mary impairment in aphasia. As such, discourse elicitation in aphasia is thought to be an 9 ecologically valid and comprehensive means to understand the language system's affected 10 and unaffected processes [Bryant et al., 2016], lending information about complex cognitive-11 12 linguistic behaviors and cognitive-communicative deficits that are not otherwise measured by standardized assessments. It follows that discourse is commonly employed in clinical and 13 research settings to assess language and language-related cognitive processes (e.g., organiza-14 tion) in persons with aphasia [Bryant et al., 2016, Linnik et al., 2016], and is used as a primary 15 outcome of speech-language intervention [Brady et al., 2016]. 16

In linguistics, the term *discourse* is typically defined as a language unit longer than a 17 single sentence, but in clinical contexts, is more narrowly defined as language beyond a single 18 simple clause, which is used for a specific purpose or function [Dipper & Pritchard, 2017]. 19 Discourse can be either monologic or dialogic, dependent on the number of interlocutors. 20 Discourse analysis comprises three components (as described by [Sherratt, 2007, Frederiksen 21 et al., 1990]): (1) how language is used (lexical and grammatical resources), typically termed 22 its 'microstructural' level, (2) what information is included, and (3) how the information is 23 structured. Levels two and three typically are thought of as 'macrostructural' or 'functional' 24 levels, i.e., pertaining to information beyond linguistic structure. For the purposes of a concise 25 introduction, we will focus on the microstructural level, as this is the crux of the analysis we 26 provide in our results section, but we will make note of evidence from macrostructural levels 27 28 where possible. It should also be noted that the evidence we provide, unless noted otherwise, 29 is from English speakers.

Here, we will restrict discussion to monologic discourse. There are several elicitation 30 genres of monologic discourse [Bryant et al., 2016]: (1) Expositional discourse (or expository 31 discourse; sometimes called 'descriptive' discourse) is elicited by providing subjects with a 32 set of structured instructions and, often, a picture or picture sequence to describe (e.g., "tell 33 me what's happening in this picture"); (2) Narrative discourse is elicited by having the subject 34 recount a personal story ('personal narrative') or retell a well-known story ('story retelling'). 35 Typically, narrative discourse does not employ visual aids, and the instructions are semi-36 structured (e.g., "tell me what you know about..."); and (3) Procedural discourse is elicited by 37 having subjects describe a set of actions or a procedure (e.g., "tell me how you might make 38 a ..."), typically without pictorial cues. Within each elicitation genre, a variety of tasks exist. 39 For example, within expositional discourse, the most commonly used tasks are single picture 40 descriptions and picture sequence descriptions [Bryant et al., 2016]. Within the narrative 41 context, there are autobiographical tasks, as well as tasks that rely more on semantic memory, 42 such as fictional story retell ("Cinderella"). 43

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, Fraser et al., 2015, Faroqi-Shah et al., 2020]. 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 con-50 straint [Bryant et al., 2017, Cruice et al., 2020]. This is not surprising, given that it is estimated 51 to take a trained transcriber anywhere from 6-12 minutes to transcribe a single minute of a 52 discourse sample from a person with aphasia [Boles, 1998] and additional time to code and 53 analyze that sample. Therefore, time is a non-negligible restraint, which may directly impair 54 comprehensive assessment of discourse. Despite this, there is mounting evidence that lan-55 56 guage sampling using multiple, varied discourse elicitation methods will provide the most comprehensive language assessment. 57

Firstly, there is evidence that language must be sampled in sufficient quantity to obtain 58 reliable estimates. For example, test-retest reliability of correct information units in speak-59 ers with aphasia, a commonly employed metric of language efficiency and productivity, is 60 low unless samples contain at least 300-400 words [Brookshire & Nicholas, 1994]. Notably, 61 62 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 AphasiaBank [MacWhin-63 ney et al., 2011] demonstrates that, during a single picture description task (e.g., Cat Rescue), 64 persons with aphasia produce on average 70 tokens (lemmas) [Stark, 2019]. Indeed, even 65 speakers without aphasia (non-brain-damaged control group matched on age and education) 66 produced, on average, only a few more tokens (~ 90) on this task [Stark, 2019]. A cursory 67 examination of the Pitt Corpus from DementiaBank [Becker et al., 1994] lends a similar result, 68 with the average tokens produced on the Cookie Theft picture [Goodglass & Kaplan, 1972] by 69 persons with probable Alzheimer's Disease at their first assessment being 81 (N=196) (this 70 analysis run for example purposes by the authors). This is also the case for RHDBank [MacWhin-71 ney et al., 2011], a database of persons with right hemisphere stroke, when retelling the Cat 72 Rescue story. Indeed, they produced on average 186 tokens (N=16) (this analysis run for ex-73 *ample purposes by the authors*). Therefore, despite a single picture description being the most 74 prevalent means of acquiring a discourse sample in clinical practice, it does not align with 75 the evidence on sampling reliability. There is a need to acquire multiple discourse samples, 76 77 to maximize 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 78 level [Boyle, 2014, Cameron et al., 2010], emphasizing that more work must be done to ascer-79 tain which discourse-extracted outcome measures are most stable and reliable. 80

Second, there is considerable evidence that elicitation method interacts with language 81 produced in the discourse; that is, simply put, the task constrains the language produced. 82 83 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 84 reflects the different genres of discourse elicited (e.g., narrative, procedure) [Sherratt, 2007, 85 Frederiksen et al., 1990]. That is, each elicitation method requires a unique frame structure. 86 It follows that each task, and each genre, will have a considerable impact on micro- and 87 macrostructural levels. 88

Wright and Capilouto (2009) found that, in 24 healthy older adults, task instructions influ-89 enced linguistic information: if subjects were told to "talk about what is going on in the 90 picture" ('describe'), subjects tended to produce fewer past-tense verbs than if they were 91 told to "look at the picture(s) and tell a story that has a beginning, middle and end" ('sto-92 rytelling') [Wright & Capilouto, 2009]. In a later study, Fergadiotis, Wright and Capilouto 93 (2011) explored differences in lexical diversity (a measure of the proportion of lexical items 94 [i.e., nouns, verbs, adjectives and some adverbs]) across discourse genres (expositional, nar-95 96 rative 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, 97 in picture sequences) [Fergadiotis & Wright, 2011], demonstrating genre-specific microstruc-98 tural information. In a larger sample size (N=86) of persons without aphasia, lexical diver-99 sity, a microlinguistic measure, was demonstrated to be influenced by discourse type and 100

age [Fergadiotis et al., 2011]. Not surprisingly, elicitation methods have also been found to 101 influence micro- and macrostructural processes of language in speakers with aphasia (a non-102 exhaustive list: [Sahraoui & Nespoulous, 2012, Brookshire & Nicholas, 1994, Easterbrook et al., 103 1982, Glosser et al., 1988, Correia et al., 1989, Roberts & Wertz, 1989, Brenneise-Sarshad et al., 104 1991, Correia et al., 1990, Olness, 2006, Stark, 2019, Doyle et al., 1998]). For example, people with 105 aphasia were found to produce more present tense verbs during a picture description (expo-106 107 sitional discourse) and more past tense verbs during a story retelling (narrative discourse) (N=12) [Olness, 2006]. Reduced syntactic complexity during procedural discourse, compared 108 with narrative discourse, has also been shown in aphasia [Ulatowska et al., 1981]. A study 109 in French-speaking persons with agrammatic speech and aphasia likewise revealed across-110 task variability in morphosyntax during spoken language (contrasting different discourse 111 elicitation methods as well as constrained sentence production), suggesting that participants 112 113 tended to adjust the morphosyntactic aspects of their speech according to task-dependent factors [Sahraoui & Nespoulous, 2012]. 114

While some of these early studies in aphasia included small sample sizes and did not 115 directly compare performance of speakers with aphasia to speakers without aphasia, there 116 has been considerable methodological improvement, with studies contrasting language in 117 118 speakers with and without aphasia across a variety of discourse genres (a non-exhaustive list including [Stark, 2019, Dalton & Richardson, 2019, Richardson & Dalton, 2016]), replicating 119 early findings. Indeed, recent work by our group evaluated a handful of microstructural 120 variables (e.g., mean length of utterance, propositional density) in a large group of speakers 121 with aphasia (N=90) and without aphasia (N=84; age and education matched) across three 122 discourse genres (narrative, expositional, procedural) [Stark, 2019]. Despite the speakers with 123 aphasia producing nearly 50% fewer words than the control group [Stark, 2019], there were 124 genre- and task-specific microstructural patterns evidenced in both the aphasia and control 125 groups. For example, all speakers produced the least syntactically complex language during 126 the procedural narrative, while all speakers produced the densest language during the story 127 128 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 129 speakers with and without aphasia, but note that genre- and task-specific effects have been 130 shown in this literature as well [Van Leer & Turkstra, 1999, Glosser & Deser, 1992, Dalton 131 & Richardson, 2019, Dalton & Richardson, 2015]. Additionally, we see task-specificity for 132 micro- and macrostructural components of discourse in children as young as 5 years old 133 [Berman, 2009], further emphasizing that the frame generation stage molds the later stages 134 of discourse production (i.e., insertion and integration of semantic information, selection and 135 topicalization of information, generation of propositions and linguistic formulation) [Sherratt, 136 2007, Frederiksen et al., 1990] from an early age. 137

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 prioritize an elicitation 144 method that best elicits discourse with particular properties of interest. Evidence from a va-145 riety of sources have demonstrated that understanding task-language interactions can inform 146 147 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 148 result of intervention, it may be best to sample language using a narrative elicitation method 149 rather than a procedural elicitation method [Stark, 2019, Ulatowska et al., 1981], as narra-150 tives typically associate with denser and more syntactically complex language. Pritchard and 151

Dipper offer another hypothetical situation to the same effect: that if a client has difficulty 152 with cohesion (a macrostructural variable), the clinician should select a discourse elicitation 153 method that is likely to use cohesion, such as a narrative discourse with multiple characters, 154 in order to test this hypothesis [Dipper & Pritchard, 2017]. While understanding task-specific 155 interactions with language is ideal, doing so does not necessarily argue against collection of 156 multiple samples (indeed, in the case of both examples, it would argue for the collection of 157 158 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 159 a single sample might. To our knowledge, there has been limited research evaluating this 160 conjecture. To introduce this concept we will provide some theoretical examples, below. 161

162

163 164

165

166

167

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 168 169 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 pref-170 erentially produce content words with limited use of functor words. But, it remains unclear 171 the extent to which the ratio of content-to-functor words, for example, is influenced by elic-172 itation method. If the primary purpose for evaluating discourse in a person with Broca's 173 aphasia is to examine content-to-functor word ratio, and the content-to-functor word ratio is 174 not a variable that is heavily influenced by elicitation method, this may be an opportunity to 175 use clinical linguistic knowledge to pick the most sensitive elicitation method (e.g., to measure 176 propositional density, choosing a narrative task rather than a procedural task, [Fromm et al., 177 2016]) and collect multiple samples using that same elicitation method, rather than employing 178 179 varied methods.

Another illustrative example that is especially pertinent to the analyses we present in this 180 project, evaluates the impact of elicitation method on language as a result of aphasia factors 181 (type and overall aphasia severity). Wernicke's aphasia, a more severe fluent-type of aphasia, 182 is characterized by production of indefinite words, paragrammatic speech, and paraphasias 183 (word errors). A speaker with Wernicke's aphasia might produce many semantically-related, 184 phonemically-related, neologistic, and/or unrelated paraphasias, suggesting impaired, in-185 consistent access to their lexical-semantic-phonological system. Characteristic of Wernicke's 186 aphasia, there are few self-corrections and revisions. Conduction aphasia, typically presenting 187 as a less severe fluent-type of aphasia, is likewise characterized by production of indefinite 188 words, paraphasias, and some paragrammatic speech, but also by *conduite d'approche*, or the 189 successive approximations at producing target words. Nadeau (2001) and others suggest that 190 *conduite d'approche* reflects self-correcting mechanism underpinned by intact lexical-semantic 191 information [Nadeau, 2001]. From this information, we may make the assumption that com-192 munication between the lexical-semantic-phonological system is less impaired in a speaker 193 with Conduction aphasia than a speaker with Wernicke's aphasia. Therefore, microstructural 194 information may not differ by discourse elicitation method in persons with Wernicke's apha-195 sia because these speakers are, in general, unable to reliably access linguistic information, 196 meaning that they are liable to produce unreliable linguistic information regardless of the 197 198 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 199 and memory), which we discussed earlier. An inability to frame resultant language (and 200 macrostructural) information may result in discourse appearing very similar (at micro- and 201 macrostructural levels) across all elicitation methods. In a similar vein, speakers with severe 202

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 208 209 speakers produce context-relevant lexical items during tasks) supports the supposition that aphasia type and overall aphasia severity may interact with task [Dalton & Richardson, 2019]. 210 During most tasks (a single picture description, a picture sequence description, a procedural 211 narrative, and a fictional narrative), persons with Wernicke's aphasia (N=19) and Broca's 212 aphasia (N=61) tended to produce the same number of main concept units per task. On the 213 other hand, persons with Anomic aphasia (N=86) produced varying numbers of main concept 214 215 units dependent on the task (ranging from 8 - 25 units), and a similar pattern was noted for the non-brain-damaged control group (N=145) (ranging from 17-55 units) [Dalton & Richardson, 216 2018]. What we cannot glean from these quantitative results is whether the lexical content 217 was similar across tasks (i.e., were they accessing similar numbers of verbs and nouns?), but 218 we can surmise that a similar amount of macrostructural information was being conveyed 219 by persons with Broca's and Wernicke's aphasia per task, whereas a task-dependent amount 220 of macrostructural information was being conveyed by persons with Anomic aphasia and 221 persons without brain damage. Research at the microstructural level, evaluating propositional 222 density, has also shown distinct interactions of discourse task and aphasia type [Fromm et al., 223 2016]. 224

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 dis-239 course task is thought to draw upon unique cognitive components; for example, fictional 240 story retelling likely draws more on semantic memory whereas narrative story retelling (au-241 tobiographical) draws more on episodic memory (as emphasized in development by [Bliss & 242 Mccabe, 2006]). A distinction in microstructure produced by each task may further empha-243 size the unique cognitive contributions required for the task. Another important contribution 244 to theoretical understanding of discourse is the extent to which discourse microstructure is 245 consistent within similar groups of people (e.g., those without brain damage). If there is intra-246 group homogeneity in microstructure for each task, this suggests that the task being employed 24 taps into shared cognitive and language resources across people, thus supporting the validity 248 249 for acquiring monologic discourse to examine more naturalistic language processes.

250

Therefore, this study has both clinical and theoretical implications.

### **251 2** Materials and Methods

### 2.1 Database

252

264

This project draws from a large database of behavioral 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, analyzing each speaker's first administration of the discourse protocol.

### 262 2.2 Subjects

### 263 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 265 used to qualify presence, overall aphasia severity, and type of aphasia. To create an over-266 all aphasia severity metric (the Aphasia Quotient [AQ] of the WAB), four sub-component, 267 composite scores are acquired: auditory verbal comprehension, naming, spontaneous speech 268 and repetition. Auditory verbal comprehension composite score contains scores from yes/no 269 questions, auditory word recognition, and sequential commands subtests. Naming composite 270 score contains scores from object naming, word fluency, sentence completion and responsive 271 speech subtests. Spontaneous speech composite score contains scores about the informational 272 content and fluency of a picture description subtest and a short conversational section. Finally, 273 repetition composite score contains scores from a single subtest comprising single-word and 274 sentence-level stimuli. These four sections allow us to derive the AQ, which is a score from o 275 to 100. According to the WAB manual, an AQ < 93.8 indicates presence of aphasia. The WAB 276 also provides guidelines for assigning overall aphasia severity based on the AQ score. An 277 278 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 279 not necessarily reflective of discourse ability – it is a composite, standardized score meant to 280 reflect overall aphasia severity having taken into account repetition, auditory comprehension, 281 282 naming, and spontaneous speech information.

Additionally, composite scores of auditory verbal comprehension, naming, and repeti-283 284 tion, 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, 285 Broca's, Conduction, Wernicke's, Transcortical Sensory, Transcortical Motor, Isolation (Mixed 286 Transcortical) and Anomic. Additionally, people having an AQ of >93.8 are considered to 287 be 'not aphasic by WAB' (in some cases, called 'latent aphasia') but, given the presence of a 288 chronic brain injury, often present with residual language problems (i.e., most do not score a 289 perfect 100 on the assessment). Indeed, all persons in the AphasiaBank database, regardless of 290 WAB score, were identified as having aphasia by a clinician and/or seeking aphasia services 291 in a university clinic or aphasia center. For the purposes of this study, we did not include data 292 from people with Transcortical motor aphasia, Transcortical sensory aphasia, Mixed transcor-293 294 tical 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. 295 We did include individuals who were not considered aphasic by WAB standards (AQ >93.8) 296 but also who did not score a perfect 100 on the exam; we consider this our 'Not Aphasic 297

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, standardized test of aphasia.

Below is a brief description of language production characteristics associated with aphasia types included in the current study [Goodglass & Kaplan, 1972, Goodglass et al., 2000, Davis, 2007]:

- *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 characterized 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 characterized 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 characterized 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 characterized 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, standardized batteries, but persons with this type of aphasia may self-identify as having aphasia, or be identified as having aphasia by clinicians.

327 328 329

330

331

332

333

303 304

305 306

307

308

309 310

311

312

313

314

315

316

317

318

319

320

321

322

323

324 325

326

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.

334 2.2.2 All Speaker Characteristics

Because we were interested in modeling microstructure (specifically, lexical variables) and 335 needed sufficient data to do so, we excluded all speakers (including controls) who produced 336 fewer than 20 tokens (lemmas) per task. The monologic discourse elicitation protocol from 337 AphasiaBank was not completed in its entirety for every participant, and therefore there was a 338 different number of speakers (control and aphasia) for each task. Therefore, it is best to refer to 339 Table 1, which describes, in detail, demographic information for included participants. Table 340 1 shows the number of speakers that had data for each discourse task, separated by group 341 342 and aphasia type; years of education; relevant standardized testing scores; age at testing; and sex. This table also provides an overview for each discourse task by showing total number of 343 tokens, mean length of utterance (in words) and total utterances. 344

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

Data	Cat	Cinderella	Important Event	Sandwich	Window
Total n	224	237	74	215	238
Education (years)	15.2 (2.3)	15.2 (2.3)	15.3 (2.8)	15.2 (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)
Age at test (years)	56.1 (23.1)	55.5 (22.9)	76.1 (7)	55.5 (23.3)	55.6 (22.8)
Females	o.58	0.59	0.42	0.57	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)
Tokens	109 (46.7)	447.2 (265.8)	338 (327.7)	95 (57.1)	85 (46.2)
Utterances	10.9 (5.6)	46.2 (31.8)	37.9 (42.7)	10.8 (5.9)	8.9 (6.6)

# Aphasia participants

Data	Cat	Cinderella	Important Event	Sandwich	Stroke	Window
Total n	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 <sup>r</sup> 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)

Table 1: Demographics and basic task-related language information for controls (top) and aphasia group (bottom), broken down by task. Wherever possible, mean and standard deviaition (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 analyzed 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).

### **2.3.1** Expositional 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 seconds, 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?"

### 363 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 seconds, 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?"

<sup>370</sup> See Figure 1 for Window and Cat picture stimuli.

### **2.3.2** Narrative Genre

369

### 372 **Cinderella** - *Fictional Story Retell*

Present picture book. "I'm going to ask you to tell a story. Have you ever heard the story of 373 Cinderella? Do you remember much about it? These pictures might remind you of how it 374 goes. Take a look at the pictures and then I'll put the book away, and ask you to tell me 375 the story in your own words." Allow participant to look through book (assist with page turning if 376 needed). Remove book. If necessary, prompt: "Now tell me as much of the story of Cinderella as 377 you can. You can use any details you know about the story, as well as the pictures you just 378 looked at." If participant gives a response of fewer than three utterances, or seems to falter, allow 10 379 seconds, then prompt: "What happened next?" or "Go on." Continue until participant concludes 380 story or it is clear s/he has finished. 381

### <sup>382</sup> 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 seconds, prompt*: "For instance, you could tell me about a trip you took or something about your family or your work – anything."



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

### 387 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 seconds, prompt:* "Try to tell me about the day you had your stroke."

### <sup>392</sup> 2.3.3 Procedural Genre

388

### 393 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 seconds, 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 % ).

### <sup>402</sup> 2.4 Microstructural (linguistic) information derived from discourse

Language samples were transcribed in CHAT format at the collection site or by Aphasia-Bank 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 compos-406 ite measures of language (e.g., noun-verb ratio, mean length of utterance, lexical diversity), 407 analyzed independently of one another. Variable independence at the linguistic level is not 408 the rule (e.g., number of nouns produced tends to highly relate to number of verbs pro-409 duced; number of verbs produced is highly related to number of inflections produced). Here, 410 we wanted to model many (interrelated) linguistic variables using multivariate methodology 411 (described in 2.5). To do so, we derived the following variables from each speaker's dis-412 course samples: word class (i.e., open-class, closed-class), part of speech (i.e., noun, verb) and 413 morphosyntactic form (i.e., past tense, present participle). We first used CLAN's gem func-414 tion to separate each speaker's transcript by task (e.g., Cat, Cinderella) [MacWhinney, 2000]. 415 Then, morphological and grammatical dependency information was automatically tagged us-416 ing the mor command (overall, the accuracy of mor tagging for AphasiaBank transcripts is 417 418 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 419 command was used to extract selected linguistic variables (specifically, we only extracted in-420 formation from the persons with aphasia and not the experimenter/other interlocutor). An 421 example of a coded utterance from our sample, of describing the Window picture: 422

423 424 425

426

427

428

429

430

431

432 433

434

435

436

437

438

439 440

441

442

443

444

445

446

447

448

449

450

451 452

453

454

455

456

\*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;

Controls

Task	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)	
Aphasia						
Task	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)

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.

457	•	open class words;
458	•	closed class words;

459 460 The distribution of these variables by group (aphasia, control) and task is shown in Table 2.

### **2.5** Statistical Analysis

### 462 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 visualize the resulting model differ.

### 2.5.2 Between-class analysis to describe the relationship between task and lan guage microstructure

To analyze 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 visualize the relationships between the categories and the category means, we turn to 477 between-class analysis (BCA). In BCA, the category means are projected down from the initial 478 *p*-dimensional space to a lower-dimensional space in such a way that the distances between 479 the category means in the lower-dimensional spaces reflect as well as possible the original dis-480 tances between the categories in the high-dimensional space. We use a two-dimensional space 481 482 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 ap-483 proximation of the true distances between the category means, categories that have similar 484 language composition are plotted near to each other, while categories that have dissimilar 485 language composition are positioned farther away. 486

In addition to the representation of the category means in a lower-dimensional space, BCA 487 also provides information about the relationships between the variables (e.g., language data) 488 and the axes in the BCA plot. If a category has mean  $x = (x_1, \ldots, x_p)$ , then its position on 489 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 490 in the position of the category means on the BCA axes. For instance, if two category means 491 have similar values for all but variable i and  $v_{1i} > 0$ , the category with the larger value of 492 variable *i* will have a larger score on the first BCA axis. For variable *i*, we can make a point 493 whose position on the first BCA axis is  $v_{1i}$  and whose position on the second BCA axis is 494  $v_{2i}$ . This representation of the variables gives us some insight into the variables that are likely 495 responsible for differences between the categories. 496

In addition to BCA for visualization, we can summarize the relationships between cat-497 egories by computing the average distance between category means. In the *p*-dimensional 498 space, we can compute means within categories (for instance, task type, aphasia type, combi-499 nations of the two) and we can then use the relative distances between the category means as 500 a measure of the relative similarities of language produced by different groups and/or in re-50<sup>.</sup> sponse to different tasks. These distances can then be used in a hypothesis-testing framework 502 (e.g., PERMANOVA) to assess significance of the overall distances between the categories, or 503 we can create bootstrap confidence intervals for the between-category distances. 504

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

507 While BCA is appropriate for modeling and visualizing the relationship between language 508 microstructure and a categorical variable, it is not applicable to describing the relationship 509 between language microstructure and a continuous variable such as aphasia AQ. To do so, 510 we used a combination of multivariate regression of the linguistic variables on aphasia AQ 511 subscore and the other relevant variables and used PCA with respect to instrumental variables 512 (PCA-IV) to visualize 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

Variable	Mean	SD
SpontSp	15.01	3.66
AudVbl	8.44	1.48
Rep	6.85	2.47
Naming	7.18	2.35

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

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

521 522

523

524

525

527

528

529

530

531

533

534

535

536

537

That space and the *p* response variables, and
 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 visualize the fitted values from our model, as well as the relationship between the fitted values and the predictor variables.

### <sup>538</sup> 2.5.4 Quantifying within-class variance

The final set of analyses we performed have to do with the variance of microstructural (lin-539 guistic) information within each class. To quantify the overall spread of the samples within 540 each class in the multi-dimensional space, we used tr( $\hat{\Sigma}_i$ ), where  $\hat{\Sigma}$  is a  $p \times p$  giving the 541 estimated covariance within class *i*. In one dimension, this would simply be the estimated 542 within-class variance. In more than one dimension, as we have here, this quantity is an ag-543 gregate measure of variance across all of the variables. The other statistic we used to quantify 544 the within-class variance was the correlation between the scores on the first and second BCA 545 axes within each class. Again, if  $\hat{\Sigma}_i$  is the estimated covariance matrix within class *i* and  $v_1$ 546 and  $v_2$  are the loadings of the variables on the first and second PCA-IV axes, our statistic is 547  $v_1^T \hat{\Sigma}_i v_2 / \sqrt{v_1^T \hat{\Sigma}_i v_1 \cdot v_2^T \hat{\Sigma}_i v_2}$ . This measure is complementary to  $tr(\hat{\Sigma}_i)$  in that it tells us about 548 the shape of the distribution within class *i*, while  $tr(\hat{\Sigma}_i)$  tells us about the overall size of the 549 distribution within class *i*. 550

### 551 2.5.5 Data Considerations

Before applying any of the downstream methods, the data were transformed to remove the 552 effect of certain demographic variables and to bring the data more in line with the normality 553 assumptions implicit in several of the methods. A started log transformation was applied 554 to each language variable. Note that a log transformation would allow us to read off ratios 555 between variables as differences between the log-transformed variables (for instance, the log 556 ratio of plurals to nouns would be the difference between log plurals and log nouns), and the 557 558 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 559 number of words were regressed out of the transformed microstructure matrix. BCA was 560 then performed using the ade4 package in R. The confidence intervals provided are bootstrap 561 intervals, with the resampling done by subject to preserve the hierarchical structure of the 562 data. 563

### <sup>564</sup> 2.5.6 Data and Code Availability

AphasiaBank data is supported by NIH-NIDCD grant Ro1-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.

### <sup>569</sup> 3 Results and Discussion

### 570 Hypotheses

571

572

574

575

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 hypothesized, 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.
- The interaction of task with discourse microstructure would differ by aphasia type. We
  hypothesized that overall aphasia severity would play some role, in that the interaction of task and linguistic variables in 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.
- 5833. Finally, we anticipated that aphasia types associated with more severe language impair-<br/>ments (e.g., Broca's aphasia, Wernicke's aphasia) would not demonstrate as strong of a<br/>task x microstructure interaction, in that both within-genre tasks and across-genre tasks<br/>(e.g., Cinderella compared with Window) would demonstrate considerable overlap in<br/>linguistic variables. This hypothesis speaks to our theory that linguistic information is<br/>inherently limited in this population (i.e., severely impoverished language due to apha-<br/>sia), and may not change with task demands.

### 590 3.1 Group Comparison

For each task and group (aphasia, control), we statistically compared demographic and basic 591 language information using t-tests with Bonferroni correction. For years of education, there 592 was not a significant difference for subjects included in the Important Event task (p=.48), 593 Cinderella (p=.17) or Window (p=.11); there was a significant difference for Cat (p=.046) and 594 Sandwich (p=.01). For age (years) at testing, there was a significant difference for all tasks 595 (p<.05), as the aphasia group had a tendency to be older. For sex, there was not a signifi-596 cant difference for Important Event (p=.96) or Sandwich (p=.20) but there was a significant 597 598 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 apha-599 sia produced fewer total tokens. 600

- <sup>601</sup> For these reasons, we controlled for these demographics and language information in <sup>602</sup> 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 con-605 trols. The loadings of the linguistic variables in BCA space are shown in the sub-box in Fig 2, 606 where the two axes explain a total of 90.84% of the variance. This reduced lower-dimensional 607 space robustly reflects the linguistic variables of interest. Note that the y-axis tends to reflect 608 609 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 610 to the right and nouns, present participle, and determiners to the left. It should be noted 611 that CLAN MOR program will tag semi-auxiliary constructions (e.g., 'are going to make', 'are 612 going to need') as present participles (top left corner). One can directly map this sub-box to 613

the larger plot, showing, for example, that the Sandwich task loads left of center 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 616 Cinderella tasks cluster together in the BCA space (as demonstrated in Table 4, there is consid-617 erable overlap in their class mean confidence intervals), suggesting a similar linguistic profile. 618 Similar to what was observed with the two narrative tasks, the two expositional tasks cluster 619 620 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 621 intervals for the Sandwich task do not overlap with those of any other task, signifying a true 622 dissociation in linguistic microstructure from both narrative and expositional genres. 623

We next evaluated within-genre dissociation of tasks – that is, despite being a similar genre 624 (e.g., narrative), does each task have unique microstructure? While discourse microstructure 625 626 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 627 positions on each axis. On Axis II, the Cinderella task loads somewhat negatively (M=-0.20 [-628 0.27,-0.14])while the Important Event loads somewhat positively (M=0.36 [0.29,0.45]). On Axis 629 I, the Cinderella task loads somewhat negatively (M=-0.13 [-0.18,-0.07]) while the Important 630 Event loads near zero (M=0.03 [-0.06,0.12]). Unlike the Cinderella and Important Event tasks, 631 the Cat and Window tasks are not significantly differently positioned in either Axis I or Axis II. 632 Both tasks similarly (overlapping confidence intervals) load positively on Axis I (Cat, M=0.87 633 [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]; 634 Window, M=-0.45 [-0.54,-0.37]). The Sandwich task is positioned negatively along Axis I (M=-635 1.74 [-1.80, -1.67]) and slightly negatively along Axis II (M=-1.03 [-1.10,-0.97]). 636

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

	Important Event	Cinderella	Cat	Window
Cinderella	1.19 (0.98, 1.43)			
Cat	1.70 (1.59 <i>,</i> 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)

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

#### <sup>644</sup> 3.2.2 Evidence from the aphasia group

We were interested in modeling language usage by a combination of task and either aphasia 645 type or AQ subcomponent score. To this end, we performed multivariate regressions using 646 the four AQ subcomponent scores (naming, repetition, spontaneous speech, comprehension), 647 type of aphasia, and task, either on their own or in combination with the others. Task alone 648 explained a smaller fraction of the variance among the aphasia group: 10.4% (9.6%, 11.8%). 649 Aphasia type alone explained a smaller fraction of the variance than task: 4.9% (4.2%, 7.5%). 650 Task and aphasia type together explained 18.1% (18.5%, 22.8%) of the variance. Although the 651 distances between classes are on average smaller in the aphasia groups, the smaller proportion 652 of variance explained by task in the aphasia group than the control group is at least in part due 653



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.

	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)

Table 5: Distances between tasks for the aphasia group, measured in the full space. 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])

	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)

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.

to a large within-group variance by task among the aphasia groups, which we will highlight 654 below. Additive models with task and AQ scores or task and type of aphasia explained similar 655 fractions of the variance in the linguistic variables. Models with an interaction between AQ 656 scores and task or between aphasia type and task did not explain substantially more of the 657 variance in the linguistic variables than did the corresponding additive models. As AQ score 658 is a linear combination of the four individual AQ subcomponents, we could also test whether a 659 model with task and the four AQ subcomponents explained substantially more of the variance 660 than a model with task and the overall AQ score: we found that it did (p = .02). This led us 661 to the model with additive terms for task and the four AQ subcomponents as the best model 662 of language usage. 663

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 666 and task by aphasia type, just as we did for controls in Figure 2. Note that narrative tasks 667 (three in the case of aphasia: Important Event, Stroke, Cinderella) occupy a space in the top 668 left-to-center quadrant of the BCA space. The two expositional tasks (Cat, Window) occupy 669 a far-right quadrant in the BCA space. One can immediately appreciate that the distances 670 between the tasks is not as large for any of the aphasia types as for the control group (Table 671 6). Notably, and similar again to the control group, the Sandwich task occupies the lower 672 left-hand quadrant of the BCA space, and, for every aphasia type group, the Sandwich task 673 did not overlap with any other tasks. 674

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 visualization 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 686 relationship between the two axes, suggesting that overall aphasia severity may explain a large 687 portion of the covariance shown in the aphasia group. That is, those individuals with the 688 most severe aphasia demonstrate less distinction between tasks (i.e., produce similar linguistic 689 information in dissimilar tasks, as is shown in Figure 3 and Table 6). It is also helpful to 690 appreciate the shape of the groups in the bottom left of Figure 3, where there is variation 691 692 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 693 Anomic group covariance structure most resemble the control group's (Table 7). 694

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

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 expositional or narrative genres.

#### 703 3.2.3 Linguistic information is elicitation method-specific

695

696

697 698

699

700

701

702

We created a two-dimensional space representative of the full space of linguistic variables for 704 our control group and for our aphasia group, separately. Importantly, these two language 705 spaces (as shown in Figures 2 and 3) resemble one-another. For instance, the x-axis (Axis I) 706 appears to relate to morphosyntax and the y-axis (Axis II), lexical information. This is, of 707 course, an over-simplification for the purposes of understanding the multidimensional space 708 represented by our linguistic variables, but it does aid in interpretation. Additionally, in the 709 aphasia group, we took into account variables drawn from a neuropsychological battery (the 710 711 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 712 difference in microstructure for each discourse genre and task. For regression coefficients of 713 linguistic variables and neuropsychological variables on task in the aphasia group, refer to 714 Table 8. 715

The procedural discourse (Sandwich) loads in the bottom left quadrant of BCA space for 716 both subject groups. As such, we can surmise that its morphosyntax is mostly present tense, 717 and that its lexical information is predominantly nouns and prepositions. As noted in Table 718 8, which demonstrates the regression coefficients for the aphasia group, we can appreciate 719 the extent to which nouns, prepositions and conjunctions load most positively, and the high 720 negative loadings of past tense, participles (present and past), and auxiliaries. This loading 721 jives with prior research by our group suggesting that the Sandwich task, representative of 722 the procedural genre, produces morphosyntactically simple phrasing with a reliance on nouns 723 [Stark, 2019]. 724

The narrative discourse genre (Important Event, Cinderella, and Stroke [aphasia group 725 only]) loads in the upper half of the y-axis (Axis II) and relatively central on the x-axis (Axis 726 I) in BCA space. We interpret this to mean a mix of past and present tense usage and, in 727 terms of lexical information, fewer nouns produced. Again referring to Table 8 for loadings of 728 linguistic variables onto narrative genre tasks for the aphasia group, we see positive loadings 729 730 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 731 and adverbs for these tasks. In contrast, linguistic variables loading most positively for the 732 Cinderella task include plurals, past tense and adjectives, as the other narrative genres did, but 733 also nouns and determiners and pronouns. Pronouns are understandably higher in this task, 734



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 is p = 17 fitted values, one for each variable. These fitted values are projected down into a lower-dimensional space that maximizes 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.

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 <i>,</i> 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 <i>,</i> 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)

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

likely due to subject-related local coherence (e.g., talking about Cinderella often, resulting in
 referencing 'her' often).

The expositional 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 742 with linguistic variables as well as neuropsychological variables (i.e., language subtests from 743 the WAB, including auditory verbal comprehension, spontaneous speech, naming and repe-744 tition). Therefore, we are also able to evaluate the extent to which underlying language pro-745 cesses (tapped by these four subtests) explain discourse microstructure by genre and by task. 746 747 The bottom right hand 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 748 we look at the bottom left of Figure 3, where the tasks are plotted by aphasia type, we are able 749 to appreciate the extent to which each task is explained by the neuropsychological variables in 750 tandem with the linguistic variables. We can surmise that the procedural task, which loads in 751 the far bottom left quadrant of neuropsychological and language variable space, likely relies 752 753 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 754 content words and relatively simple morphosyntax (e.g., [Ulatowska et al., 1981, Stark, 2019], 755 suggests that procedural discourse may be a task that engages lexical-semantic ability but not 756 morphosyntactic ability. It also suggests that those performing better on the naming subtest 757 of the WAB (and those who, by proxy, have relatively accessible lexical-semantic processes), 758 may be those driving the linguistic pattern we see in both aphasia and control groups, i.e., the 759 production of content words. There is also a sanity check built into this analysis, which is that 760 the Spontaneous Speech subtest from the WAB ('Spont' in the BCA space) should be logically 761 most associated with the expositional discourse genre tasks (Cat, Window). This is because 762 the Spontaneous Speech score is acquired by rating informational content and fluency of the 763 subject's speaking during a different picture description task (the Picnic Scene), as well as 764 during a basic conversational interview [Kertesz, 2007]. Therefore, there should be similarity 765 between one single picture description expositional task (Picnic Scene) and others (Cat) if in-766 deed there is some generalizability across testing procedures and time, which is what we see 767 here. Auditory Verbal Comprehension and Repetition scores do not appear to be particularly 768 associated with microstructure on any of the tasks. 769

### 770 **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



Figure 4: Covariances (top, visualized 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.

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 neighbors. Meanwhile, both the Cat and the Window expositional tasks do not have significantly different locations on either axis for either subject group
(aphasia, control), suggesting that expositional 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 780 produced by the control group is relatively homogeneous. Therefore, while there is some 781 variance among control speakers (as demonstrated across many prior studies) [Fergadiotis & 782 Wright, 2011, Dalton & Richardson, 2019], neurotypical speakers produced relatively similar, 783 task-specific microstructure. This is an important finding because it provides a normative 784 comparison for the aphasia group; confirms the need for consistency of task instructions and 785 demands within a study; and demonstrates that linguistic information logically and reliably 786 shifts according to task demands. On the other hand, there was large covariance within 787 each aphasia type, showing differential task-specific microstructure according to aphasia type. 788 Overall aphasia severity was the most influential of the aphasia characteristics, mediating the 789 relationship between microstructure and task. 790

To summarize the findings in reference to our original hypotheses: (i) Each genre (narrative, expositional, procedural) elicits relatively unique linguistic information (with some overlap, i.e., expositional 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

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
15.35 PAST PASTP PRESP prep. adj. adv. conj. det. pro. open class closed class	0.029 0.050 0.007 0.034 0.072 -0.006 0.049 0.009 0.063 0.042 0.014 0.025	0.032 0.067 -0.011 -0.049 0.077 -0.056 -0.035 -0.014 -0.007 0.016 -0.005 0.012	0.037 0.038 0.016 0.014 0.038 0.009 0.000 0.000 0.002 0.003 0.002 0.003	-0.061 -0.070 0.000 -0.019 -0.081 0.012 -0.051 -0.038 0.010 -0.108 0.029 -0.039	0.243 0.252 0.039 -0.089 -0.098 0.309 -0.128 0.152 -0.579 0.183 -0.043 -0.088	0.115 0.096 -0.157 0.073 -0.193 0.030 -0.033 -0.002 -0.481 0.091 -0.088 -0.038	0.077 0.201 0.098 -0.083 -0.101 0.151 -0.115 -0.016 0.107 -0.080 0.026 -0.016	-0.038 -0.066 0.116 0.412 0.015 -0.211 0.179 0.130 0.491 -0.353 0.016 0.044	0.075 0.367 0.184 0.425 0.120 -0.020 -0.110 -0.079 0.471 -0.010 0.062 0.024	-0.471 -0.850 -0.279 -0.738 0.258 -0.259 0.207 -0.185 -0.009 0.168 0.026 0.074

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

*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.

at least partially related to aphasia characteristics (type and overall aphasia severity). The 796 task x aphasia (type, overall aphasia severity) interaction likely drives the heterogeneity in 79 the aphasia group as a whole and the within-type covariance. Further, individuals whose 798 aphasia types tend to be associated with greater overall aphasia severity (e.g., BrocaâAZs, 799 Wernicke $\hat{a}\hat{A}\hat{Z}s$  aphasia) did not demonstrate task dissociation as strongly. There is some nu-800 ance to interpreting the influence of overall aphasia severity on microlinguistic information 801 extracted from spoken discourse. Overall aphasia severity is representative of more than just 802 spoken language ability (indeed, it incorporates scores from auditory comprehension, nam-803 ing, repetition, and spontaneous speech), and the result that persons with severe aphasia tend 804 to demonstrate microlinguistic information that is not as strongly dissociable by task must 805 be interpreted through the lens of aphasia type and characteristics. For example, this finding 806 may be driven in one way by individuals with severe Broca's aphasia who have difficulty pro-807 ducing a variety of linguistic information, or by individuals with severe Wernicke's aphasia 808 who have difficulty producing meaningful, well-structured (grammatical) information. These 809 two things may therefore be occurring, regardless of task, thus resulting in a similarity of 810 microlinguistic structure across tasks for individuals with overall severe aphasia. It should be 811 noted that nearly half of our Wernicke's sample, and roughly 35% of our Broca's sample had 812 overall severe aphasia, meaning that the rest of the individuals with Broca's and Wernicke's 813 aphasia presented with overall moderate aphasia. Note, too, that we did not include individ-814 uals with Global aphasia in this analysis, as there were very few in the database. Future work 815 with a larger sample of individuals with Global aphasia and severe Broca's and Wernicke's 816 aphasia could test our hypothesis, that overall aphasia severity results in microlinguistic in-817 formation that is similar across discourse elicitation methods. As it stands, the interaction of 818 overall aphasia severity and type must be taken into account. 819

### 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.

### <sup>827</sup> Clinical implications

Perhaps the most innovative portion of the study, we evaluated the extent to which apha-828 sia characteristics, like type and overall aphasia severity, mediated task- and genre-specific 829 830 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 831 method. In summary, we determined that there are genre- and task-specific microstructural 832 patterns regardless of aphasia type and/or overall aphasia severity. These results are indica-833 tive of a few things. While severe aphasia does seem to blur elicitation method boundaries 834 835 (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 836 speakers with severe Broca's or Wernicke's aphasia, we would expect, at the very least, dif-837 ferent microstructural information to be produced during procedural discourse compared 838 with other discourse genres. But, these results beg the question of how language sampling 839 strategies might differ in those with severe aphasia, as we showed that linguistic informa-840 841 tion was not particularly different for narrative or expositional tasks for these individuals. One might argue that, for those with severe aphasia, acquiring a sample from several dis-842 course genres may not be necessary to comprehensively evaluate that individual's language 843 abilities. But, we would hesitate to argue as such because each discourse genre (and task) 844 likely produces informative, additional language and communicative information beyond mi-845 crostructural information. Indeed, recall that linguistic information (microstructure) is one 846 of three hypothesized levels of analysis that can and should be done on discourse [Sher-847 ratt, 2007, Frederiksen et al., 1990], with the other levels assessing more macrostructural / 848 functional information. For instance, research has shown that certain discourse genres, like 849 expositional and narrative, produce unique gesture types and gesturing loads (i.e., number 850 of linguistic-accompanying gestures, or standalone gestures) [Sekine & Rose, 2013]. Addi-851 tionally, individuals with severe aphasia may still be able to produce task-relevant words 852 (e.g., components of core lexicon and/or main concepts), even if the linguistic structure is the 853 same [Dalton & Richardson, 2019, Dalton & Richardson, 2015]. Therefore, collecting language 854 samples across genres provides information about microstructure as well as information that 855 856 is very useful for understanding communicative ability as a whole.

Our results further highlight wide intra-group (or, within-group) variability. That is, 857 speakers with a shared aphasia type (e.g., Broca's aphasia) produced variable linguistic infor-858 mation during a task. This observation has also been appreciated for language function. Main 859 concept production is one means of evaluating language function by evaluating the ability of 860 the speaker to cover the main topics that should logically be produced during the task (e.g., 861 in the Cinderella story, speakers should discuss the glass slipper, the fairy godmother, coming 862 home at midnight, etc., [Richardson & Dalton, 2016]. In a recent large study evaluating main 863 concept production across several discourse elicitation methods from AphasiaBank in speak-864 ers with and without aphasia, there was considerable within-group heterogeneity [Dalton 865 & Richardson, 2019]. Like with the microstructural within-group variation across discourse 866 genres shown here, the within-group heterogeneity for main concept production was much 867

larger for aphasia types that typically include both moderate and severe aphasia presentation
 (e.g., Broca's, Wernicke's) [Dalton & Richardson, 2019]. Altogether, our work in microstruc ture 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 prioritizing 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] [Fraser et al., 2013, Le & Provost, 2016, Le et al., 2018, Themistocleous & Kokkinakis, 2018] as well as checklists to be used during the live task [Dalton et al., 2020].

#### 878 Transdiagnostic implications

We expect that our findings (i.e., that within-genre microstructure is similar and across-genre 879 microstructure is dissimilar) would hold when investigated in populations of developmentally 880 disparate speakers, due to frame selection being at the earliest level of discourse hierarchy, 881 and thus influencing downstream (micro- and macrostructural) processes [Frederiksen et al., 882 1990, Sherratt, 2007]. Bliss and McCabe (2006) discuss the notion that personal narratives (like 883 our 'Important Event' here), when used to elicit language samples in children, tend to also be 884 characterized by use of the past tense and first and/or third person [Bliss & Mccabe, 2006]. 885 Children reciting a fictional narrative (like a story retell, akin to our 'Cinderella' story) may 886 elicit more utterances and be viewed as an overall 'easier' task because of the given, struc-887 tured information (e.g., the story components in a picture book). But, this fictional narrative 888 may elicit more indefinite speech because of the shared referents (e.g., both child and experi-889 menter have seen the story book), which is very similar to the greater proportion of pronoun 890 and indefinite words used in the Cinderella task (compared with the other narrative tasks) 891 in our results. Together, these findings further clarify different aspects of frame selection, the 892 first step in this kind of discourse. On the one hand, there are task instructions, which we 893 know can influence frame selection even when the same stimuli is presented [Wright & Capi-894 louto, 2009], and further, the shared referents between interlocutor and recipient/addressee 895 further refine the frame. Therefore, our findings regarding discourse microstructure and task 896 interaction replicate ideas about language sampling across development, and emphasize the 897 importance of collecting language samples using multiple, varied discourse genres (which, in 898 themselves, have variable instructions and referents) to comprehensively examine microstruc-899 tural elements. It is not far-fetched to make the conjecture that it is equally important to 900 acquire multiple discourse genres to assess other aspects of language, communication and 901 cognition (e.g., different memory, attention and executive function systems at work in tandem 902 with language). 903

In the field of aphasiology - and indeed, across many populations where behavioral in-904 tervention is considered best practice (e.g., speech-language therapy) - we are tasked with 905 demonstrating that behavioral therapy generalizes to more natural language processes. Speech-906 language therapy has a tendency to focus at the level of isolated processes, like noun/verb 907 retrieval. There has been a considerable issue with demonstrating generalizability of ther-908 apy success to discourse in populations such as aphasia [Boyle, 2011, Rose et al., 2015], in 909 that very few behavioral interventions have actually been shown to generalize to discourse. 910 There are a few explanations for this. A contributing factor is the extraction of insensitive 911 912 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 913 retrieval) [Rose et al., 2015]. There is a considerable amount of work to do to understand how 914 isolated processes being trained in therapy are best represented by variables extracted from 915 discourse. One such way of doing so is to evaluate patterns of discourse (e.g., microstructural 916

variables in a high-dimensional space) rather than single measures of discourse, as we have 917 done here. There have been efforts to model multi-level discourse outcomes [Marini et al., 918 2011] but each level has comprised single outcome measures (e.g., mean length of utterance) 919 and not variables modeled co-dependently. We emphasize moving away from (arbitrary) dis-920 course outcome selection toward comprehensive modeling of all available variables. While 921 we focused on examining microstructure, the analyses we describe here are possible for any 922 923 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 924 that other, more macrostructural variables, such as coherence and cohesion [Galetto et al., 925 2013], main concept [Dalton & Richardson, 2019] and/or core lexicon [Dalton & Richardson, 926 2015] analysis could be modeled in such a fashion, allowing us to appreciate how each of 927 these variables describes a part of the functional language being used during each task. 928

929 An ultimate goal would be to use this statistical modeling to marry structure and function (or, micro- and macrostructural properties), thus providing a layered, high dimensional sys-930 tem that allows us to understand the relation between, and contribution of unique linguistic 931 variables during each task. Further, analyses such as this are appropriate for transdiagnostic 932 populations [Husain, 2017]. For instance, language is affected in a variety of neurogenic and 933 neuropsychiatric disorders, such as post-stroke aphasia, post-traumatic brain injury language 934 impairments, major depressive disorder and primary progressive aphasia. Analyzing spoken 935 discourse in each of these populations, and modeling this information in a high-dimensional 936 space, would allow us to sensitively explore shared (and disparate) patterns of language, as 937 well as the interaction between language severity, disorder, and discourse elicitation method. 938 This is a key future direction that we are presently exploring. 939

### 940 Theoretical implications

Another explanation for the lack of generalization from isolated processes being trained in 941 942 therapy to discourse-level improvements is that we do not yet understand the cognitivelinguistic processes contributing to discourse tasks. For instance, we discussed how narrative 943 discourse relies on processes of long-term memory, and dependent on the task, the memory 944 system being relied upon may be autobiographical or semantic. Further, multiword environ-945 ments like discourse come with processes of language that may not be necessary for the pro-946 cesses being trained during therapy. For example, suppression of word alternatives (whether 947 they are semantic, lexical, or phonological) co-occurs with morphosyntactic processes in mul-948 tiword environments, whereas suppression of word alternatives in isolated tasks does not. 949 While we have yet to understand how the cognitive-linguistic complexity of discourse best 950 relates to isolated tasks, what we have shown here is that each discourse task/genre seems to 951 preferentially tap into a part of the language system. For instance, we showed that speakers 952 tend to preferentially produce more content words and simplest syntax during procedural 953 discourse compared with narrative and expositional discourse. This suggests that procedural 954 discourse may be able to pick up on changes in content word retrieval and production, if in-955 deed that is being trained in therapy. As it stands, researchers and clinicians may be choosing 956 insensitive or inappropriate discourse tasks from which to extract language measures, but we 957 hope that the present study, and other work highlighted in the introduction, are helpful for 958 careful selection of discourse tasks. 959

Another important point that our results emphasize is the importance of task-specific interpretation in papers utilizing 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 impor-966 tant implications for interpretation of studies employing monologic discourse. For example, 967 we should not generalize the findings from a study employing fictional story retelling (e.g., 968 Cinderella) to findings from a study employing autobiographical story telling (e.g., Impor-969 tant Event). Indeed, the emphasis on different memory systems may be a primary factor 970 explaining the dissociation in linguistic information produced across narrative tasks, as well 971 972 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 973 fields that acquire discourse as a means of understanding a person's linguistic (and indeed, 974 cognitive) abilities in a more naturalistic setting than provided by standardized testing. Fur-975 ther, language can be understood also at a broader (not just deeper) level by using multiple 976 discourse genres. We believe that the push by others [Nicholas & Brookshire, 1995, Brookshire 977 978 & Nicholas, 1994, Fergadiotis & Wright, 2011, Ulatowska et al., 1981, Olness, 2006, Dalton & Richardson, 2015, Dalton et al., 2020, Richardson & Dalton, 2016] and ourselves [Stark, 2019] 979 for the acquisition of multiple, varied discourse genres when sampling discourse gains more 980 traction here, specific to evaluating microstructure. 981

Our results highlight that task instructions and, in following, experimental fidelity is crit-982 ical. Fidelity is, of course, hyper-critical for all experiments, but it becomes increasingly 983 challenging to ensure fidelity across studies where instructions are semi-structured and open 984 to interpretation, like during monologic discourse and other naturalistic tasks. Our work and 985 others [Wright & Capilouto, 2009, Stark, 2019] provide evidence for the establishment of fi-986 delity during all stages of language sampling, from acquisition to transcription to coding to 987 analysis. A recently founded working group, FOQUSAphasia ('FOstering QUality of Spoken 988 discourse research in Aphasia; www.foqusaphasia.com; [Stark et al., 2020]) has endeavored 989 to create initiatives to foster fidelity and standardization for discourse in aphasia. Without 990 having these standards, it is difficult to reproduce or replicate prior findings in the field, and 991 to draw conclusions (e.g., meta-analyses, scoping reviews) about spoken discourse across a 992 993 variety of settings (e.g., assessment, therapy).

Finally, this study demonstrates that monologic discourse retains consistent microstructure 994 within subject groups. Recall that members of the control group, as well as members of the 995 aphasia group with mildest language impairment (i.e., Not Aphasic by WAB and mild anomic 996 groups) produced relatively homogeneous genre- and task-specific linguistic information (i.e., 997 smaller within-group covariance). This finding suggests that the tasks being employed here 998 produce consistent linguistic elements when given to people with no language impairment 999 or mild language impairment, and also demonstrates consistency across subjects of varying 1000 education, age, literacy, sex and presence of brain injury. This gives considerable credence 1001 to the use of these monologic discourse in experimental settings, allowing us to bridge the 1002 gap between every day communication (e.g., conversation) and isolated language tasks (e.g., 1003 confrontation naming). In general, standardized norms for discourse-level language are lack-1004 ing in the field of aphasiology (and in development [Applebee, 1978, Peterson & McCabe, 1005 2013]). But, the accumulation of evidence, including results from this study, suggests that 1006 we may be able to create standardized norms for aspects of discourse. This will improve our 1007 understanding of discourse change (i.e., in development, after brain injury, over the course of 1008 neurodegenerative disease). 1009

#### 1010 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 compar-

isons had small group sizes (e.g., sixteen speakers with Broca's aphasia in the Sandwich task). 1015 Further, few of the speakers were bi- or multilingual and most speakers had relatively simi-1016 lar levels of education (ranging from 12-20 years, typically). Therefore, our sample may not 1017 generalize to populations that are traditionally bi- or multilingual, or to samples with lower 1018 socioeconomic status and education. A priority should be on replication of our findings in 1019 diverse samples. 1020

1021

Our analysis relied on semi-automatic morphological tagging done by the CLAN program [MacWhinney, 2000]. While the morphological tagging has been validated in Aphasi-1022 aBank prior research, it may still be the case that there is incorrect word-level tagging due 1023 to inherent ambiguity in linguistic information when few tokens were produced. We would 1024 argue that this would also be the case if done by the 'gold standard' (e.g., human raters). 1025 Replication of these results in a new sample will speak to the quality and reproducibility of 1026 1027 task x microstructure interaction.

Finally, this analysis was restricted to only a single level of discourse (microstructure, or the 1028 language used in the discourse), and did not evaluate further levels, such as macrostructure 1029 / functional information [Frederiksen et al., 1990, Sherratt, 2007]. We intend to conduct future 1030 work linking together micro- and macrostructural levels, as has been suggested by others 1031 [Sherratt, 2007, Dipper & Pritchard, 2017, Marini et al., 2011]. 1032

### **References**

1034	Applebee, A. N. (1978). The child's concept of story: Ages two to seventeen.
1035	Becker, J. T., Lopez, O. L., Saxton, J., Mcgonigle, K. L., Alzheimer, M, & Program, R.
1036	(1994). The natural history of Alzheimer's disease description of study cohort
1037	and accuracy of diagnosis. Archives of Neurology, 51(6), 585–594.
1038	Berman, R. (2009). Trends in research on narrative development. In S. Foster-Cohen
1039	(Ed.), Language Acquisition (pp. 294–318). London, UK: Palgrave Macmillan.
1040	Bliss, L. S. & Mccabe, A. (2006). Comparison of discourse genres: Clinical implica-
1041	tions. <i>Contemporary issues in communication sciences and disorders</i> , 33.
1042	Boles, L. (1998). Conversational discourse analysis as a method for evaluating
1043	progress in aphasia: A case report. Journal of Communication Disorders, 31, 261-
1044	274.
1045	Boyle, M. (2011). Discourse treatment for word retrieval impairment in aphasia: The
1046	story so far. <i>Aphasiology</i> , 25(11), 1308–1326.
1047	Boyle, M. (2014). Test-retest stability of word retrieval in aphasic discourse. Journal of
1048	Speech, Language, and Hearing Research, 57(June), 966–978.
1049	Brady, M., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and
1050	language therapy for aphasia following stroke. <i>Cochrane Database of Systematic</i>
1051	Reviews, (6).
1052	Brenneise-Sarshad, R., Nicholas, L. E., & Brookshire, R. H. (1991). Effects of appar-
1053	ent listener knowledge and picture stimuli on aphasic and non-brain-damaged
1054	speakers' narrative discourse. Journal of Speech and Hearing Research, 34(1), 168-
1055	176.
1056	Brookshire, R. & Nicholas, L. (1994). Speech sample-size and test-retest stability of
1057	connected speech measures for adults with aphasia. Journal of Speech, Language
1058	and Hearing Research, 37(2), 399–407.
1059	Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in
1060	aphasia: A review of the literature. Clinical Linguistics & Phonetics, 30(7), 489-
1061	518.
1062	Bryant, L., Spencer, E., & Ferguson, A. (2017). Clinical use of linguistic discourse
1063	analysis for the assessment of language in aphasia. Aphasiology, 31(10), 1105-
1064	1126.
1065	Cameron, R. M., Wambaugh, J. L., & Mauszycki, S. C. (2010). Individual variability
1066	on discourse measures over repeated sampling times in persons with aphasia.
1067	Aphasiology, 24(6-8), 671–684.
1068	Correia, L., Brookshire, R. H., & Nicholas, L. E. (1989). The effects of picture content on
1069	descriptions by aphasic and non-brain damaged speakers. Clinical Aphasiology,
1070	18, 447–462.
1071	Correia, L., Brookshire, R. H., & Nicholas, L. E. (1990). Aphasic and non-brain-
1072	damaged adults' descriptions of aphasia test pictures and gender-biased pic-
1073	tures. Journal of Speech and Hearing Disorders, 55(4), 713–720.
1074	Cruice, M., Botting, N., Marshall, J., Boyle, M., Hersh, D., Pritchard, M., & Dipper,
1075	L. (2020). UK speech and language therapists' views and reported practices of
1076	discourse analysis in aphasia rehabilitation. International Journal of Language and
1077	Communication Disorders, (pp. 1–26).

Dalton, S. G., Kim, H., Richardson Heather, J. D., & Wright, H. (2020). A compendium 1078 of core lexicon checklists. *Seminars in speech and language*, 41(1), 45–60. 1079 Dalton, S. G. & Richardson, J. (2018). A large-scale comparison of main concept 1080 production between persons with aphasia and persons without brain injury. 1081 American Journal of Speech-Language Pathology, (pp. 1–28). 1082 Dalton, S. G. & Richardson, J. D. (2015). Core-Lexicon and main-concept production 1083 during picture-sequence description in adults without brain damage and adults 1084 with aphasia. American Journal of Speech-Language Pathology, 24(4). 1085 Dalton, S. G. & Richardson, J. D. (2019). A large-scale comparison of main concept 1086 production between persons with aphasia and persons without brain injury. 1087 American Journal of Speech-Language Pathology, 28, 293–320. 1088 Davis, G. (2007). Aphasiology: Disorders and clinical practice. Pearson, 2 edition. 1089 Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predic-1090 tions from a connectionist model. *Journal of Memory and Language*, 27(2), 124–142. 1091 Dipper, L. T. & Pritchard, M. (2017). Discourse: Assessment and therapy. In Advances 1092 in Speech-language Pathology. InTech. 1093 Doyle, P. J., McNeil, M. R., Spencer, K. A., Goda, A. J., Cottrell, K., & Lustig, A. P. 1094 (1998). The effects of concurrent picture presentations on retelling of orally 1095 presented stories by adults with aphasia. In Aphasiology, volume 12 (pp. 561– 1096 574).: Routledge. 1097 Easterbrook, A., Brown, B. B., & Perera, K. (1982). A comparison of the speech of 1098 adult aphasic subjects in spontaneous and structured interactions. International 1099 Journal of Language & Communication Disorders, 17(3), 93–107. 1100 Faroqi-Shah, Y., Treanor, A., Ratner, N. B., Ficek, B., Webster, K., & Tsapkini, K. (2020). 1101 Using narratives in differential diagnosis of neurodegenerative syndromes. *Jour-*1102 nal of Communication Disorders, (pp. 105994). 1103 Fergadiotis, G. & Wright, H. (2011). Lexical diversity for adults with and without 1104 aphasia across discourse elicitation tasks. *Aphasiology*, 25(11), 1414–1430. 1105 Fergadiotis, G., Wright, H. H., & Capilouto, G. J. (2011). Productive vocabulary across 1106 discourse types. *Aphasiology*, 25(10), 1261–1278. 1107 Fraser, K., Rudzicz, F., Graham, N., & Rochon, E. (2013). Automatic speech recognition 1108 in the diagnosis of primary progressive aphasia. SLPAT 2013, 4th Workshop on 1109 Speech and Language Processing for Assistive Technologies, (pp. 47–54). 1110 Fraser, K. C., Meltzer, J. A., & Rudzicz, F. (2015). Linguistic features identify 1111 Alzheimer's disease in narrative speech. Journal of Alzheimer's Disease, 49(2), 1112 407-422. 1113 Frederiksen, C. H., Bracewell, R. J., Breuleux, A., & Renaud, A. (1990). The cogni-1114 tive representation and processing of discourse: Function and dysfunction. In 1115 *Discourse Ability and Brain Damage* (pp. 69–110). Springer-Verlag. 1116 Fromm, D., Forbes, M., Holland, A., & Dalton, G. (2017). Discourse characteristics in 1117 aphasia beyond the Western Aphasia Battery cutoff. 26(August), 1–7. 1118 Fromm, D., Greenhouse, J., Hou, K., Russel, G., Cai, X., Forbes, M., Holland, A., & 1119 MacWhinney, B. (2016). Automated proposition density analysis for discourse 1120 in aphasia. Journal of Speech, Language and Hearing Research, 59, 1123–1132. 1121 Galetto, V., Kintz, S., West, T., Mrini, A., & Wright, H. (2013). Measuring global 1122

1123	coherence in aphasia. <i>Procedia-Social and Behavioral Sciences</i> , 94, 198–199.
1124	Glosser, G. & Deser, T. (1992). A comparison of changes in macrolinguistic and mi-
1125	crolinguistic aspects of discourse production in normal aging. Journal of Geron-
1126	tology, 47(4), P266–P272.
1127	Glosser, G., Wiener, M., & Kaplan, E. (1988). Variations in aphasic language behaviors.
1128	Journal of Speech and Hearing Disorders, 53(2), 115–124.
1129	Goodglass, H. & Kaplan, E. (1972). Boston Diagnostic Aphasia Examination. Philadel-
1130	phia, Pa: Lea & Febiger.
1131	Goodglass, H., Kaplan, E., & Barresi, B. (2000). Boston Diagnostic Aphasia Examination.
1132	Pearson, 3 edition.
1133	Husain, M. (2017). Transdiagnostic neurology: Neuropsychiatric symptoms in neu-
1134	rodegenerative diseases. Brain, 140, 1535–1536.
1135	Kertesz, A. (2007). Western Aphasia Battery - Revised. San Antonio, TX: The Psycholog-
1136	ical Corporation.
1137	Le, D., Licata, K., & Mower Provost, E. (2018). Automatic quantitative analysis of
1138	spontaneous aphasic speech. Speech Communication, 100, 1–12.
1139	Le, D. & Provost, E. M. (2016). Improving automatic recognition of aphasic
1140	speech with AphasiaBank. Proceedings of the Annual Conference of the In-
1141	ternational Speech Communication Association, INTERSPEECH, 08-12-September-
1142	2016(March), 2681–2685.
1143	Levelt, W. (1989). Speaking: From intention to articulation. Cambridge, MA: MIT Press.
1144	Linnik, A., Bastiaanse, R., & Höhle, B. (2016). Discourse production in aphasia: A
1145	current review of theoretical and methodological challenges.
1146	MacWhinney, B. (2000). The CHILDES Project: Tools for Analysing Talk. Mahwah, NJ:
1147	Lawrence Erlbaum Associates Inc., 3 edition.
1148	MacWhinney, B., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for study-
1149	ing discourse. Aphasiology, 25(11), 1286–1307.
1150	MacWhinney, B., Fromm, D., Holland, A., Forbes, M., & Wright, H. (2010). Automated
1151	analysis of the Cinderella story. <i>Aphasiology</i> , 24(0), 856–868.
1152	Marini, A., Andreetta, S., del Tin, S., & Carlomagno, S. (2011). A multi-level approach
1153	to the analysis of narrative language in aphasia. <i>Aphasiology</i> , 25(11), 1372–1392.
1154	Nadeau, S. E. (2001). Phonology: A review and proposals from a connectionist per-
1155	spective. <i>Brain and Language</i> , 79(3), 511–579.
1156	Nicholas, L. & Brookshire, R. (1995). Presence, completeness, and accuracy of main
1157	concepts in the connected speech of non-brain-damaged adults and adults with
1158	aphasia. Journal of Speech, Language and Hearing Research, 38, 145–156.
1159	Olness, G. S. (2006). Genre, verb, and coherence in pictureâARelicited discourse of
1160	adults with aphasia. <i>Aphasiology</i> , 20(2-4), 175–187.
1161	Peterson, C. & McCabe, A. (2013). <i>Developmental psycholinguistics: Three ways of looking</i>
1162	at a childâAZs narrative. Springer Science & Business Media.
1163	Richardson, J. D. & Dalton, S. G. (2016). Main concepts for three different discourse
1164	tasks in a large non-clinical sample. <i>Aphasiology</i> , 30(1), 45–73.
1165	Roberts, J. A. & Wertz, R. T. (1989). Comparison of spontaneous and elicited oral-
1166	expressive language in aphasia. <i>Clinical Aphasiology</i> , 18, 479–488.
1167	Rose, M. L., Mok, Z., Carragher, M., Katthagen, S., & Attard, M. (2015). Comparing

1168	multi-modality and constraint-induced treatment for aphasia: A preliminary
1169	investigation of generalisation to discourse. <i>Aphasiology</i> , 30(6), 678–698.
1170	Sahraoui, H. & Nespoulous, JL. (2012). Across-task variability in agrammatic per-
1171	formance. Aphasiology, 26(6), 785–810.
1172	Sekine, K. & Rose, M. L. (2013). The relationship of aphasia type and gesture produc-
1173	tion in people with aphasia. <i>American Journal of Speech-Language Pathology</i> , 22(4),
1174	662–672.
1175	Sherratt, S. (2007). MultiâĂŘlevel discourse analysis: A feasible approach. Aphasiol-
1176	<i>ogy</i> , 21(3-4), 375–393.
1177	Stark, B. C. (2019). A comparison of three discourse elicitation methods in aphasia
1178	and age-matched adults: Implications for language assessment and outcome.
1179	American Journal of Speech-Language Pathology, 28(3), 1067–1083.
1180	Stark, B. C., Dutta, M., Murray, L. L., Bryant, L., Fromm, D., MacWhinney, B., Ram-
1181	age, A. E., Roberts, A., den Ouden, D. B., Brock, K., McKinney-Bock, K., Paek,
1182	E. J., Harmon, T. G., Yoon, S. O., Themistocleous, C., Yoo, H., Aveni, K., Gutier-
1183	rez, S., & Sharma, S. (2020). Standardizing assessment of spoken discourse in
1184	aphasia: A working group with deliverables. American Journal of Speech-Language
1185	Pathology.
1186	Themistocleous, H. & Kokkinakis, D. (2018). THEMIS-SV: Automatic classification of
1187	language disorders from speech signals Linguistic and extra-linguistic parame-
1188	ters for early detection of cognitive impairment.
1189	Ulatowska, H. K., North, A. J., & Macaluso-Haynes, S. (1981). Production of narrative
1190	and procedural discourse in aphasia. <i>Brain and Language</i> , 13(2), 345–371.
1191	Van Leer, E. & Turkstra, L. (1999). The effect of elicitation on discourse coherence and
1192	cohesion in adolescents with brain injury. Journal of Communication Disorders, 32,
1193	327-349.
1194	Wright, H. H. & Capilouto, G. J. (2009). Manipulating task instructions to change
1195	narrative discourse performance. <i>Aphasiology</i> , 23(10), 1295–1308.

### <sup>1196</sup> Supplementary Material

$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	Anomic							
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	1 intennie	Important Event	Stroke	Cinderella	Cat	Window		
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Stroke	0.60 (0.44, 0.76)						
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	Cinderella	1.03 (0.90, 1.18)	0.94 (0.80, 1.09)					
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Cat	2.07 (1.88, 2.28)	1.90 (1.70, 2.08)	1.41 (1.25, 1.58)				
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	Window	1.92 (1.75, 2.11)	1.74 (1.57, 1.91)	1.36 (1.21, 1.51)	0.94 (0.75, 1.13)			
Broca's       Important Event       Stroke       Cinderella       Cat       Window         Stroke $0.74 (0.52, 0.96)$ $1.11 (0.80, 1.43)$ $1.16 (0.90, 1.47)$ $2.01 (1.66, 2.42)$ $1.82 (1.48, 2.21)$ $1.33 (1.02, 1.70)$ Window $1.79 (1.41, 2.17)$ $1.69 (1.30, 2.08)$ $1.34 (0.97, 1.73)$ $1.13 (0.84, 1.46)$ Sandwich $1.87 (1.44, 2.33)$ $1.85 (1.47, 2.30)$ $1.64 (1.26, 2.09)$ $1.96 (1.48, 2.51)$ $2.18 (1.70, 2.71)$ Conduction       Important Event       Stroke       Cinderella       Cat       Window $0.87 (0.69, 1.07)$ $0.96 (0.81, 1.13)$ $1.32 (1.03, 1.61)$ Undow $0.87 (0.69, 1.07)$ $0.96 (0.81, 1.13)$ Cat $0.77 (0.52, 0.91)$ $0.46 (1.27, 1.92)$ $1.54 (1.29, 1.81)$ $1.32 (1.03, 1.61)$ Undow         Window $1.66 (1.40, 1.95)$ $1.48 (1.25, 1.70)$ $1.11 (0.83, 1.43)$ $1.04 (0.80, 1.29)$ Sandwich $1.75 (1.47, 2.02)$ $1.70 (1.47, 1.92)$ $1.85 (1.57, 2.11)$ $2.22 (1.90, 2.52)$ $2.29 (1.89, 2.64)$ NotAphasicByWAB       Important Event       Stroke       Cinderella       Cat       Window         Stroke $0.87 (0.62, 1.11)$ <	Sandwich	2.14 (1.88, 2.38)	1.89 (1.67, 2.10)	1.91 (1.69, 2.11)	2.66 (2.46, 2.85)	2.71 (2.49, 2.93)		
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$						7 ( 1), )))		
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	Broca's							
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	bioca 3	Important Event	Stroke	Cinderella	Cat	Window		
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Stroke	0.74 (0.52, 0.96)						
$\begin{array}{c cccc} Cat & 2.01 (1.66, 2.42) & 1.82 (1.48, 2.21) & 1.33 (1.02, 1.70) \\ Window & 1.79 (1.41, 2.17) & 1.69 (1.30, 2.08) & 1.34 (0.97, 1.73) & 1.13 (0.84, 1.46) \\ Sandwich & 1.87 (1.44, 2.33) & 1.85 (1.47, 2.30) & 1.64 (1.26, 2.09) & 1.96 (1.48, 2.51) & 2.18 (1.70, 2.71) \\ \hline \\ \hline \\ \hline \\ Conduction & & & & & & & & & & & & & & & & & & &$	Cinderella	1.11 (0.80, 1.43)	1.16 (0.90, 1.47)					
Window $1.79$ (1.41, 2.17) $1.69$ (1.30, 2.08) $1.34$ (0.97, 1.73) $1.13$ (0.84, 1.46)         Sandwich $1.87$ (1.44, 2.33) $1.85$ (1.47, 2.30) $1.64$ (1.26, 2.09) $1.96$ (1.48, 2.51) $2.18$ (1.70, 2.71)         Conduction       Important Event       Stroke       Cinderella       Cat       Window         Stroke $0.71$ (0.52, 0.91) $0.96$ (0.81, 1.13) $1.32$ (1.03, 1.61) $1.04$ (0.80, 1.29)         Cat $1.75$ (1.42, 2.09) $1.54$ (1.29, 1.81) $1.32$ (1.03, 1.61) $1.04$ (0.80, 1.29)         Sandwich $1.75$ (1.42, 2.09) $1.54$ (1.29, 1.81) $1.32$ (1.03, 1.61) $1.04$ (0.80, 1.29)         Sandwich $1.75$ (1.47, 2.02) $1.70$ (1.47, 1.92) $1.85$ (1.57, 2.11) $2.22$ (1.90, 2.52) $2.29$ (1.89, 2.64)         NotAphasicByWAB       Important Event       Stroke       Cinderella       Cat       Window         Grid $1.02$ (0.73, 1.39) $1.02$ (0.88, 1.21) $1.55$ (1.29, 1.89) $1.02$ (0.82, 1.24)         Cat $1.70$ (1.38, 2.04) $1.56$ (1.33, 1.83) $1.13$ (0.83, 1.47) $1.02$ (0.82, 1.24)         Sandwich $2.60$ (2.29, 2.87) $2.17$ (1.46, 2.11) $1.55$ (1.29, 1.89) $1.02$ (0.82, 1.24)	Cat	2.01 (1.66, 2.42)	1.82 (1.48, 2.21)	1.33 (1.02, 1.70)				
Sandwich       1.87 (1.44, 2.33)       1.85 (1.47, 2.30)       1.64 (1.26, 2.09)       1.96 (1.48, 2.51)       2.18 (1.70, 2.71)         Conduction       Important Event       Stroke       Cinderella       Cat       Window         Stroke       0.71 (0.52, 0.91)       0.96 (0.81, 1.13)       1.32 (1.03, 1.61)       1.04 (0.80, 1.29)         Cat       1.75 (1.42, 2.09)       1.54 (1.29, 1.81)       1.32 (1.03, 1.61)       1.04 (0.80, 1.29)         Sandwich       1.66 (1.40, 1.95)       1.48 (1.25, 1.70)       1.11 (0.83, 1.43)       1.04 (0.80, 1.29)         Sandwich       1.75 (1.47, 2.02)       1.70 (1.47, 1.92)       1.85 (1.57, 2.11)       2.22 (1.90, 2.52)       2.29 (1.89, 2.64)         NotAphasicByWAB       Important Event       Stroke       Cinderella       Cat       Window         Stroke       0.87 (0.62, 1.11)       1.02 (0.88, 1.21)       1.26 (0.73, 1.39)       1.02 (0.88, 1.21)         Cat       1.70 (1.38, 2.04)       1.56 (1.33, 1.83)       1.13 (0.83, 1.47)       Window         Window       1.93 (1.56, 2.32)       1.76 (1.46, 2.11)       1.55 (1.29, 1.89)       1.02 (0.82, 1.24)         Sandwich       2.60 (2.29, 2.87)       2.17 (1.94, 2.40)       2.26 (2.02, 2.52)       2.73 (2.44, 3.04)       2.99 (2.71, 3.26)         Wernicke's       Import	Window	1.79 (1.41, 2.17)	1.69 (1.30, 2.08)	1.34 (0.97, 1.73)	1.13 (0.84, 1.46)			
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	Sandwich	1.87 (1.44, 2.33)	1.85 (1.47, 2.30)	1.64 (1.26, 2.09)	1.96 (1.48, 2.51)	2.18 (1.70, 2.71)		
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$			5 ( 17, 5 )		) (   , ) ,			
$\begin{tabular}{ l l l l l l l l l l l l l l l l l l l$	Conduction							
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	Conduction	Important Event	Stroke	Cinderella	Cat	Window		
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	Stroke	0.71 (0.52, 0.91)						
Cat       1.75 (1.42, 2.09)       1.54 (1.29, 1.81)       1.32 (1.03, 1.61)         Window       1.66 (1.40, 1.95)       1.48 (1.25, 1.70)       1.11 (0.83, 1.43)       1.04 (0.80, 1.29)         Sandwich       1.75 (1.47, 2.02)       1.70 (1.47, 1.92)       1.85 (1.57, 2.11)       2.22 (1.90, 2.52)       2.29 (1.89, 2.64)         NotAphasicByWAB       Important Event       Stroke       Cinderella       Cat       Window         Stroke       0.87 (0.62, 1.11)       1.02 (0.88, 1.21)       1.33 (0.83, 1.47)       1.02 (0.82, 1.24)         Cat       1.70 (1.38, 2.04)       1.56 (1.33, 1.83)       1.13 (0.83, 1.47)       1.02 (0.82, 1.24)         Sandwich       2.60 (2.29, 2.87)       2.17 (1.94, 2.40)       2.26 (2.02, 2.52)       2.73 (2.44, 3.04)       2.99 (2.71, 3.26)         Wernicke's       Important Event       Stroke       Cinderella       Cat       Window         Stroke       0.90 (0.64, 1.21)       1.18 (0.89, 1.55)       2.60 (2.29, 2.17)       2.40       2.26 (2.02, 2.52)       2.73 (2.44, 3.04)       2.99 (2.71, 3.26)         Wernicke's       Important Event       Stroke       Cinderella       Cat       Window         Stroke       0.90 (0.64, 1.21)       1.18 (0.89, 1.55)       1.42 (0.75, 1.60)       1.42 (0.65, 1.82)       2.42 (0.66, 1.82)	Cinderella	0.87 (0.69, 1.07)	0.96 (0.81, 1.13)					
Window1.66 (1.40, 1.95)1.48 (1.25, 1.70)1.11 (0.83, 1.43)1.04 (0.80, 1.29)Sandwich1.75 (1.47, 2.02)1.70 (1.47, 1.92)1.85 (1.57, 2.11)2.22 (1.90, 2.52)2.29 (1.89, 2.64)NotAphasicByWABImportant EventStrokeCinderellaCatWindowStroke0.87 (0.62, 1.11)1.02 (0.73, 1.39)1.02 (0.88, 1.21)Cat1.70 (1.38, 2.04)1.56 (1.33, 1.83)1.13 (0.83, 1.47)Window1.93 (1.56, 2.32)1.76 (1.46, 2.11)1.55 (1.29, 1.89)Sandwich2.60 (2.29, 2.87)2.17 (1.94, 2.40)2.26 (2.02, 2.52)Sandwich2.60 (2.29, 2.87)2.17 (1.94, 2.40)2.26 (2.02, 2.52)Wernicke'sImportant EventStrokeCinderellaCat1.29 (1.05, 1.60)1.18 (0.89, 1.55)Cat1.52 (1.56, 2.18)1.56 (1.26, 1.94)Mindow1.92 (1.05, 1.60)1.18 (0.89, 1.55)Cat1.72 (1.92, 1.88)1.92 (0.94, 1.60)Window1.92 (1.05, 1.60)1.56 (1.26, 1.94)Mindow1.92 (1.92, 1.80)1.52 (0.96, 1.82)	Cat	1.75(1.42, 2.09)	1.54 (1.29, 1.81)	1.32 (1.03, 1.61)				
Sandwich1.75 $(1.47, 2.02)$ 1.70 $(1.47, 1.92)$ 1.85 $(1.57, 2.11)$ 2.22 $(1.90, 2.52)$ 2.29 $(1.89, 2.64)$ NotAphasicByWABImportant EventStrokeCinderellaCatWindowStroke0.87 $(0.62, 1.11)$ 1.02 $(0.88, 1.21)$ 1.13 $(0.83, 1.47)$ Cat1.70 $(1.38, 2.04)$ 1.56 $(1.33, 1.83)$ 1.13 $(0.83, 1.47)$ Window1.93 $(1.56, 2.32)$ 1.76 $(1.46, 2.11)$ 1.55 $(1.29, 1.89)$ 1.02 $(0.82, 1.24)$ Sandwich2.60 $(2.29, 2.87)$ 2.17 $(1.94, 2.40)$ 2.26 $(2.02, 2.52)$ 2.73 $(2.44, 3.04)$ 2.99 $(2.71, 3.26)$ Wernicke'sImportant EventStrokeCinderellaCatWindowStroke0.90 $(0.64, 1.21)$ 1.18 $(0.89, 1.55)$ CatUindowStroke1.29 $(1.05, 1.60)$ 1.18 $(0.89, 1.55)$ 1.24 $(0.04, 1.60)$ Window1.92 $(1.57, 2.18)$ 1.56 $(1.26, 1.94)$ 1.14 $(0.75, 1.60)$ Window1.91 $(1.96, 2.12)$ 1.92 $(1.92, 1.48)$ 1.92 $(0.94, 1.82)$	Window	1.66 (1.40, 1.95)	1.48(1.25, 1.70)	1.11 (0.83, 1.43)	1.04 (0.80, 1.29)			
NotAphasicByWAB       Important Event       Stroke       Cinderella       Cat       Window         Stroke $0.87 (0.62, 1.11)$ $(0.73, 1.39)$ $1.02 (0.88, 1.21)$ $(0.87, 1.13, 2.04)$ $1.56 (1.33, 1.83)$ $1.13 (0.83, 1.47)$ Window $1.93 (1.56, 2.32)$ $1.76 (1.46, 2.11)$ $1.55 (1.29, 1.89)$ $1.02 (0.82, 1.24)$ Sandwich $2.60 (2.29, 2.87)$ $2.17 (1.94, 2.40)$ $2.26 (2.02, 2.52)$ $2.73 (2.44, 3.04)$ $2.99 (2.71, 3.26)$ Wernicke's       Important Event       Stroke       Cinderella       Cat       Window         Stroke $0.90 (0.64, 1.21)$ $1.18 (0.89, 1.55)$ $1.14 (0.75, 1.60)$ $1.37 (0.96, 1.82)$ Window $1.72 (1.65, 2.18) = 1.56 (1.26, 1.94) = 1.42 (0.94, 1.60)$ $1.27 (0.96, 1.82)$	Sandwich	1.75(1.47, 2.02)	1.70(1.47, 1.92)	1.85(1.57, 2.11)	2.22(1.90, 2.52)	2.29 (1.89, 2.64)		
NotAphasicByWAB           Important Event         Stroke         Cinderella         Cat         Window           Stroke $0.87 (0.62, 1.11)$	_							
$\frac{\text{Important Event Stroke}}{\text{Stroke}} = \frac{\text{Cinderella}}{\text{Cat}} = \frac{\text{Cat}}{\text{Window}}$ $\frac{\text{Stroke}}{\text{Stroke}} = \frac{0.87 (0.62, 1.11)}{0.2 (0.73, 1.39)} = 1.02 (0.88, 1.21)$ $\frac{\text{Cat}}{\text{Cat}} = 1.70 (1.38, 2.04) = 1.56 (1.33, 1.83) = 1.13 (0.83, 1.47)$ $\text{Window} = 1.93 (1.56, 2.32) = 1.76 (1.46, 2.11) = 1.55 (1.29, 1.89) = 1.02 (0.82, 1.24)$ $\text{Sandwich} = 2.60 (2.29, 2.87) = 2.17 (1.94, 2.40) = 2.26 (2.02, 2.52) = 2.73 (2.44, 3.04) = 2.99 (2.71, 3.26)$ $\frac{\text{Wernicke's}}{\text{Stroke}} = \frac{\text{Important Event Stroke}}{0.90 (0.64, 1.21)} = \frac{\text{Stroke}}{1.18 (0.89, 1.55)} = \frac{\text{Cinderella}}{1.29 (1.05, 1.60)} = 1.18 (0.89, 1.55) = 1.25 (0.04, 1.60)$ $\frac{1.72 (1.26, 2.18)}{1.72 (1.26, 2.13)} = 1.52 (1.26, 1.94) = 1.25 (0.04, 1.60)$	NotAphasicByWAB							
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	r to u spruoroz	Important Event	Stroke	Cinderella	Cat	Window		
$\begin{array}{c cccc} Cinderella & 1.02 & (0.73, 1.39) & 1.02 & (0.88, 1.21) \\ Cat & 1.70 & (1.38, 2.04) & 1.56 & (1.33, 1.83) & 1.13 & (0.83, 1.47) \\ Window & 1.93 & (1.56, 2.32) & 1.76 & (1.46, 2.11) & 1.55 & (1.29, 1.89) & 1.02 & (0.82, 1.24) \\ Sandwich & 2.60 & (2.29, 2.87) & 2.17 & (1.94, 2.40) & 2.26 & (2.02, 2.52) & 2.73 & (2.44, 3.04) & 2.99 & (2.71, 3.26) \\ \hline \\ \hline \\ Wernicke's & & \\ \hline \\ Wernicke's & & \\ \hline \\ Stroke & 0.90 & (0.64, 1.21) & \\ Cinderella & 1.29 & (1.05, 1.60) & 1.18 & (0.89, 1.55) & \\ Cat & 1.85 & (1.56, 2.18) & 1.56 & (1.26, 1.94) & 1.14 & (0.75, 1.60) & \\ \hline \\ Window & 1.73 & (1.26, 2.12) & & 1.52 & (1.26, 1.94) & 1.25 & (0.06, 1.82) & \\ \hline \\ \hline \\ Window & 1.73 & (1.26, 2.12) & & 1.52 & (1.22, 1.82) & 1.25 & (0.06, 1.82) & \\ \hline \end{array}$	Stroke	0.87 (0.62, 1.11)						
Cat $1.70 (1.38, 2.04)$ $1.56 (1.33, 1.83)$ $1.13 (0.83, 1.47)$ Window $1.93 (1.56, 2.32)$ $1.76 (1.46, 2.11)$ $1.55 (1.29, 1.89)$ $1.02 (0.82, 1.24)$ Sandwich $2.60 (2.29, 2.87)$ $2.17 (1.94, 2.40)$ $2.26 (2.02, 2.52)$ $2.73 (2.44, 3.04)$ $2.99 (2.71, 3.26)$ Wernicke'sImportant Event StrokeCinderellaCatWindowStroke $0.90 (0.64, 1.21)$ $1.18 (0.89, 1.55)$ $1.14 (0.75, 1.60)$ Cat $1.85 (1.56, 2.18)$ $1.56 (1.26, 1.94)$ $1.14 (0.75, 1.60)$ Window	Cinderella	1.02 (0.73, 1.39)	1.02 (0.88, 1.21)					
Window Sandwich $1.93 (1.56, 2.32)$ $1.76 (1.46, 2.11)$ $1.55 (1.29, 1.89)$ $1.02 (0.82, 1.24)$ Sandwich $2.60 (2.29, 2.87)$ $2.17 (1.94, 2.40)$ $2.26 (2.02, 2.52)$ $2.73 (2.44, 3.04)$ $2.99 (2.71, 3.26)$ Wernicke'sImportant EventStrokeCinderellaCatWindowStroke $0.90 (0.64, 1.21)$ $1.18 (0.89, 1.55)$ $1.29 (1.05, 1.60)$ $1.18 (0.89, 1.55)$ Cat $1.85 (1.56, 2.18)$ $1.56 (1.26, 1.94)$ $1.14 (0.75, 1.60)$ Window $1.72 (1.26, 2.12)$ $1.52 (1.26, 1.94)$ $1.25 (0.06, 1.82)$	Cat	1.70 (1.38, 2.04)	1.56 (1.33, 1.83)	1.13 (0.83, 1.47)				
Sandwich $2.60 (2.29, 2.87)$ $2.17 (1.94, 2.40)$ $2.26 (2.02, 2.52)$ $2.73 (2.44, 3.04)$ $2.99 (2.71, 3.26)$ Wernicke'sImportant EventStrokeCinderellaCatWindowStroke $0.90 (0.64, 1.21)$ $1.18 (0.89, 1.55)$ $1.18 (0.89, 1.55)$ Cat $1.85 (1.56, 2.18)$ $1.56 (1.26, 1.94)$ $1.14 (0.75, 1.60)$ Window $1.72 (1.26, 2.12)$ $1.52 (1.26, 2.18)$	Window	1.93 (1.56, 2.32)	1.76 (1.46, 2.11)	1.55 (1.29, 1.89)	1.02 (0.82, 1.24)			
Wernicke's       Important Event       Stroke       Cinderella       Cat       Window         Stroke       0.90 (0.64, 1.21)       0.90 (0.64, 1.21)       0.90 (0.64, 1.21)       0.90 (0.64, 1.21)         Cinderella       1.29 (1.05, 1.60)       1.18 (0.89, 1.55)       0.90 (0.64, 1.21)       0.90 (0.64, 1.21)         Window       1.72 (1.05, 1.60)       1.18 (0.89, 1.55)       0.90 (0.64, 1.21)       0.90 (0.64, 1.21)         Window       1.72 (1.26, 2.18)       1.56 (1.26, 1.94)       1.14 (0.75, 1.60)       1.25 (0.06, 1.82)	Sandwich	2.60 (2.29, 2.87)	2.17 (1.94, 2.40)	2.26 (2.02, 2.52)	2.73 (2.44, 3.04)	2.99 (2.71, 3.26)		
Wernicke's         Important Event         Stroke         Cinderella         Cat         Window           Stroke         0.90 (0.64, 1.21)								
Important Event         Stroke         Cinderella         Cat         Window           Stroke         0.90 (0.64, 1.21)	Warnisko's							
Stroke         0.90 (0.64, 1.21)         Cinderella         1.29 (1.05, 1.60)         1.18 (0.89, 1.55)           Cat         1.85 (1.56, 2.18)         1.56 (1.26, 1.94)         1.14 (0.75, 1.60)           Window         1.72 (1.26, 2.12)         1.52 (1.22, 1.82)         1.25 (0.06, 1.82)	Wernicke S	Important Event	Stroke	Cinderella	Cat	Window		
Cinderella $1.29 (1.05, 1.60)$ $1.18 (0.89, 1.55)$ Cat $1.85 (1.56, 2.18)$ $1.56 (1.26, 1.94)$ $1.14 (0.75, 1.60)$ Window $1.52 (1.26, 2.12)$ $1.52 (1.22, 1.82)$ $1.25 (0.04, 1.60)$	Stroke	0.90 (0.64, 1.21)						
Cat $1.85 (1.56, 2.18)$ $1.56 (1.26, 1.94)$ $1.14 (0.75, 1.60)$ Window $1.52 (1.26, 2.12)$ $1.52 (1.22, 1.82)$ $1.25 (0.04, 1.60)$ $1.25 (0.06, 1.82)$	Cinderella	1.29 (1.05. 1.60)	1.18 (0.80, 1.55)					
Window = 1 = 2 (1 - 26 - 2 + 12) = 1 = 2 (1 - 22 + 182) = 1 = 2 (0 - 04 + 160) = 1 = 2 (0 - 06 + 182)	Cat	1.85 (1.56, 2.18)	1.56 (1.26, 1.04)	1.14 (0.75, 1.60)				
	Window	1.72 (1.36, 2.13)	1.53(1.23, 1.82)	1.25 (0.04, 1.60)	1.35 (0.06, 1.82)			
Sandwich 2.17 (1.76, 2.66) 1.91 (1.52, 2.36) 1.82 (1.38, 2.32) 2.09 (1.61, 2.54) 2.28 (1.84, 2.74)	Sandwich	2.17 (1.76, 2.66)	1.91 (1.52, 2.36)	1.82 (1.38, 2.32)	2.09 (1.61, 2.54)	2.28 (1.84, 2.74)		

Table S1: Distances between tasks for each aphasia type. Distances in the full space, bootstrap confidence intervals based on subject-level resampling.



Figure S1: Means  $\pm$  one standard error for demographic and some linguistic variables by task and aphasia type.

*FREQ.tokens* represents the mean of the log-transformed number of tokens, so exponentiating the values gives the geometric mean of the raw number of tokens.

AQ = Aphasia Quotient extracted from the Western Aphasia Battery. A number ranging from o to 100 where 100 is no aphasia and <93.8 indicates presence of aphasia.

*Aphasia duration* = Number of years with aphasia.

*MLU.Words* = Mean length of utterance, shown in words.

*Total.Utts* = Total utterances produced.

*Sexfemale* = Proportion of females.

*Apraxia.of.SpeechY* = Presence of apraxia of speech, which was done by clinical judgment.

*DysarthriaY* = Presence of dysarthria, which was done by clinical judgment.

*Aphasia.EtiologySTR* = Proportion of participants experiencing aphasia as a result of a stroke, rather than other etiology (e.g., traumatic brain injury).