ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Characterizing Verb Argument Structure in Aphasia Using Dependency Parsers

Laura Giglio,[a,b] Julius Fridriksson,[a] and Dirk Bart den Ouden[a]

[a] Department of Communication Sciences and Disorders, Arnold School of Public Health, University of South Carolina, Columbia [b] Institute for the Interdisciplinary Study of Language Evolution, University of Zurich, Switzerland

ARTICLE INFO

Article History:
Received May 2, 2025
Revision received July 17, 2025
Accepted February 20, 2026

Editor-in-Chief: Julie A. Washington
Editor: Stephen M. Wilson

https://doi.org/10.1044/2026_JSLHR-25-00332

ABSTRACT

Purpose: Verb argument structure (VAS) is often impaired in poststroke aphasia. Previous studies investigated VAS impairments using constrained experiments or less often narrative speech, which relies on time-consuming manual annotations of VAS in speech transcriptions. Here, we aimed to develop and validate an automated approach to quantify VAS use in narrative speech using natural language processing (a dependency parser) and to apply it to a new data set.
Method: First, we validated this approach by comparing manually annotated measures of VAS from a previous study, with automatically annotated VAS measures developed here. We then applied this approach to a new data set of participants with aphasia (n = 106), whose narrative discourse samples were compared to a control group from AphasiaBank to (a) replicate previous findings and (b) assess VAS impairments in agrammatism specifically.
Results: The analysis of the new data set using automatically annotated VAS revealed that participants with Broca's aphasia and, according to grammatical distinctions, agrammatic participants were more likely to select verbs used with fewer arguments, on average, and to produce fewer arguments than controls.
Conclusions: This study reveals that dependency parsers are suitable to characterize VAS use in spontaneous speech with virtually absent manual labor, confirming the feasibility of developing clinically applicable tools for complex language analysis. The results of the study additionally show that participants with aphasia differ from controls in their VAS sensitivity and abilities.

Impairments to verb production are a prominent feature of poststroke aphasia. Verbs are associated with an argument structure, which defines the relations between the entities taking part in the event described by the verb (Giglio, Hagoort, & Ostarek, 2024; Jackendoff, 1992; Levin & Rappaport Hovav, 2005). There have been several investigations of verb and argument structure impairments in aphasia, both using constrained tasks, such as the Northwestern Assessment of Verbs and Sentences (NAVS; Cho-Reyes & Thompson, 2012), and using elicitations of narrative speech, such as through retelling of the Cinderella story (MacWhinney et al., 2011). Quantitative and qualitative characterizations of verb argument structure (VAS) in aphasia, however, require intensive manual annotations that are not always feasible in clinical practice and in research. In this study, we developed a new quantitative way of characterizing VAS for over 4,000 verbs and used this method to characterize VAS in aphasia with very limited manual input and annotations. This automated approach has the potential of being used in clinical settings as well, as VAS impairments could be assessed with automated annotations from a single discourse task.

Correspondence to Laura Giglio: laura.giglio@uzh.ch. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

VAS in Aphasia

Aphasia is an acquired language disorder. It can be caused by brain lesions following stroke, neurodegenerative diseases such as primary progressive aphasia, traumatic brain injury, tumor, or infections. In poststroke aphasia, each individual presents with a different behavioral profile, but most individuals have difficulties with noun and verb retrieval and are usually classified as being fluent or nonfluent, depending not only on speech rate but also on lexical retrieval and grammaticality of connected speech (Clough & Gordon, 2020). In addition, some persons with aphasia (PWAs) can be classified as having a specific sentence-level deficit, such as expressive agrammatism (henceforth termed *agrammatism*). Agrammatism is associated with fragmented or simple utterances, morphosyntactic errors, omission of verbs, and function words (Faroqi-Shah, 2023; Matchin et al., 2020). In spontaneous speech, patients with agrammatism are found to omit verbs and morphemes, produce inflection errors (Casilio et al., 2019; Goodglass et al., 1993), and produce errors in VAS (Malyutina et al., 2016). Participants with agrammatism often have impairments with verb retrieval, relative to noun retrieval (E. Bates et al., 1991; Miceli et al., 1988; Thompson, 2019), although verb production is difficult for all aphasia types, partly due to verbs having higher processing demands than nouns (Mätzig et al., 2009).

Verbs can vary in complexity based on their argument structure. Verbs can be used with different numbers of arguments, defined as verb valency, such as a single argument (e.g., agent in intransitive verbs such as "jump"), two arguments (e.g., agent and patient in transitive verbs such as "hit"), or three arguments (e.g., agent, theme, and recipient in verbs such as "give"). According to the "argument structure complexity" hypothesis, verbs with more arguments are more difficult to produce for speakers with agrammatism, also when they are produced in isolation (e.g., with picture naming), because the lexical–syntactic representations of verbs encode the number and type of arguments, making verb access more effortful with more arguments (Thompson, 2003; Thompson et al., 1997). Speakers with agrammatic aphasia are found to have difficulties with the production (but not comprehension) of verbs with more arguments (e.g., Kim & Thompson, 2000; Thompson et al., 2012), as well as verbs in noncanonical structures (such as unaccusatives, where a patient argument is mapped to subject role: "the vase fell," or passive sentences; Cho-Reyes & Thompson, 2012; Lee & Thompson, 2004; Thompson et al., 1997).

Verbs also differ in whether their arguments are obligatory ("Mary kicked the ball" vs. "*Mary kicked") or optional ("Mary was reading the book" vs. "Mary was reading") and in how many ways their arguments can be represented syntactically (e.g., "give" with indirect object and direct object, or direct object and prepositional object). Some verbs are particularly flexible in how many argument configurations they can be used with. For example, a verb such as "think" can be used without a direct object or with a clausal complement (e.g., "he was thinking"; "he thought that . . ."). Shapiro and Levine (1990) found that verbs with more argument configurations have greater processing demands in comprehension because all argument structure configurations are activated when accessing their lexical representations. The same effect was found in participants with agrammatic aphasia, suggesting that their lexical representations were intact. However, with respect to speech production, the evidence is less conclusive, with some studies finding no differences on argument optionality (Barbieri et al., 2019, 2024; Meltzer-Asscher et al., 2015), some studies finding more difficulties for optionally transitive verbs relative to obligatory transitive verbs (Cho-Reyes & Thompson, 2012), and other studies finding facilitatory effects for verbs with more subcategorization options (Malyutina & den Ouden, 2017; Malyutina et al., 2016; Malyutina & Zelenkova, 2020).

Most of these studies tested VAS with controlled experimental designs. In these cases, cues to a verb's argument structure are provided by a picture (with entities interacting to depict an event). An eye-tracking study found that, while participants with agrammatism have intact access to verb meaning and can integrate arguments with the verb, they are impaired in the retrieval of argument structure when the arguments are not provided (Mack et al., 2013). It is therefore also important to investigate verb production in spontaneous speech, which can provide information about (a) which verbs are available to participants without visual probes and (b) whether the argument structure is respected in speech without visual cues to necessary entities in an action. A previous study investigated VAS in narrative speech through retellings of the Cinderella story available on AphasiaBank (MacWhinney et al., 2011; Malyutina et al., 2016). Malyutina et al. (2016) found that argument structure, measured with transitivity and subcategorization options, does not hinder verb selection in spontaneous speech, but participants with Broca's aphasia were found to use those verbs less correctly in their argument structure realizations. That is, they would not only underuse the options offered by the verbs' optional argument and subcategorization structures but also violate obligatory constraints on VAS. Therefore, they argued that VAS complexity does not hinder verb access, but it affects verb use in the sentence.

Three main limitations were noted for the study by Malyutina et al. (2016). First, the characterization of the transcribed speech samples in terms of verb and VAS use was done manually. This required time-consuming work that is not applicable to clinical settings. A previous study that attempted to automate the marking of sentence complexity features using Computerized Language Analysis (CLAN) software found that automated commands for VAS measures were not comparable to manual coding from the Northwestern Narrative Language Analysis hand coding (Fromm et al., 2020). It is therefore fundamental to pursue alternative options for the automatic coding of argument structure. Second, the manual annotations followed VAS categorizations based on dictionary entries.

This assumed that both impaired and unimpaired speeches follow categorical constraints on VAS (i.e., on how many arguments are used), while in practice, argument use may be more flexible. For example, the verb "run" is usually considered to be intransitive, but it can be used transitively in many constructions (e.g., "running an analysis"). The flexibility of verb use was also noted as a problem for automated annotation with CLAN in a previous study, as it can also reduce reliability in manual annotations (Fromm et al., 2020; Hsu & Thompson, 2018). Third, the study did not explicitly characterize agrammatism but used Broca's aphasia as a proxy for agrammatism, since agrammatism most often occurs in Broca's aphasia, (but Broca's aphasia does not always include agrammatism [den Ouden et al., 2019; Goodglass, 1993]). In the current study, we therefore sought to apply a more clinically feasible quantification-based approach to the assessment of VAS use by characterizing VAS based on actual use in available corpora. In addition, we characterized VAS use in agrammatism relative to other speakers with aphasia and controls.

Automatic Extraction of VAS With Dependency Parsers

As mentioned, the analysis of VAS in spontaneous speech requires detailed annotations of speech transcripts to identify (a) all the verbs used, (b) all the arguments used, and (c) instances of missing or incorrect arguments (Malyutina et al., 2016). Malyutina et al. (2016) used a dictionary to get examples of the different ways each verb can be used and, as such, of their subcategorization options. This requires time-consuming manual annotations, which apply only to a limited set of verbs. Here, we used a dependency parser from CoreNLP (Chen & Manning, 2014; Manning et al., 2014) to identify all the verbs and arguments used by each participant without manual annotations (Lopopolo et al., 2021; Tesnière, 2015). However, we were not able to focus on instances of missing or incorrect arguments since these still require manual annotation. The use of parsers has grown in recent years to quantify incremental linguistic processing in discourse both from a receptive point of view and in spontaneous production (e.g., Brennan, 2016; Coopmans et al., 2025; Giglio, Ostarek, et al., 2024; Li & Hale, 2019). A dependency parser is also used for the creation of the grammatical tier in CLAN software for the Codes for the Human Analysis of Transcripts (CHAT) protocol, which is widely used in aphasia research (MacWhinney, 2000).

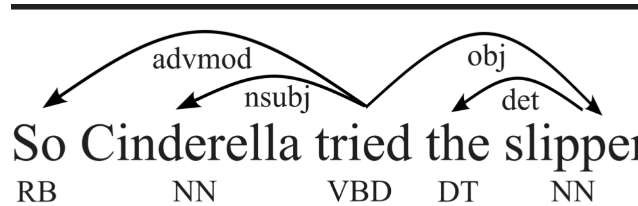
Dependency parsers organize sentences in a dependency structure of relations between pairs of words, a head and a dependent. Finite verbs are often the heads of a dependency structure, and all words are linked to other words by means of dependency relations, which correspond to grammatical functions (see Figure 1). Of most relevance here, the dependency parser can isolate verbs in a sentence and find the words that are their dependency relations. By labeling the dependency relations, it also allows for focusing on the relations that are relevant for argument structure such as "nsubj" and "obj" in the example in Figure 1, but not "advmod." We therefore took advantage of this approach to extract all verbs and their dependents/arguments from speech samples, without manual annotations. The parser is, however, not able to mark missing or incorrect arguments, so we did not focus on errors in this study.

In addition, to understand whether a verb's argument structure affects its use, verbs need to be characterized in their standard argument structure. While this is straightforward for some verbs with obligatory arguments, in many cases the use of verbs is more flexible than assumed by categorical accounts (Perek, 2015). For example, the verb "to run" is usually described as an intransitive verb that takes one argument, as when it refers to a running action carried out by a single agent. However, "run" can be used in phrasal constructions or transitively to refer to more abstract or metaphorical actions, such as "running an analysis" or "running out of salt." This flexibility is not problematic in picture descriptions, where a clear semantic use associated with a specific argument structure is depicted. For assessments of verb use in spontaneous speech, it is instead useful to characterize the variety of verb uses when assessing argument structure complexity. Therefore, we used dependency parsers to code VAS in a large corpus and get a measure of argument structure based on general usage that we could use as an index of argument structure complexity in the narrative speech production of PWAs.

The Current Study

We applied a quantitative perspective to argument structure characterization and verb use in discourse in

Figure 1. Example of the dependency structure that is identified by the dependency parser from CoreNLP for an example sentence. The verb is marked as the head of the sentence, and all other words are dependency relations attached to the verb or to their respective heads (e.g., "slipper" is head for "the"). Parts of speech: RB = adverb; NN = noun; VBD = verb, past tense; DT = determiner. Dependency relations: advmod = adverbial modifier; nsubj = noun subject; obj = object; det = determiner.



aphasia. The general goal was to reduce manual annotations to a minimum to allow for a broad and objective use of VAS characterizations, which may ultimately contribute to clinically applicable tools for the automated assessment of language samples. The automated characterization of VAS was also needed to ensure as many verbs as possible were coded, allowing for reliable measures of verb characteristics, including for infrequent verbs. The automated characterization of VAS additionally allowed for a quantitative perspective on verb use, which can more flexibly capture patterns of verb use by healthy controls, independently from rigid categorizations of grammatical use.

In addition to an automated quantification of VAS characteristics, we aimed to quantify VAS use in discourse in aphasia automatically using dependency parsers. Since dependency parsers are subject to errors and this method has not been used before, to our knowledge, we validated the approach by comparing our automatically derived measures of VAS with manual annotations. We then used these VAS characteristics in a new data set and asked whether the verb characteristics developed here and extracted from a large corpus, rather than manually, are similarly appropriate to characterize verb choice in PWAs. We focused on two analyses of interest. First, we asked whether measures of VAS complexity affect verb choice in aphasia. Malyutina et al. (2016) found that speakers with Broca's aphasia deviated from other aphasia types in their use of transitive verbs. While all other groups used more intransitive than transitive verbs, speakers with Broca's aphasia used similar numbers of transitive and intransitive verbs. Speakers with anomia and Wernicke's aphasia were instead found to use verbs with more subcategorization options more frequently than controls. The results of this analysis therefore suggested that verbs with more complex representations—be it transitivity or subcategorization options—were equally, if not more, accessible than less complex verbs for PWAs.

The second question focused on verb use. The previous results indicated that speakers with Wernicke's, Broca's, and conduction aphasia used fewer arguments than controls. Speakers with Broca's aphasia also used less diverse constructions than controls. Because of the limited feasibility of manually annotating all verbs, Malyutina et al. (2016) limited this analysis to 30 verbs with a low number of SOs and 30 verbs with a high number of SOs. By contrast, here we were able to run this analysis on all verbs used. Finally, Malyutina et al. also tested inflection accuracy, argument-type accuracy, and argument-number accuracy. We were not able to replicate these three accuracy analyses, since they required manual annotation.

We ran these analyses on a data set that included additional information on participants, compared to the previously used data set from AphasiaBank (MacWhinney et al., 2011), including whether they were agrammatic or presented with no grammatical deficits. In addition to relying on Western Aphasia Battery–Revised (WAB-R) aphasia classifications (Kertesz, 2006), as done by Malyutina et al. (2016), we clustered participants based on their grammatical profile. In fact, although VAS impairments have been hypothesized to be present in agrammatism specifically (Thompson, 2003), they have often been studied in participants with more general nonfluent aphasia or Broca's aphasia, which are then taken as proxies for grammatical problems (Gerona et al., 2024; Malyutina et al., 2016; Nedergaard et al., 2020; Zhang & Hinzen, 2022). Here, we asked whether a focus on participants with agrammatism specifically, instead of broad aphasia categorizations, could highlight different patterns of verb use that may be lost by the focus on Broca's aphasia. We additionally compared agrammatic to nonagrammatic participants and controls to have a full picture of the types of impairments across groups while controlling for severity (Faroqi-Shah, 2023). Finally, we investigated whether VAS production is affected by additional dimensions of neurogenic impairments, such as fluency, apraxia of speech, and dysarthria.

Therefore, the current study had several aims. First, we aimed to develop usage-based measures of VAS characteristics using dependency parsers, leading to the creation of a VerbBank. Second, we aimed to validate the VerbBank measures and the quantification of VAS in aphasia using dependency parsers. We compared the VerbBank measures and the argument structure measures with previously manually annotated ones from a study by Malyutina et al. (2016). Third, we aimed to take advantage of these measures to further investigate VAS in aphasia in a new data set, which included a classification for the presence of agrammatism, which was absent in the previous study. We asked whether VAS impairments were specific to agrammatism or present across grammatical and aphasia-type distinctions.

Materials and Method

Participants

Participants were recruited as part of the Predicting Outcomes of Language Rehabilitation in Aphasia study (Kristinsson et al., 2021, 2023). Participants were eligible for recruitment if they were between 21 and 80 years old, if they had chronic aphasia (> 12 months poststroke) due to left-hemisphere stroke, were speakers of English as their primary language for over 20 years, were willing and able to provide informed consent, and were able to undergo magnetic resonance imaging scanning. Participants were excluded if they had severely limited speech output as defined by the WAB-R (Spontaneous Speech score of 0–1) or severely limited auditory comprehension (WAB-R Auditory Comprehension score of 0–1). The study was carried out at the University of South Carolina and the Medical University of South Carolina, and the study procedures were approved by the institutional review board (IRB) at both universities (IRB Approval Nos. 00053559 and 00058579, respectively). Participants provided informed consent to participate in the study. Participants who scored 93.8 and higher in the WAB-R assessment ("not aphasic" by WAB-R [NABW]) were also included in this study, although they were considered control participants in the overall study and did not undergo language therapy. We included a total of 106 participants who had baseline discourse assessments (specifically Cinderella) available (see Table 1 for demographics and clinical characteristics). We additionally included control participants who did not have a stroke from AphasiaBank. We used 263 control participants available on AphasiaBank in March 2024.

Agrammatism Coding and Aphasia Covariates

Participants were coded as agrammatic by one of the authors with over 25 years of experience in aphasia and linguistics research (D.B.d.O.). Participants were classified as agrammatic if their speech in Cinderella was characterized by morphosyntactic omissions and reduced sentence complexity (see Table 2). Some participants were additionally marked as "undetermined" if their speech output was too limited or too affected by motor speech problems for a confident categorization. We included this "undetermined" group in our analysis to understand whether there was some cutoff in the verb variables we used, under which participants had severely limited speech (note that the total number of words itself may not be indicative, as the repetition of the same word sequence may result in a relatively high number of words but low input for assessment). Participants with agrammatism had lower mean length of utterance in words ($t = 5.5$, $df = 35.6$, $p < .001$) and in morphemes ($t = 5.5$, $df = 37.6$, $p < .001$), as well as fewer verbs per utterance ($t = 4.9$, $df = 24.8$, $p < .001$) than nonagrammatic participants (for averages per group, see Table 2).

During baseline assessments, participants were additionally characterized in fluency (based on the WAB-R, with Fluency scores below 5 considered as nonfluent), apraxia, and dysarthria (binary, either with or without apraxia and dysarthria, respectively). Apraxia and dysarthria were assessed using the Apraxia of Speech Rating Scale (Strand et al., 2014).

Discourse Procedure

Participants took part in baseline neuroimaging, medical, cognitive, and linguistic assessments. They then took part in speech-language treatment, which is unrelated to the purposes of the current study. Baseline linguistic assessments included discourse production. Here, we focused on the Cinderella narrative. The protocol from AphasiaBank was used for elicitation. Participants were reminded of the major events in the story by looking at a picture book without words. They were then asked to retell the events without the picture support while being recorded for later transcription. The speech samples were then transcribed by trained graduate research assistants following CHAT guidelines. Transcriptions included division in utterances, marking of fillers, repairs, repetitions, and lexical errors. Assessment of intrarater and interrater reliabilities was completed using intraclass correlation coefficients on 10% of the discourse samples and was rated good/excellent on all discourse measures.

Linguistic Annotations

VerbBank

First, we made a database of verb use from a large corpus (Corpus of Contemporary American English; see supplementary information on the OSF, VerbBank

Table 1. Participant demographics by aphasia types defined by the Western Aphasia Battery–Revised (WAB-R).

Aphasia type	*n*	Female *n* (%)	Age in years *M* (range)	WAB-R AQ *M* (range)	Agrammatic *n* (%)	AOS *n* (%)	Dysarthria *n* (%)
Control	263	139 (52.9)	55.54 (18–89)	n/a	n/a	n/a	n/a
Anomia	29	13 (44.8)	59.4 (35–76)	85.7 (71–93.1)	1 (3.4)	10 (34.5)	5 (17.2)
Broca's	33	11 (33.3)	60.6 (37–80)	49.8 (22.1–74)	13 (39.4)	24 (72.7)	11 (33.3)
Conduction	15	10 (66.6)	63 (29–78)	66 (39.8–88)	2 (13.3)	7 (46.7)	3 (20)
Global	4	1 (25)	58.7 (49–67)	27.8 (25.2–31.3)	0	3 (75)	1 (25)
NABW	17	10 (58.8)	59.6 (42–78)	97.4 (94.4–100)	0	3 (17.6)	5 (29.4)
Transcortical motor	1	0	60	78.2	0	0	0
Wernicke's	7	2 (28.6)	64.4 (57–71)	40.9 (29.8–67.8)	0	1 (14.3)	2 (28.6)

Note. AQ = Aphasia Quotient; AOS = apraxia of speech; NABW = not aphasic by WAB-R.

Table 2. Participant demographics by grammatical classification.

Grammatical classification	Agrammatic	Not agrammatic	Undetermined
n	16	58	13
Female, *n* (%)	5 (31)	29 (50)	3 (23)
Age in years, *M* (range)	55.8 (29–76)	61.7 (35–80)	61.5 (46–77)
WAB-R AQ, *M* (range)	57.5 (40.6–82.3)	70.4 (25.3–93.1)	36.8 (22.1–71.4)
Fluency, *n* (%)	3 (18.7)	45 (77.6)	2 (15.4)
AOS, *n* (%)	9 (56.3)	27 (46.6)	9 (69.2)
Dysarthria, *n* (%)	1 (6.3)	13 (22.4)	7 (53.9)
MLU_Words, *M* ± *SEM*	4.8 ± 0.4	7.5 ± 0.3	3.2 ± 0.4
MLU_Morphemes, *M* ± *SEM*	5.3 ± 0.4	8.3 ± 0.4	3.5 ± 0.4
VpU, *M* ± *SEM*	0.7 ± 0.1	1.3 ± 0.05	0.5 ± 0.1

Note. *n* = number of participants classified as fluent (Fluency), with apraxia of speech (AOS), or dysarthria; WAB-R AQ = Western Aphasia Battery–Revised Aphasia Quotient; MLU_Words = mean length of utterance in words; *SEM* = standard error of the mean; MLU_Morphemes = mean length of utterance in morphemes; VpU = verbs per utterance.

Appendix, for details). This bank was created automatically using Python (i.e., with limited manual input, all steps mentioned were scripted to be run automatically in Python or Java). By not relying on manual annotations, we could base the VerbBank on very large texts, increasing its reliability. We used CoreNLP's dependency parser (Version 4.5.6; Chen & Manning, 2014; Manning et al., 2014) to establish dependency relations as an index of argument structure. For the purposes of this analysis, we focused on dependency relations that could indicate an argument relation (see Table 3). Oblique relations are used for prepositionally marked nouns, pronouns, and noun phrases functioning as noncore arguments and adjuncts. Examples of arguments marked by obliques are the prepositional dative (e.g., "he gives the book *to the girl*") and the agent of a passive verb ("the cat was chased *by the dog*"). Although obliques often reflect an adjunct relation, we chose to include them in this first part of the analysis, since they can be arguments to some verbs (there is no way to distinguish arguments from adjuncts with the names of these dependency relations).

We scanned each sentence for the presence of verbs. For each verb, we looked for dependency relations among the ones listed in Table 3. This gave us a list of all the verbs used in the text and all their dependency relations of interest. We then summarized the information available into the following variables (see supplementary information, VerbBank Appendix, for more details):

1. Times_used: number of times each verb was used in the data set (verb frequency in the corpus)
2. MeanArgs: average number of arguments associated with each instance of verb use (a measure of verb valency that is influenced by argument optionality)
3. SO: a proxy for subcategorization options, by counting all the different patterns of argument use, for example, "he asks a question": [subj obj], "he asked her to leave": [subj iobj xcomp], "he asked her for help": [subj iobj obl], "he asked a question to her": [subj obj obl]. These are four subcategorization options for the verb "ask." These subcategorization options measure both the number of argument configurations (also called "valency frames") and the number of syntactic subcategorization options (or strict subcategorizations; cf. Malyutina & den Ouden, 2017; Shapiro & Levine, 1990).

For both MeanArgs and SO, we computed two variables: one including obliques and the other excluding obliques. We also included baseline variables for each verb: log frequency from SUBTLEX-US, concreteness (Brysbaert et al., 2014), and phoneme length (Coltheart, 1981). The correlations between all these variables can be seen in Figure 2. To illustrate, data for verbs with the highest frequency and median frequency in the VerbBank are presented in Table S1. The VerbBank is freely available to download on the Open Science Framework (https://doi.org/10.17605/OSF.IO/9ESKJ).

Argument Structure in Cinderella

We preprocessed the Cinderella narratives from all participants, which had been transcribed using CHAT guidelines (MacWhinney, 2000). Using code, we excluded all CHAT-specific transcription (e.g., error types), and we replaced errors with their target (when available) to facilitate dependency parsing, since this analysis was deliberately independent of paraphasias. We also excluded repetitions and repairs (only keeping the target) and excluded discourse markers, such as "I mean" and "you know," when they were marked as such. We then ran the dependency

Table 3. Dependency relations extracted from the output of the dependency parser to identify verb argument structure.

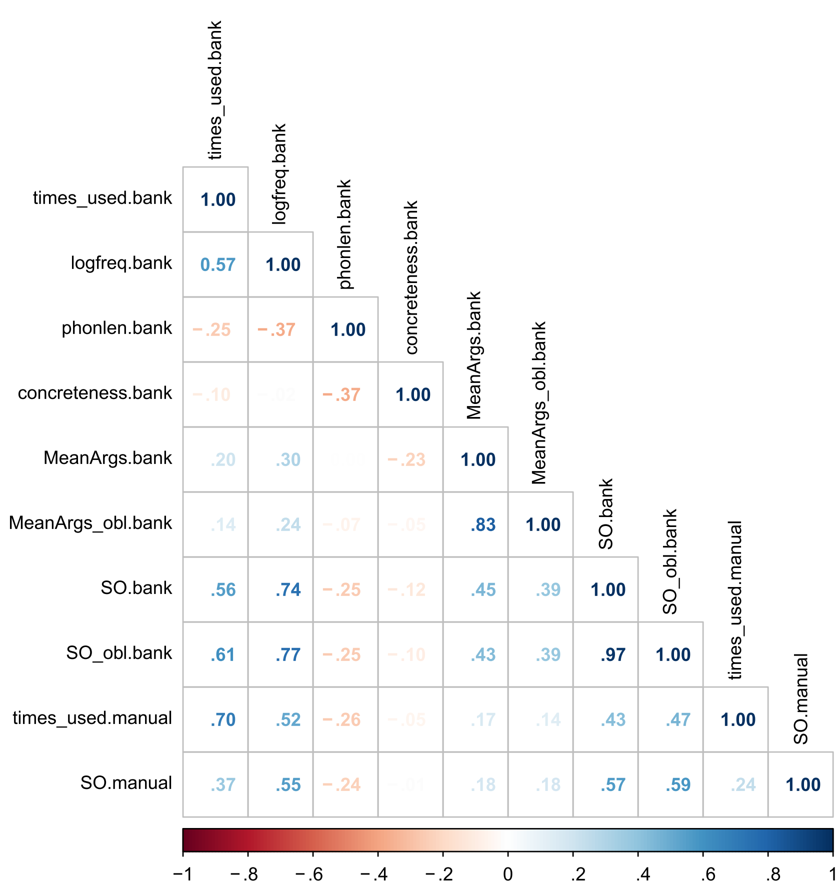
Dependency relation	Argument	Coded as	Example sentence
nsubj	Noun subject	subject	*The boy* hugs the girl.
nsubj:pass	Noun subject in a passive construction	subject	*The girl* was given a book by the boy.
nsubj:xsubj	Subject of an xcomp	subject	He wanted *her* to leave.
csubj	Clausal subject	subject	*That he was quiet* annoyed Mary.
obj	Object	object	The boy hugs *the girl*.
ccomp	Clausal complement	object clause	He thought *that she was right*.
xcomp	Open clausal complement	object clause	He wanted her *to leave*.
iobj	Indirect object	indirect object	The boy gave *her* the book.
obl	Indirect argument/adjunct	oblique	He put the book *on the table*//He read *on the couch.*

Note. The words in italics in the example sentences refer to the argument for that type of dependency relation.

parser, obtaining a list of tokens (words) and dependents for each sentence (see supplementary information, VerbBank Appendix, Dependency Parser for more details).

The approach to measure VAS was then the same as for VerbBank. For every verb identified in a sentence, we looked for dependents with one of the relations listed in Table 3 and saved which dependency relations were used for each instance of verb use, excluding auxiliaries and "be" in all cases. This gave us a list of all verbs used by each participant to narrate Cinderella, together with the arguments used in each case. For each of the verbs

Figure 2. Correlation matrix for all the verb variables extracted from the corpus using the dependency parser and baseline length, frequency and concreteness values used in the analysis (all marked with ".bank"), and values used by Malyutina et al. (2016) and annotated manually (".manual"). logfreq = log frequency; phonlen = phoneme length; MeanArgs = average number of arguments associated with each verb; obl = oblique; SO = a proxy for subcategorization options.



used, in addition to the arguments that were used in each specific case, we included the VAS characteristics obtained with VerbBank.

Approach Validation

To validate the VAS characteristics from VerbBank, as well as the ability of the parser to capture argument structure relations in discourse, we ran two analyses. First, we asked how similar the VerbBank VAS characteristics (MeanArgs and SOs) were to the original manual annotations from the previous study (Malyutina et al., 2016). Second, we tested the accuracy of the parser in isolating the number of arguments produced in spontaneous speech. For a replication of the original results with the automatically annotated VAS measures, see supplementary information, Replication Appendix (Figures S1–S3 and Tables S3–S5). In short, the automatic approach captured differences between controls and aphasia types similar to the original manually annotated results.

VerbBank Measures Validation

We compared the VerbBank MeanArgs and SO values with manual annotations available for verbs analyzed in a previous study (Malyutina et al., 2016). Correlations and distributions for automatic and manually annotated variables can be seen in Figures 2 and 3. We coded for transitivity in VerbBank by marking as intransitive verbs that were never used with obj, ccomp, or xcomp. This resulted in 24 verbs (vs. 101 in Malyutina, out of 485 verbs that had been manually annotated). Therefore, the parser was more likely to assign an object to a verb than a human coder. While this difference may indicate parser errors, it may also highlight cases where a verb is generally considered to be intransitive, while it can, in fact, also be used transitively (e.g., "run (a meeting/a race)"). Intransitive verbs according to VerbBank were nevertheless the ones used with fewer arguments (see Figure 3, small circles). Verbs coded as intransitive by Malyutina et al. (2016) were used with fewer arguments (MeanArgs), on average, than transitive verbs (see Figure 3; with obliques: transitive MeanArgs = 1.61 ± 0.015, intransitive MeanArgs = 1.42 ± 0.045, $t(24) = 6.1$, $p < .0001$; without obliques: transitive MeanArgs = 1.27 ± 0.016, intransitive MeanArgs = 1.04 ± 0.039, $t(25) = 8.5$, $p < .0001$). Therefore, verbs that were originally manually annotated as being (obligatorily) intransitive were used, on average, with fewer arguments than verbs manually annotated as (optionally) transitive.

The automatic annotation approach found more subcategorization options (ranging from 1 to 37) than the manual approach (ranging from 1 to 8). This difference was likely due to the automatic approach being fully quantitative, marking all different cases of verb use, including, for example, missing subjects (e.g., in coordinated sentences). The manual approach, instead, was categorical, defining subcategorization options based on dictionary entries, without considering virtually all ways verbs are used. While not perfectly aligned, the two subcategorization option measures were similar between the two verb banks, especially with obliques being considered. The correlation between the measures was above 0.5 (see Figure 2; with obliques, $r = .59$, $t = 16$, $p < .0001$; without obliques, $r = .57$, $t = 15$, $p < .0001$). A linear regression showed that manual SOs significantly predicted VerbBank SOs (see Figure 3; with obliques: $\beta = 2.21$, $SE = 0.14$, $p < .001$; without obliques: $\beta = 1.26$, $SE = 0.08$, $p < .001$). Therefore, while not identical, the manually annotated and dependency-annotated verb banks were broadly aligned.

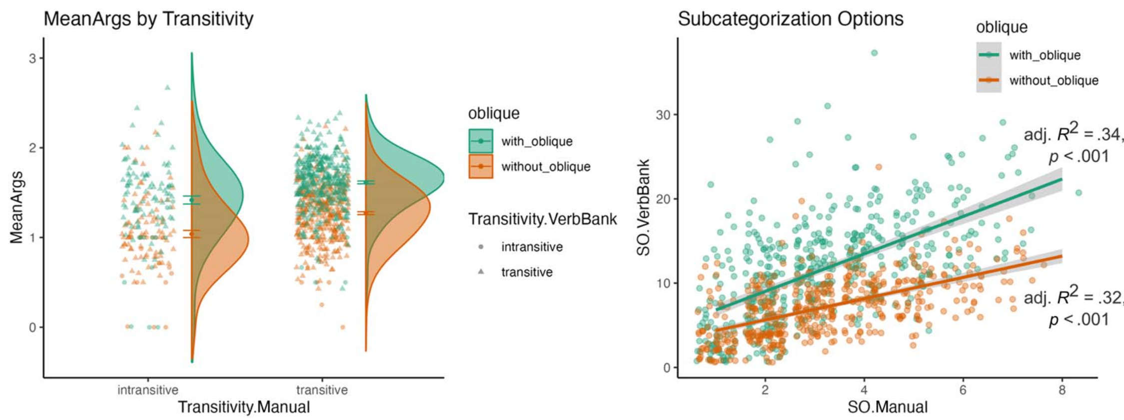
Parser Accuracy

We checked whether the parser was able to isolate the arguments produced in spontaneous speech and especially aphasic speech appropriately. We randomly extracted one verb per participant, leading to 369 verbs. The number of arguments that would be expected with a correct dependency parse for each of these sentences was manually annotated. We then ran an intraclass correlation coefficient analysis between the manually annotated number of arguments and the automatic annotation, which led to an intraclass correlation coefficient of .81 (good reliability).

Data Analysis

We ran the following analyses: verb production (assessed with verb tokens, verb types, and verb type–token ratio [TTR]), verb choice (assessed by the impact of MeanArgs and SOs on verb selection), number of arguments produced, and number of subcategorization options produced (see Table 4 for an overview of the analyses and their results). All of the analyses followed the same structure. First, participants with aphasia were compared to controls for each dependent variable of interest. In this analysis, participants with aphasia were classified according to WAB-R aphasia types (using treatment contrast coding comparing controls against each aphasia type). Second, participants classified by grammatical deficits were compared to controls (using treatment contrast coding comparing controls to "not agrammatic," "agrammatic," and "undetermined" participants, while NABW participants were excluded from this analysis). The single participant with transcortical motor aphasia was excluded from the aphasia type analysis but included in the grammatical deficit analysis. Third, for each variable of interest, an additional analysis was run on PWAs only to understand whether agrammatic participants differed from nonagrammatic participants after controlling for baseline

Figure 3. (Left panel) MeanArgs (*y*-axis, violin plot) from VerbBank for verbs coded manually either as transitive or intransitive by Malyutina et al. (2016) on the *x*-axis. Intransitive verbs are used with fewer arguments than in transitive verbs on average. The color indicates whether MeanArgs includes obliques or not. The shape of the points indicates whether each verb was intransitive or transitive according to VerbBank. (Right panel) Subcategorization options coded manually by Malyutina et al. (2016) versus automatic annotation in VerbBank. MeanArgs = average number of arguments associated with each verb; SO = a proxy for subcategorization options.



predictors. This analysis grouped participants by grammatical deficits and included aphasia severity (WAB-R Aphasia Quotient [AQ]), fluency, apraxia of speech, and dysarthria as covariates. Note that the addition of these aphasia-specific covariates was not possible in the models comparing PWAs with controls, since not all of these variables had a value for controls. Assigning an arbitrary value to controls would increase the collinearity of the models, since all covariates would align with the aphasia group classification (as controls would not vary on these covariates). For all analyses, we report the statistics for the main effects and interaction in text using *F* scores or chi-square, depending on the analysis, while specific model results and group comparisons are reported in tables in supplementary information.

Verb Use Analysis

First, we assessed general measures of verb production to understand overall verb use before focusing on argument structure. Participants were assessed on the number of verb tokens they spoke, the number of verb types they used, and the verb TTR, using linear regression in R. Verbs were extracted from the linguistic annotations of the dependency parser.

Verb Choice Analysis

Next, we determined whether verb characteristics affected the likelihood of a verb being used by participants. We ran this analysis using logistic mixed-effects models (*glmer*, from *lme4* Version 1.1-35.2; D. Bates et al., 2015). We used a logistic model, instead of the linear *lmer* model run in the previous study (Malyutina et al., 2016), because we were primarily interested in what factors drove the selection of verbs rather than how many times each verb was used (which would be modeled with linear models). We selected all verbs used by at least three participants (to facilitate convergence), leading to 255 verbs used in the analysis. We then marked each verb as being used or not used by each participant. The logistic model then predicted the probability of use of each verb based on verb concreteness, phoneme length, verb frequency, MeanArgs, SO (with obliques, normalized to help convergence), and percentage of oblique use. Model collinearity checks with variance inflation factors resulted in values below 5, indicating low collinearity between factors. We additionally included interactions between frequency, MeanArgs, SO, and obliques with the aphasia group (WAB-R aphasia type in one analysis, grammatical types in another analysis). We included by-participant and by-verb random effects. Factor significance was determined using R package *car* (Version 3.1-2; Fox et al., 2021). Post hoc comparisons were run using *emmeans* (Version 1.10-1; Lenth et al., 2022) and *interactions* (Version 1.1.5; Long, 2024). We did not run this analysis with aphasia covariates, as the model had too many convergence issues.

Argument Structure Analysis

Next, we assessed the argument structure produced by participants. We compared how many arguments were used for each verb across the aphasia groups. We counted the arguments produced for each verb (i.e., one value, number of arguments, for each verb produced by each participant), excluding obliques, since they may measure adjuncts rather than arguments, and compared use with linear mixed-effects models (*lmer* in *lme4*), predicting the number of arguments, with aphasia group, verb frequency, MeanArgs associated with each verb, concreteness, and length as fixed factors. We additionally included by-participant and by-verb random factors.

Table 4. Summary of the analyses and results reported.

Variable	Analysis	Significant effects
Verb tokens	lm(tokens ~ aphasia_type)	Anomia < controls Broca's < controls Conduction < controls Global < controls NABW < controls Wernicke's < controls
	lm(tokens ~ gramm_group)	Agramm < controls Non-agramm < controls Undetermined < controls
	lm(tokens ~ gramm_group + AQ + fluency + AOS + dysarthria)	Fluency
Verb types	lm(types ~ aphasia_type)	Anomia < controls Broca's < controls Conduction < controls Global < controls NABW < controls Wernicke's < controls
	lm(types ~ gramm_group)	Agramm < controls Non-agramm < controls Undetermined < controls
	lm(types ~ gramm_group + AQ + fluency + AOS + dysarthria)	Undetermined < non-agramm Fluency
Verb TTR	lm(TTR ~ aphasia_type)	Broca's > controls
	lm(TTR ~ gramm_group)	Undetermined > controls
	lm(TTR ~ gramm_group + AQ + fluency + AOS + dysarthria)	n.s.
Verb choice	glmer(verb_binary ~ phonlength + concreteness + logfrequency * aphasiatype + MeanArgs * aphasiatype + SOs * aphasiatype + Obl * aphasiatype + (1|lemma) + (1|s))	MeanArgs: Broca's < controls MeanArgs: global < controls SOs: global > controls Obl: Broca's < controls Obl: conduction < controls Obl: global < controls Frequency: anomia > controls Frequency: Broca's > controls Frequency: conduction > controls Frequency: Wernicke's > controls
	glmer(verb_binary ~ phonlength + concreteness + logfrequency * grammgroup + MeanArgs * grammgroup + SOs * grammgroup + Obl * grammgroup + (1 | lemma) + (1 | s))	MeanArgs: agramm < controls MeanArgs: undetermined < controls SO: undetermined > controls Obl: agramm < controls Obl: undetermined < controls Frequency: non-agramm > controls Frequency: undetermined > controls
	glmer(verb_binary ~ phonlength + concreteness + logfrequency * grammgroup + MeanArgs * grammgroup + SOs * grammgroup + Obl * grammgroup + AQ + fluency + AOS + dysarhtria + (1 | lemma) + (1 | s))	Did not converge
Number of arguments produced	lmer(n_args ~ aphasiatype + logfrequency + concrenetess + phonlength + MeanArgs + (1 | lemma) + (1 | s))	Broca's < controls Concreteness MeanArgs
	lmer(n_args ~ grammgroup + logfrequency + concrenetess + phonlength + MeanArgs + (1 | lemma) + (1 | s))	Agramm < controls Undetermined < controls Concreteness MeanArgs
	lmer(n_args ~ grammgroup + logfrequency + concrenetess + phonlength + MeanArgs + AQ + fluency + AOS + dysarthria + (1 | lemma) + (1 | s))	Agramm < non-agramm Fluency MeanArgs

(*table continues*)

Table 4. *(Continued).*

Variable	Analysis	Significant effects
SOs produced	lm(n_SO ~ aphasiatype + tokens + phonlength + logfrequency + concreteness + MeanSO)	Broca's < controls Conduction < controls Global < controls
	lm(n_SO ~ grammgroup + tokens + phonlength + logfrequency + concreteness + MeanSO)	Agramm < controls Non-agramm < controls Undetermined < controls
	lm(n_SO ~ grammgroup + tokens + phonlength + logfrequency + concreteness + MeanSO + AQ + fluency + AOS + dysarthria)	Undetermined < non-agramm

Note. For each variable of interest, the model used together with predictors is shown, as well as the significant effects, the details of which can be found in relative supplementary tables. For the first two models of each variable, the reference level is the control group; for the third model, including aphasia covariates, the reference level is the nonagrammatic group. lm = linear regression; NABW = "not aphasic" by Western Aphasia Battery–Revised; AQ = Aphasia Quotient; AOS = apraxia of speech; TTR = type–token ratio; glmer = logistic linear mixed-effects model; lmer = linear mixed-effects model; MeanArgs = average number of arguments associated with each verb; SO = a proxy for subcategorization options; obl = oblique; n.s. = not significant.

We also assessed subcategorization options produced. We counted how many different configurations of arguments were used in total across verbs by each participant, leading to one value (total number of subcategorization options produced) per participant. We then ran linear regression to predict the total number of subcategorization options produced, with aphasia group, verb frequency, length and concreteness, average number of SOs for the verbs used by each participant, and number of verbs used by each participant as predictors. We included this last variable, as the number of subcategorization options was necessarily bounded by how many verbs were used.

Results

Verb Use

The average verb tokens, types, and TTR for each aphasia type and grammatical group are presented in Figure 4 (see Supplemental Tables S6–S8 for statistical results). The analysis by aphasia types revealed that all aphasia types produced significantly fewer verb types and tokens than controls, including NABW (tokens: $F(6) = 15.7$, $p < .001$; types: $F(6) = 33.9$, $p < .001$). Verb TTR instead differed only in participants with Broca's aphasia and was marginally significant in global aphasia and NABW relative to controls, $F(6) = 2.8$, $p < .012$. Importantly, these groups had higher verb TTR than controls, confirming that higher TTR is not always indicative of "better" performance, as noted in previous work as well (Fergadiotis et al., 2013).

Grammatical group (controls, agrammatic, not agrammatic, undetermined) was a significant predictor of verb tokens, types, and TTR (tokens: $F(3) = 28.4$, $p < .001$; types, $F(3) = 62.5$, $p < .001$; TTR, $F(3) = 3.9$, $p < .009$). All aphasic groups produced overall fewer verb tokens and types than controls (*p*s < .0001). With respect to lexical diversity, only undetermined participants significantly differed from controls, showing a higher verb TTR ($p < .002$).

We ran a follow-up linear regression to understand whether AQ, fluency, apraxia, and dysarthria were significant predictors of produced verb types and tokens among the PWAs. Fluency was a significant predictor of verb types, $F(1) = 9.3$, $p < .004$, with more verb types produced by fluent participants than nonfluent participants. There was a significant difference between nonagrammatic and undetermined participants in verb types ($p = .035$). Fluency was also a significant predictor of verb tokens, $F(1) = 4.8$, $p < .03$. While agrammatic and undetermined participants numerically produced fewer verb tokens than nonagrammatic participants, this difference was not significant after controlling for confounding variables (AQ, fluency, apraxia, and dysarthria). No variables specific to aphasia significantly predicted verb TTR (except for a marginally significant effect for AQ, $F(1) = 2.9$, $p = .088$).

Verb Choice

Next, we assessed whether the choice of verbs by controls and participants with aphasia was affected by the argument structure characteristics of verbs. First, we asked whether there were differences between participants categorized by WAB-R aphasia types and controls (see Figure 5 and Table S9). We found a main effect of frequency, suggesting that more frequent verbs were more likely to be selected ($\chi^2 = 53.6$, $p < .0001$), and a main effect of SO ($\chi^2 = 4.8$, $p < .03$), also showing a general tendency to select verbs with more SOs. There were also interactions between frequency and aphasia type ($\chi^2 = 52.5$, $p < .0001$), MeanArgs and aphasia type ($\chi^2 = 25.1$, $p < .0004$), SO and aphasia type ($\chi^2 = 14.4$, $p < .026$), and obliques and aphasia type ($\chi^2 = 63.7$, $p < .0001$). Participants with anomia, Broca's

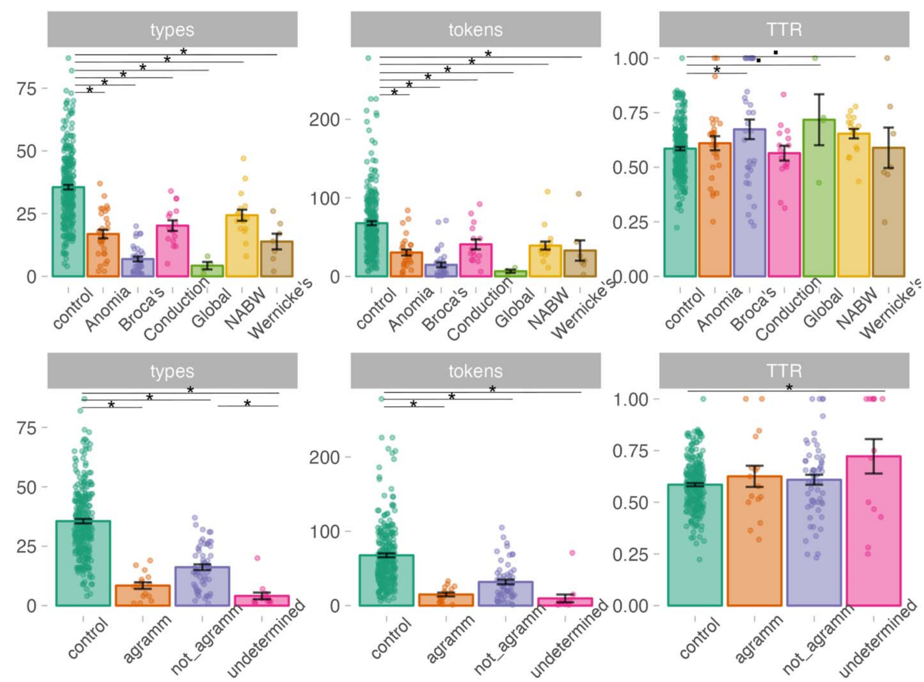


Figure 4. Verb types, tokens, and type–token ratio (TTR) for each participant (shown in individual points) and for each aphasia type (top) and grammatical group (bottom). Bars indicate group mean for variables indicated by gray bar on top, with error bars indicating standard error of the mean. In WAB-R aphasia types, comparisons were tested only between types and controls. In grammatical groups, comparisons were tested between groups and controls and separately between agrammatic (Agramm)/undetermined and nonagrammatic (Not_agramm) participants after controlling for aphasia covariates. *p < .05, ■p < .01. WAB-R = Western Aphasia Battery–Revised; NABW = "not aphasic" by WAB-R.

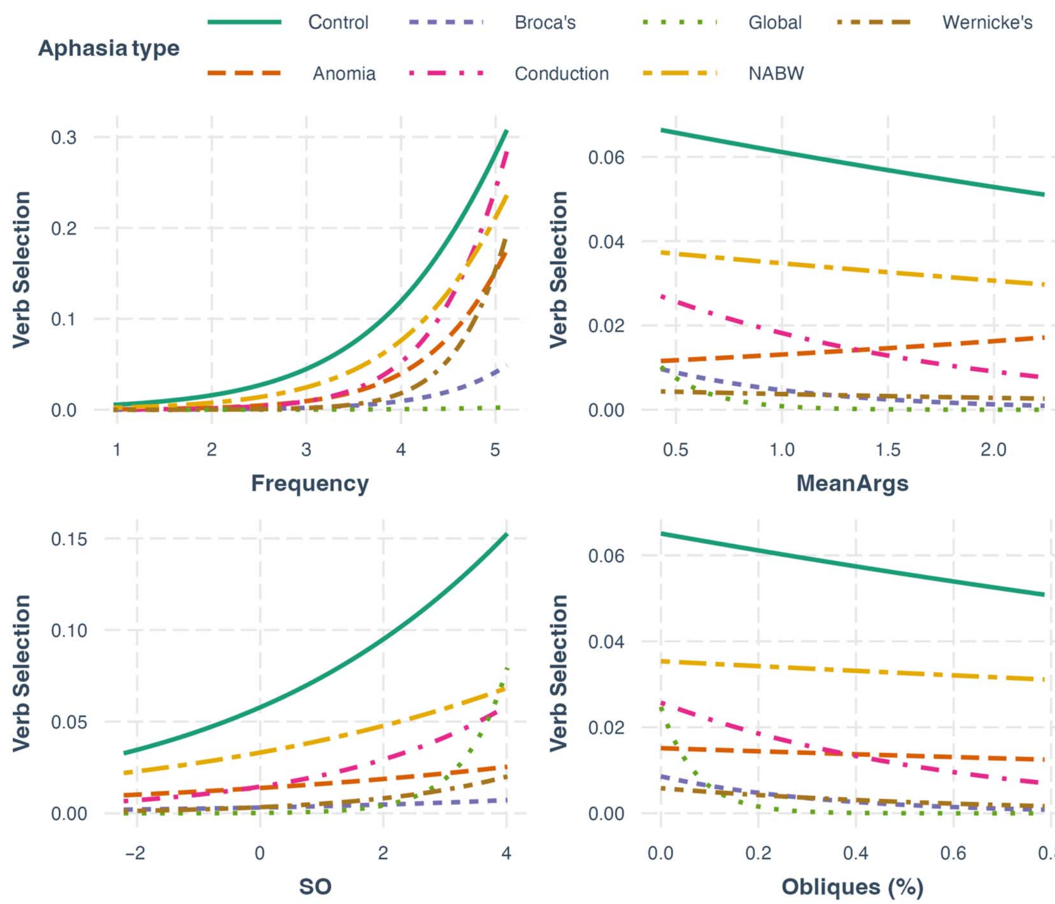
aphasia, conduction aphasia, and Wernicke's aphasia were more sensitive to verb frequency than controls, as they were more likely to select more frequent verbs than less frequent verbs relative to controls ($ps < .006$). Participants with Broca's and global aphasia were relatively more likely to select verbs with fewer arguments (MeanArgs) than controls ($ps < .004$), who were not strongly affected by MeanArgs in their verb selection. Participants with global aphasia were more sensitive to SO than controls, as they were more likely to select verbs with more SOs ($p < .002$). Finally, the percentage of oblique use for each verb was not a strong predictor of verb selection overall, but participants with Broca's, conduction, and global aphasia were more likely to select verbs that are used with fewer obliques relative to controls ($ps < .04$).

We then also investigated differences between controls and participants categorized by their grammatical deficits (i.e., agrammatic, not agrammatic, or undetermined). There was again a significant main effect of frequency ($\chi^2 = 64.8$, $p < .0001$; see Figure 6 and Table S10). There was also a significant main effect of SO ($\chi^2 = 7.7$, $p < .006$), since verbs with more subcategorization options were more likely to be selected overall. Frequency ($\chi^2 = 49.4$, $p < .0001$), MeanArgs ($\chi^2 = 30.5$, $p < .0001$), SO ($\chi^2 = 9.2$, $p < .027$), and obliques ($\chi^2 = 59.0$, $p < .0001$) also significantly interacted with grammatical group. Participants with no grammatical deficits and undetermined participants were more sensitive to frequency than controls ($ps < .003$), meaning that they had an even stronger preference for high-frequency verbs than controls, while participants with agrammatism did not significantly differ from controls. The number of MeanArgs a verb is associated with did not affect the likelihood of a verb being selected by controls, but agrammatic and undetermined participants had a stronger preference for verbs with fewer arguments over verbs with more arguments relative to controls ($ps < .001$). Undetermined participants also had a stronger preference for verbs with more SOs than controls ($p < .006$). Control participants were not strongly influenced by the usage ratio of obliques per verb, while agrammatic and undetermined participants were less likely to choose verbs often used with obliques than verbs used with fewer obliques ($ps < .001$).

Argument Structure Analysis

We asked whether participants used verb arguments in different ways (see Figure 7). We counted the number of arguments produced by each participant for each verb (excluding obliques and as a consequence prepositional

Figure 5. Slopes for the likelihood of verb selection by verb argument structure characteristics in each group. NABW = not aphasic by Western Aphasia Battery–Revised; MeanArgs = average number of arguments associated with each verb; SO = a proxy for subcategorization options.



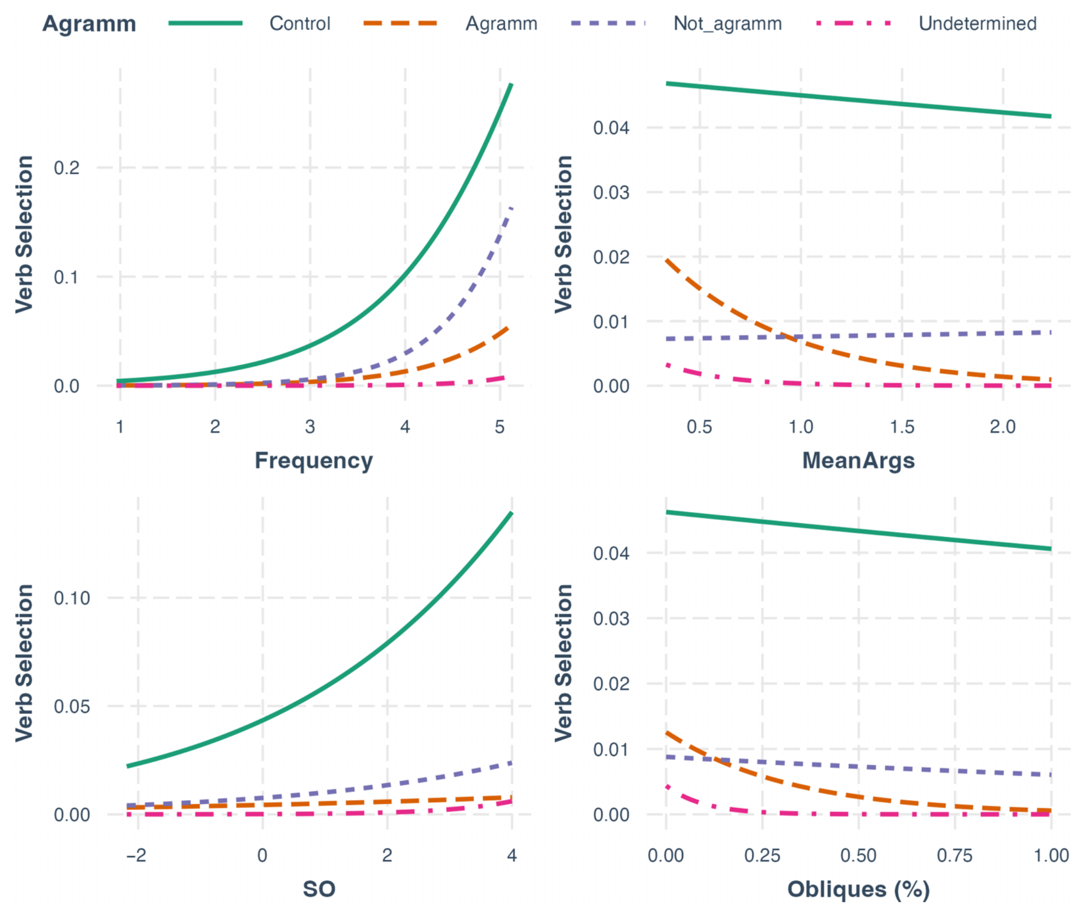
objects) and asked whether it differed by aphasia type (see Table S11). The model revealed that, unsurprisingly, the MeanArgs of a verb is a significant predictor of how many arguments are produced for each verb (χ^2 = 158.5, p < .0001). Verb concreteness was also significantly associated with the number of arguments produced (χ^2 = 9.6, p < .002), with more arguments produced with less concrete verbs. Aphasia type was also a significant predictor (χ^2 = 25.5, p < .0003), with participants with Broca's aphasia producing significantly fewer arguments than controls (p < .0001).

The analysis on grammatical groups highlighted similar effects for MeanArgs and concreteness in predicting the number of arguments produced (see Table S12) and revealed a significant effect of grammatical distinctions (χ^2 = 28.1, p < .0001), with participants with agrammatism and undetermined participants producing significantly fewer arguments than controls (p < .03). The follow-up model on participants with aphasia including covariates showed that fluency was a significant predictor of number of arguments produced (χ^2 = 5.2, p < .023; see Table S13), and grammatical group was marginally significant (χ^2 = 5.3, p < .069). Participants with agrammatism produced significantly fewer arguments than nonagrammatic participants (p = .038).

Subcategorization Options

The number of subcategorization options produced by each participant was significantly predicted by the number of tokens, $F(1)$ = 171.9, p < .0001, marginally by concreteness, $F(1)$ = 3.0, p < .085, with more SOs with higher concreteness, and aphasia type, $F(6)$ = 2.9, p < .012. Participants with Broca's, conduction, and global aphasia produced significantly fewer SOs than controls (ps < .05; see Figure 7 and Table S14). The grammatical model again showed that verb tokens were a significant predictor of SOs used (verbs: $F(1)$ = 178.1, p < .0001), as well as grammatical group, $F(3)$ = 6.9, p < .0003. All groups produced significantly fewer SOs than controls (ps < .02;

Figure 6. Slopes for the likelihood of verb selection by each grammatical group based on frequency, MeanArgs and SO, and percentage of oblique use for each verb. MeanArgs = average number of arguments associated with each verb; SO = a proxy for subcategorization options; Agramm = agrammatic participants; Not_agramm = nonagrammatic participants.



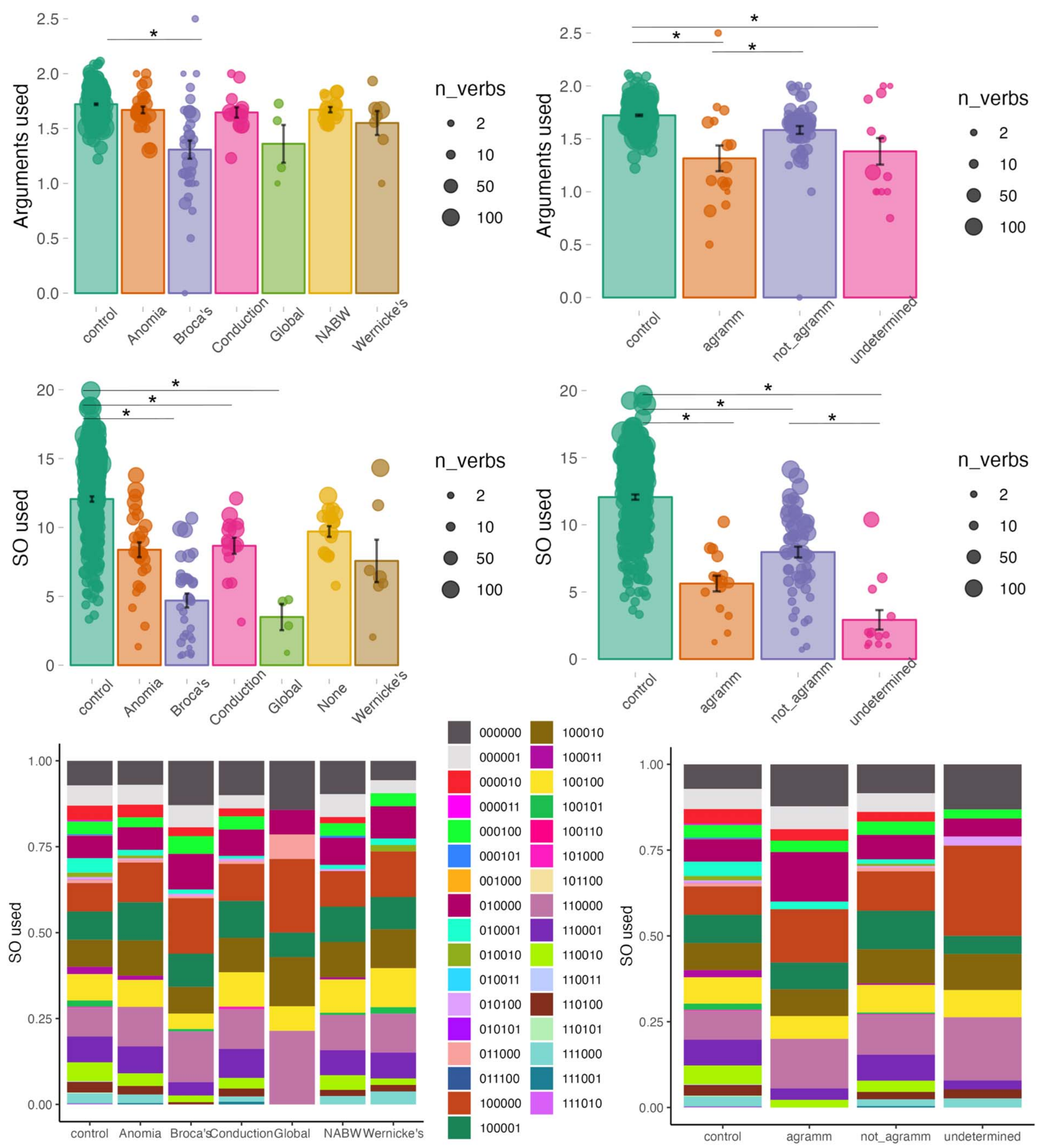
see Table S15). In a separate model with only PWAs and including covariates (Table S16), grammatical group was significant, $F(2) = 3.3$, $p = .044$, with a significant difference between undetermined and nonagrammatic participants ($p < .02$). Severity, fluency, apraxia, and dysarthria were not significantly related to the number of subcategorization options produced.

Discussion

In this study, we developed an approach to quantify VAS in discourse automatically using dependency parsers, which we validated by comparing automatic annotations of VAS with manual annotations. We then applied it to a new data set to explore VAS impairments across participants, with a specific focus on agrammatism. We found large differences between controls and PWAs in the number of verbs spoken, in the types of verbs selected, and in the number and configuration of arguments used. Differences between controls and PWAs were not specific to one aphasia type. VAS characteristics affected verb choice, especially in participants with Broca's and global aphasia or, in the analysis on agrammatism, in agrammatic and undetermined participants. The number of arguments produced was lower in Broca's aphasia or, in the analysis on agrammatism, in agrammatic and undetermined participants. Participants with Broca's, conduction, and global aphasia produced fewer subcategorization options than controls, and in the analysis on agrammatism, all groups differed from controls. The results of this study, therefore, have several implications for VAS production in aphasia and for the understanding of grammatical deficits.

VAS characteristics had an effect on verb selection, especially in nonfluent participants, as defined by the WAB-R (i.e., Broca's and global aphasia here). These participants were more likely to select verbs with fewer

Figure 7. Argument structure use. (Top) Number of arguments produced by different participants (individual participant means in single points) across aphasia groups and grammatical groups (mean and standard error of the mean presented on bar plot). (Middle) Total number of subcategorization options used by each participant (across all verbs) in individual points (sized by number of verb types produced). Group averages and standard error presented in bars, for aphasia types and grammatical groups. The number of verb tokens produced by each participant is visualized with the size of the individual points, indicating that, in most cases, outliers are due to very few verbs being produced or analyzed. In WAB-R aphasia types, comparisons were tested only between type and controls. In grammatical groups, comparisons were tested between groups and controls and separately between agrammatic (agramm)/undetermined and nonagrammatic (not_agramm) participants after controlling for aphasia covariates. *p < .05. (Bottom) SOs used by each group (in proportion). The legend indicates which arguments were used in each SO (0 for argument absent and 1 for argument used) for the following arguments: subj, obj, iobj, ccomp, xcomp, and obl (cf. Table 1 for an explanation of each argument). Notable configurations (highlighted in the plot) are 100000 (just subject), 100001 (subject and oblique), 110000 (subject and object), and 000000 (no arguments). WAB-R = Western Aphasia Battery–Revised; SO = a proxy for subcategorization options; NABW = "not aphasic" by WAB-R.



arguments and fewer obliques. A similar effect was found for agrammatism when categorizing participants by grammatical deficits. This finding fits well with the argument structure complexity hypothesis, which argues that verbs with more complex argument structure are more difficult to access (Barbieri et al., 2024; Thompson, 2003). It is also possible that nonfluent participants strategically selected verbs that require fewer arguments to reduce the amount of speech output needed, as well as working memory load (Faroqi-Shah, 2023; Fedorenko et al., 2023; Kolk & Heeschen, 1990; Kolk & Van Grunsven, 1985). In fact, participants with Broca's aphasia and agrammatic participants were also found to produce fewer arguments than controls. The elicitation of narrative speech does not force speakers to use particularly complex structures, so that representational deficits cannot be straightforwardly dissociated from processing or strategic adaptation deficits with this method.

Therefore, similar effects were found for nonfluent participants and agrammatic and undetermined participants in the two analyses. In our sample of participants, Broca's or nonfluent aphasia was shown to be an appropriate proxy for agrammatism. The only variable in which we saw differences between these categorizations of participants was in verb frequency. Participants with agrammatism did not differ from controls in their sensitivity to verb frequency, while most other groups did differ from controls. This could suggest that participants with agrammatism may not have problems with accessing verbs, since they seemed to use verbs with similar frequency to controls, but may make selections for strategic purposes. In any case, there seems to be much variability within both participant groups, whether selected by aphasia types or by grammatical deficits (as can be seen in the distribution of individual points in most figures), suggesting that Broca's aphasia and agrammatism are not homogeneous syndromes, at least with respect to VAS variables.

In this study, we also decided to include participants who were not easily categorized as either agrammatic or nonagrammatic (i.e., undetermined). These participants are often excluded in studies of grammatical deficits (e.g., den Ouden et al., 2019; Matchin et al., 2020). Due to the purely quantitative approach of this study, we preferred including them to understand if their performance better resembled agrammatic or nonagrammatic profiles. These were most often participants with very limited speech output or with much repetition and very limited variety of output, making it hard to determine whether morphological processing was correct and sentence structure was affected. Undetermined participants here were participants with severe aphasia (36 as average AQ) and categorized as having Broca's (n = 9), global (n = 2), and Wernicke's (n = 2) aphasia. The findings in both verb choice and argument structure analyses would seem to suggest that these participants are impaired in VAS in a similar way to agrammatic participants. However, it is impossible to know whether this is just a consequence of aphasia severity or a true specific impairment in verb access or knowledge of argument structure. A more specific investigation using both production and comprehension, such as the NAVS, may be able to clarify this question (cf. den Ouden et al., 2019)

It should be noted that, while both undetermined and agrammatic participants were significantly different from controls in many measures of verb use, they were not consistently different from nonagrammatic participants after controlling for baseline aphasia covariates, such as fluency and severity. In fact, agrammatic and undetermined participants had, on average, more severe aphasia than nonagrammatic participants and were more likely to be nonfluent. Undetermined participants were significantly different from nonagrammatic participants in the number of verb types and the number of subcategorization options produced, while agrammatic participants differed from nonagrammatic participants in the number of arguments produced (numerical differences were present for more measures, but they were not significant after regressing out confounding variables). Agrammatic and undetermined participants also seemed to differ from nonagrammatic participants in their sensitivity to MeanArgs for verb selection, but note that this analysis was not corrected for confounds due to model convergence issues. Therefore, it seems that participants with agrammatism are reliably different from other PWAs only in the number of arguments they use and perhaps in their verb choice. The fact that similar differences were found with undetermined participants does raise the question whether participants with agrammatism are truly affected in their VAS, relative to PWAs, or whether any differences are due to severity and production difficulties.

It is interesting to note that participants with Wernicke's aphasia did not show any clear deficits in VAS, as, for most measures, they were not different from controls. They were found to have a stronger preference for frequent verbs, which is unsurprising as more frequent verbs are more accessible, and they produced substantially fewer verbs than controls. As can be seen in most figures, however, there was substantial variability also within this group, potentially due to differences in severity. In previous work, participants with Wernicke's aphasia were found not only to be insensitive to the argument structure properties of verbs, in contrast to controls and participants with Broca's aphasia (Shapiro et al., 1993), but also to produce errors in argument structure in a similar way to participants with Broca's aphasia in Korean (Sung, 2016). Therefore, it is possible that, in the current study, these participants appeared unimpaired because dependency parsers did not allow for the assessment of accuracy in argument production.

A group that did not show as much variability was the not-aphasic group by WAB-R. This group was included as a case that is usually understudied, due to the very low aphasia severity that marks it as "not aphasic" (by WAB-R standards), even if it is recognized that such stroke survivors may have impactful language and communication problems (although less severe than other WAB-R aphasia groups), which are undetected by standardized assessments. Importantly, while this group was not impaired in VAS per se, verb production was significantly lower than among controls, suggesting that these participants are facing speech or language difficulties, even if less pronounced than the other groups. The types of verbs selected were comparable to controls, as well as the number of arguments produced. There was a trend for higher verb TTR, suggesting that they are less likely to repeat verbs than controls. Therefore, participants with AQ higher than 93.8 are able to access the same verbs as controls and use them appropriately, but language production may nevertheless constitute a more effortful process for them than for control speakers, as indicated by the lower number of verbs produced. These results mirror previous findings that highlight that this group of participants performs better than PWAs but is also significantly different from controls in many measures of connected speech (Dalton & Richardson, 2015; Fromm et al., 2017).

We now turn to how our results compare with Malyutina et al.'s (2016) findings. Malyutina et al. found that participants with Broca's aphasia had a reduced preference for intransitive verbs relative to controls. However, our results showed that, compared to controls, participants with Broca's and global aphasia had a stronger preference for verbs with fewer arguments. Rather than differences between data sets, this difference in results may be accounted for by the different analysis approaches between the two studies, as motivated in the Verb Choice section in the Method section (see also supplementary information, Replication Appendix). The logistic model in the current study focused on the likelihood of selecting verbs (and thus asks about accessibility more strictly), while the continuous linear model in the replication asked about the number of times verbs were produced (i.e., it was influenced by how often verbs were repeated). As a consequence, Malyutina et al. suggest that transitive and intransitive verbs are produced in total (and repeated) similarly by people with Broca's aphasia, but based on our results, intransitive verbs are more likely to be selected (irrespective of how many times each verb is used). Rerunning the analysis on Malyutina et al.'s data with a logistic model reveals the same pattern found in the current study (Figure S4 and Table S17). Therefore, the results of the two studies align when analyzed in the same way. Overall, the results show that verbs with fewer arguments are more likely to be selected by nonfluent and agrammatic participants, as suggested previously by studies finding a retrieval cost for verbs with more arguments (Barbieri et al., 2024; Cho-Reyes & Thompson, 2012). The number of arguments associated with a verb, therefore, either hinders verb retrieval or influences strategic choices to reduce processing load (Faroqi-Shah, 2023; Fedorenko et al., 2023; Thompson, 2003).

Limitations

Contrary to most VAS variables extracted with the dependency parser, the subcategorization options quantified by the parser did not behave as expected. Subcategorization options as developed here (both the VerbBank SO measure and the subcategorization options measured in the speech of participants analyzed here) were strongly influenced by the variability of sentence production in discourse and by sentence boundary annotations. For example, in many cases, a verb would have among its options cases of missing subject (e.g., a verb used with a direct object but without a subject). This was potentially related to the fact that the parser was trained on text. A shared subject in coordinated sentences was not always recognized as the subject of the second verb. Gerunds are also often not assigned a subject. Moreover, in CHAT coding, sentences were often split even if the subject of the second utterance was the same as the first (making it impossible for the parser to assign a subject to the verb since the parser works on sentences individually). Importantly, though, this behavior was present in both controls and PWAs. Therefore, we preferred to limit adjustments to correct for these cases here to retain a real measure of the flexibility of argument use. This approach helps us understand whether PWAs produced fewer arguments and grammatical subjects in particular by comparing them with the control baseline for how often grammatical subjects were missing (e.g., Figure 7). However, future studies may decide to deal with this differently, for example, by excluding cases in which a subject is absent (assuming a subject is always present even if not explicitly uttered, e.g., by coordination) and focusing only on the other arguments (see Table 1 for other dependency relations).

The decision to keep all of these subcategorization options, excluding any manual input, led to a confound that subcategorization options were heavily influenced by the number of times a verb was used. Subcategorization options were therefore highly correlated with verb frequency. Effectively, the high flexibility of the use of some verbs may have better reflected verb weight rather than argument structure complexity. Verb weight refers to the semantic specificity of verbs. Light verbs are very frequent verbs whose meaning often depends on the context (e.g., get, do, make). They can be used in widely varying constructions, which makes them syntactically complex. Heavy verbs, instead, are semantically more specific and are used in more stable constructions. Previous studies found that participants with worse syntactic abilities were less likely to use light verbs, while participants with semantic deficits were less likely to use heavy verbs (Barde et al., 2006; Gordon & Dell, 2003; Rezaii et al., 2024; Thorne & Faroqi-Shah, 2016). Here, we did not find a clear negative relationship between subcategorization options and verb selection in participants with agrammatism, but global and undetermined participants were more likely to select verbs with more subcategorization options, suggesting that verbs that are used in varied constructions with less specified meaning may be more accessible.

Overall, the VAS variables developed here via VerbBank were able to identify similar patterns of verb choice to those of the manual variables annotated by Malyutina et al. (2016), as well as some differences. However, all measures we developed are not transparently associated with traditional measures of argument structure. The most relevant limitation of dependency parsers and universal dependencies is that they do not mark arguments preceded by prepositions in English. MeanArgs indicates how many arguments a verb is used with, on average, but it is not able to distinguish obligatory and optional arguments. There is no consensus, however, on the relevance of argument optionality for verb complexity (Barbieri et al., 2024; Malyutina & den Ouden, 2017). In addition, SOs here seem to better reflect verb weight than more strictly intended subcategorization options, as discussed in previous studies (cf. Malyutina & den Ouden, 2017; Malyutina et al., 2016; Shapiro & Levine, 1990). It should also be noted that the measures we developed are less reliable for less frequent verbs, as we may not have appropriately captured the full dimensionality of verb use. Finally, dependency parsers are not perfect. They make errors, especially on spontaneous speech that is not as clearly organized as written text. We do not think potential errors had a large effect on this approach, since it is fully quantitative (meaning that some errors in argument count may average out). The validation of the approach also showed that any parser errors were not problematic to capture relevant patterns in the data.

Conclusions

Overall, the study presented here shows that VAS can be characterized successfully using dependency parsers without the need for manual annotations. This approach makes it possible to apply even relatively complex argument structure analyses to any speech production sample that is transcribed, which is a step forward in the development of clinically applicable automated assessments of language samples. Using this method, we showed that nonfluent participants and, more specifically, agrammatic participants are affected by VAS in their verb selection and produce fewer arguments than controls and nonagrammatic participants. VAS, therefore, is a relevant dimension on which to characterize higher level language impairments in aphasia.

Data Availability Statement

The VerbBank is openly shared on the Open Science Framework (OSF; https://doi.org/10.17605/OSF.IO/9ESKJ). The data used in the validation and the control data used in the main study are available from AphasiaBank (https://aphasia.talkbank.org/). Supplemental material, derivative data from the main study, and code to replicate the analysis are available on the OSF: https://doi.org/10.17605/OSF.IO/9ESKJ.

Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Grant P50DC014664 (principal investigator: Julius Fridriksson). The authors thank Sara Sayers for her support with data collection and organization.

References

Barbieri, E., Alessio, V., Zanobio, E., Scola, I., Luzzatti, C., & Thompson, C. K. (2024). Assessment of verb and sentence processing deficits in stroke-induced aphasia: The Italian version of the Northwestern Assessment of Verbs and Sentences (NAVS-I). *Aphasiology, 38*(3), 510–543. https://doi.org/10.1080/02687038.2023.2215494

Barbieri, E., Brambilla, I., Thompson, C. K., & Luzzatti, C. (2019). Verb and sentence processing patterns in healthy Italian participants: Insight from the Northwestern Assessment of Verbs and Sentences (NAVS). *Journal of Communication Disorders, 79,* 58–75. https://doi.org/10.1016/j.jcomdis.2019.03.001

Barde, L. H. F., Schwartz, M. F., & Boronat, C. B. (2006). Semantic weight and verb retrieval in aphasia. *Brain and Language, 97*(3), 266–278. https://doi.org/10.1016/j.bandl.2005.11.002

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, E., Wulfeck, B., & MacWhinney, B. (1991). Cross-linguistic research in aphasia: An overview. *Brain and Language, 41*(2), 123–148. https://doi.org/10.1016/0093-934X(91)90149-U

Brennan, J. R. (2016). Naturalistic sentence comprehension in the brain. *Language and Linguistics Compass, 10*(7), 299–313. https://doi.org/10.1111/lnc3.12198

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Casilio, M., Rising, K., Beeson, P. M., Bunton, K., & Wilson, S. M. (2019). Auditory-perceptual rating of connected speech in aphasia. *American Journal of Speech-Language Pathology, 28*(2), 550–568. https://doi.org/10.1044/2018_AJSLP-18-0192

Chen, D., & Manning, C. (2014). A fast and accurate dependency parser using neural networks. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 740–750). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1082

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology, 26*(10), 1250–1277. https://doi.org/10.1080/02687038.2012.693584

Clough, S., & Gordon, J. K. (2020). Fluent or nonfluent? Part A. Underlying contributors to categorical classifications of fluency in aphasia. *Aphasiology, 34*(5), 515–539. https://doi.org/10.1080/02687038.2020.1727709

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 33A*(4), 497–505. https://doi.org/10.1080/14640748108400805

Coopmans, C. W., de Hoop, H., Tezcan, F., Hagoort, P., & Martin, A. E. (2025). Language-specific neural dynamics extend syntax into the time domain. *PLoS Biology, 23*(1), Article e3002968. https://doi.org/10.1371/journal.pbio.3002968

Dalton, S. G., & Richardson, J. D. (2015). Core-lexicon and main-concept production during picture-sequence description in adults without brain damage and adults with aphasia. *American Journal of Speech-Language Pathology, 24*(4), S923–S938. https://doi.org/10.1044/2015_AJSLP-14-0161

den Ouden, D.-B., Malyutina, S., Basilakos, A., Bonilha, L., Gleichgerrcht, E., Yourganov, G., Hillis, A. E., Hickok, G., Rorden, C., & Fridriksson, J. (2019). Cortical and structural-connectivity damage correlated with impaired syntactic processing in aphasia. *Human Brain Mapping, 40*(7), 2153–2173. https://doi.org/10.1002/hbm.24514

Faroqi-Shah, Y. (2023). A reconceptualization of sentence production in post-stroke agrammatic aphasia: The synergistic processing bottleneck model. *Frontiers in Language Sciences, 2,* Article 1118739. https://doi.org/10.3389/flang.2023.1118739

Fedorenko, E., Ryskin, R., & Gibson, E. (2023). Agrammatic output in non-fluent, including Broca's, aphasia as a rational behavior. *Aphasiology, 37*(12), 1981–2000. https://doi.org/10.1080/02687038.2022.2143233

Fergadiotis, G., Wright, H. H., & West, T. M. (2013). Measuring lexical diversity in narrative discourse of people with aphasia. *American Journal of Speech-Language Pathology, 22*(2), S397–S408. https://doi.org/10.1044/1058-0360(2013/12-0083)

Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., Bolker, B., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., Graves, S., Heiberger, R., Krivitsky, P., Laboissiere, R., Maechler, M., Monette, G., Murdoch, D., Nilsson, H., . . . R-Core. (2021). *car: Companion to Applied Regression* (Version 3.0-12) [Computer software]. https://CRAN.R-project.org/package=car

Fromm, D., Forbes, M., Holland, A., Dalton, S. G., Richardson, J., & MacWhinney, B. (2017). Discourse characteristics in aphasia beyond the Western Aphasia Battery cutoff. *American Journal of Speech-Language Pathology, 26*(3), 762–768. https://doi.org/10.1044/2016_AJSLP-16-0071

Fromm, D., MacWhinney, B., & Thompson, C. K. (2020). Automation of the Northwestern Narrative Language Analysis system. *Journal of Speech, Language, and Hearing Research, 63*(6), 1835–1844. https://doi.org/10.1044/2020_JSLHR-19-00267

Gerona, J., de Kok, D., Salis, C., Webster, J., & Jonkers, R. (2024). Characterization of agrammatism in Tagalog: Evidence from narrative spontaneous speech. *Aphasiology, 39*(3), 385–417. https://doi.org/10.1080/02687038.2024.2353164

Giglio, L., Hagoort, P., & Ostarek, M. (2024). Neural encoding of semantic structures during sentence production. *Cerebral Cortex, 34*(12), Article bhae482. https://doi.org/10.1093/cercor/bhae482

Giglio, L., Ostarek, M., Sharoh, D., & Hagoort, P. (2024). Diverging neural dynamics for syntactic structure building in naturalistic speaking and listening. *Proceedings of the National Academy of Sciences, 121*(11), Article e2310766121. https://doi.org/10.1073/pnas.2310766121

Goodglass, H. (1993). *Understanding aphasia* (pp. xii, 297). Academic Press.

Goodglass, H., Christiansen, J. A., & Gallagher, R. (1993). Comparison of morphology and syntax in free narrative and structured tests: Fluent vs. nonfluent aphasics. *Cortex, 29*(3), 377–407. https://doi.org/10.1016/S0010-9452(13)80250-X

Gordon, J. K., & Dell, G. S. (2003). Learning to divide the labor: An account of deficits in light and heavy verb production. *Cognitive Science, 27*(1), 1–40. https://doi.org/10.1207/s15516709cog2701_1

Hsu, C.-J., & Thompson, C. K. (2018). Manual versus automated narrative analysis of agrammatic production patterns: The Northwestern Narrative Language Analysis and computerized language analysis. *Journal of Speech, Language, and Hearing Research, 61*(2), 373–385. https://doi.org/10.1044/2017_JSLHR-L-17-0185

Jackendoff, R. S. (1992). *Semantic structures* (Vol. 18). MIT Press.

Kertesz, A. (2006). *Western Aphasia Battery–Revised (WAB-R)* [Data set]. https://doi.org/10.1037/t15168-000

Kim, M., & Thompson, C. K. (2000). Patterns of comprehension and production of nouns and verbs in agrammatism: Implications for lexical organization. *Brain and Language, 74*(1), 1–25. https://doi.org/10.1006/brln.2000.2315

Kolk, H., & Heeschen, C. (1990). Adaptation symptoms and impairment symptoms in Broca's aphasia. *Aphasiology, 4*(3), 221–231. https://doi.org/10.1080/02687039008249075

Kolk, H., & Van Grunsven, M. M. F. (1985). Agrammatism as a variable phenomenon. *Cognitive Neuropsychology, 2*(4), 347–384. https://doi.org/10.1080/02643298508252666

Kristinsson, S., Basilakos, A., den, O. D. B., Cassarly, C., Spell, L. A., Bonilha, L., Rorden, C., Hillis, A. E., Hickok, G., Johnson, L., Busby, N., Walker, G. M., McLain, A., & Fridriksson, J. (2023). Predicting outcomes of language rehabilitation: Prognostic factors for immediate and long-term outcomes after aphasia therapy. *Journal of Speech, Language, and Hearing Research, 66*(3), 1068–1084. https://doi.org/10.1044/2022_JSLHR-22-00347

Kristinsson, S., Basilakos, A., Elm, J., Spell, L. A., Bonilha, L., Rorden, C., den Ouden, D. B., Cassarly, C., Sen, S., Hillis, A., Hickok, G., & Fridriksson, J. (2021). Individualized response to semantic versus phonological aphasia therapies in stroke. *Brain Communications, 3*(3), Article fcab174. https://doi.org/10.1093/braincomms/fcab174

Lee, M., & Thompson, C. K. (2004). Agrammatic aphasic production and comprehension of unaccusative verbs in sentence contexts. *Journal of Neurolinguistics, 17*(4), 315–330. https://doi.org/10.1016/S0911-6044(03)00062-9

Lenth, R. V., Buerkner, P., Herve, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2022). *emmeans: Estimated marginal means, aka least-squares means* (Version 1.7.3) [Computer software]. https://CRAN.R-project.org/package=emmeans

Levin, B., & Rappaport Hovav, M. R. (2005). *Argument realization*. Cambridge University Press. https://doi.org/10.1017/CBO9780511610479

Li, J., & Hale, J. (2019). Grammatical predictors for fMRI time-courses. In R. C. Berwick & E. P. Stabler (Eds.), *Minimalist parsing* (pp. 159–173). Oxford University Press. https://doi.org/10.1093/oso/9780198795087.003.0007

Long, J. A. (2024). *interactions: Comprehensive, user-friendly toolkit for probing interactions* (Version 1.2.0) [Computer software]. https://cran.r-project.org/web/packages/interactions/index.html

Lopopolo, A., van den Bosch, A., Petersson, K.-M., & Willems, R. M. (2021). Distinguishing syntactic operations in the brain: Dependency and phrase-structure parsing. *Neurobiology of Language, 2*(1), 152–175. https://doi.org/10.1162/nol_a_00029

Mack, J. E., Ji, W., & Thompson, C. K. (2013). Effects of verb meaning on lexical integration in agrammatic aphasia: Evidence from eyetracking. *Journal of Neurolinguistics, 26*(6), 619–636. https://doi.org/10.1016/j.jneuroling.2013.04.002

MacWhinney, B. (2000). *The CHILDES Project: The database* (Vol. 2). Psychology Press. https://books.google.com/books?hl=it&lr=&id=zxN648YXqHYC&oi=fnd&pg=PA1&dq=macwhinney+2000&ots=YRjU8k8KFP&sig=xHOjD702yPwW749g9LUqEH_sGYE

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

Malyutina, S., & den Ouden, D.-B. (2017). Task-dependent neural and behavioral effects of verb argument structure features. *Brain and Language, 168,* 57–72. https://doi.org/10.1016/j.bandl.2017.01.006

Malyutina, S., Richardson, J. D., & den Ouden, D. B. (2016). Verb argument structure in narrative speech: Mining AphasiaBank. *Seminars in Speech and Language, 37*(1), 34–47. https://doi.org/10.1055/s-0036-1572383

Malyutina, S., & Zelenkova, V. (2020). Verb argument structure effects in aphasia are different at single-word versus sentence level. *Aphasiology, 34*(4), 431–457. https://doi.org/10.1080/02687038.2019.1697866

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In K. Bontcheva & J. Zhu (Eds.), *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 55–60). Association for Computational Linguistics. https://doi.org/10.3115/v1/P14-5010

Matchin, W., Basilakos, A., Stark, B. C., den Ouden, D.-B., Fridriksson, J., & Hickok, G. (2020). Agrammatism and paragrammatism: A cortical double dissociation revealed by lesion-symptom mapping. *Neurobiology of Language, 1*(2), 208–225. https://doi.org/10.1162/nol_a_00010

Mätzig, S., Druks, J., Masterson, J., & Vigliocco, G. (2009). Noun and verb differences in picture naming: Past studies and new evidence. *Cortex, 45*(6), 738–758. https://doi.org/10.1016/j.cortex.2008.10.003

Meltzer-Asscher, A., Mack, J. E., Barbieri, E., & Thompson, C. K. (2015). How the brain processes different dimensions of argument structure complexity: Evidence from fMRI. *Brain and Language, 142,* 65–75. https://doi.org/10.1016/j.bandl.2014.12.005

Miceli, G., Silveri, M. C., Nocentini, U., & Caramazza, A. (1988). Patterns of dissociation in comprehension and production of nouns and verbs. *Aphasiology, 2*(3–4), 351–358. https://doi.org/10.1080/02687038808248937

Nedergaard, J. S. K., Martínez-Ferreiro, S., Fortescue, M. D., & Boye, K. (2020). Non-fluent aphasia in a polysynthetic language: Five case studies. *Aphasiology, 34*(6), 675–694. https://doi.org/10.1080/02687038.2019.1643000

Perek, F. (2015). *Argument structure in usage-based construction grammar: Experimental and corpus-based perspectives*. John Benjamins. https://books.google.com/books?hl=en&lr=&id=3x_IBwAAQBAJ&oi=fnd&pg=PR1&ots=b8ZuJ8r-mf&sig=1h6l01UHIsYfhRtXHYHatOviCEM#v=onepage&q&f=false

Rezaii, N., Hochberg, D., Quimby, M., Wong, B., Brickhouse, M., Touroutoglou, A., Dickerson, B. C., & Wolff, P. (2024). Artificial intelligence classifies primary progressive aphasia from connected speech. *Brain, 147*(9), 3070–3082. https://doi.org/10.1093/brain/awae196

Shapiro, L. P., Gordon, B., Hack, N., & Killackey, J. (1993). Verb-argument structure processing in complex sentences in Broca's and Wernicke's aphasia. *Brain and Language, 45*(3), 423–447. https://doi.org/10.1006/brln.1993.1053

Shapiro, L. P., & Levine, B. A. (1990). Verb processing during sentence comprehension in aphasia. *Brain and Language, 38*(1), 21–47. https://doi.org/10.1016/0093-934X(90)90100-U

Strand, E. A., Duffy, J. R., Clark, H. M., & Josephs, K. (2014). The Apraxia of Speech Rating Scale: A tool for diagnosis and description of apraxia of speech. *Journal of Communication Disorders, 51,* 43–50. https://doi.org/10.1016/j.jcomdis.2014.06.008

Sung, J. E. (2016). The effects of verb argument complexity on verb production in persons with aphasia: Evidence from a subject–object–verb language. *Journal of Psycholinguistic Research, 45*(2), 287–305. https://doi.org/10.1007/s10936-014-9346-y

Tesnière, L. (2015). *Elements of structural syntax*. John Benjamins. https://doi.org/10.1075/z.185

Thompson, C. K. (2019). Neurocognitive recovery of sentence processing in aphasia. *Journal of Speech, Language, and Hearing Research, 62*(11), 3947–3972. https://doi.org/10.1044/2019_JSLHR-L-RSNP-19-0219

Thompson, C. K. (2003). Unaccusative verb production in agrammatic aphasia: The argument structure complexity hypothesis. *Journal of Neurolinguistics, 16*(2–3), 151–167. https://doi.org/10.1016/S0911-6044(02)00014-3

Thompson, C. K., Lange, K. L., Schneider, S. L., & Shapiro, L. P. (1997). Agrammatic and non–brain-damaged subjects' verb and verb argument structure production. *Aphasiology, 11*(4–5), 473–490. https://doi.org/10.1080/02687039708248485

Thompson, C. K., Lukic, S., King, M. C., Mesulam, M. M., & Weintraub, S. (2012). Verb and noun deficits in stroke-induced and primary progressive aphasia: The Northwestern Naming Battery. *Aphasiology, 26*(5), 632–655. https://doi.org/10.1080/02687038.2012.676852

Thorne, J., & Faroqi-Shah, Y. (2016). Verb production in aphasia: Testing the division of labor between syntax and semantics. *Seminars in Speech and Language, 37*(1), 23–033. https://doi.org/10.1055/s-0036-1571356

Zhang, H., & Hinzen, W. (2022). Grammar in 'agrammatical' aphasia: What's intact? *PLOS ONE, 17*(12), Article e0278676. https://doi.org/10.1371/journal.pone.0278676