An objective coding scheme for grammatical production deficits in aphasia reveals a categorical divide between agrammatism and paragrammatism

Danielle Fahey,[*ǂ1] Jeremy Yeaton,[ǂ2] Brielle Stark,[3] Dirk-Bart den Ouden,[4] Gregory Hickok,[2] Julius Fridriksson,[4] & William Matchin[4]

* Indicates corresponding author
For correspondence:
32 Campus Cir.
Missoula, MT 59808
Danielle.fahey@umontana.edu

ǂ Indicates joint 1st author
1 University of Montana
2 University of California, Irvine
3 Indiana University Bloomington
4 University of South Carolina

Funding: This work was supported by a grant from the National Institute on Deafness and Other Communication Disorders (P50 DC014664) and start-up funds awarded to Danielle Fahey by the University of Montana.

Abstract
Syntactic impairments in aphasia can provide a powerful window into the neurobiology of language. Considerable research has focused on agrammatism in nonfluent aphasia, driving a strong association between frontal brain systems and syntax. However, the syndrome of paragrammatism, typically characterized by grammatical errors in fluent aphasia, has recieved far less attention. Recent work has suggested that paragrammatism is primarily associated with posterior temporal-parietal lesions, converging with functional neuroimaging evidence that these regions support hierarchical syntax. However, the holistic perceptual approaches to paragrammatism used in this work suffer from limited inter-rater reliability as well as conflating factors such as speech rate. To remedy these issues, this study reports a system to classify agrammatic and paragrammatic symptoms at the level of the utterance using an objective coding scheme, building on previous analyses by Matchin et al. (2020). Trained speech-pathology students rated transcriptions of 88 retellings of the story of Cinderella from persons with aphasia alongside transcriptions from 53 age-matched healthy controls. Each utterance was classified by the presence of errors corresponding to functional processing (i.e., hierarchical processing) or positional processing (i.e., linearization), along with non-grammatical errors. We found that patients defined holistically using the perceptual approach as agrammatic or paragrammatic doubly dissociated in rates of functional and positional processing, and that qualitative behavioral variables distinguish healthy speech from agrammatic and paragrammatic samples. We suggest that agrammatism and paragrammatism result from distinct breakdowns in syntactic structure building during speech, resulting from damage to distinct syntactic subsystems of the brain.

Keywords: discourse, aphasia, agrammatism, paragrammatism, syntax,

Introduction

Syntax – the human ability to generate novel, hierarchically structured sequences of words – is potentially a uniquely human trait, arising from novel brain structures (Friederici, 2017; Berwick & Chomsky, 2016). However, the neurobiology of syntax is the source of much controversy (Grodzinsky & Santi, 2008; Fedorenko & Kanwisher, 2009; Fedorenko et al., 2020; Matchin & Hickok, 2020). Syntactic or grammatical deficits in clinical populations are a powerful potential window into the neurobiology of syntax, allowing examination into the causal brain structures which give rise to this ability (Rorden & Karnath, 2004). In particular, grammatical deficits in post-stroke aphasia are highly prevalent (Caplan et al., 2007; Goodglass, 1993, 1997; Iorga et al., 2021; Thompson, 2019; Turkstra et al., 2010; Ullman et al., 2005) and have been identified for nearly as long as aphasia has been studied. Remarkably, perhaps, they remain poorly understood, limiting the inferences that can be gleaned from these syndromes about syntactic mechanisms in the brain. The study described here aims to provide deeper insights into the syntactic mechanisms which, when impaired, produce grammatical deficits in aphasia, which can be used in turn for produce lesion-deficit correlation studies.

Two terms are currently in use to describe productive grammatical deficits in aphasia: agrammatism and paragrammatism. Agrammatism, a term generally describing grammatical deficits resulting in loss of (primarily) productive syntax and/or morphosyntax, was coined by Adolf Kussmaul in 1877. Paragrammatism, a term generally describing grammatical deficits resulting in the ungrammatical production of syntax and/or morphosyntax, was described in 1914 by psychiatrist and neurologist Karl Kleist. Though these terms are still in use, there is great debate about the characteristics of disordered spoken language that are used to identify these syndromes, the nature of the underlying linguistic deficits that produce them, and their lesion correlates.

Grammatical deficits in lexical syntax, linear ordering, morphosyntax, and hierarchical structuring have all been described in aphasia patients in different contexts (Blumstein et al., 1980; Cho-Reyes & Thompson, 2012; Fahey et al., 2024; Gleason et al., 1975; Goodglass et al., 1993; Goodglass & Hunt, 1958; Grodzinsky, 2000; Matchin et al., 2020; Thompson et al., 1997; Vigliocco et al., 2011). However, not all of these grammatical features are addressed specifically across current descriptions of agrammatism or paragrammatism. Morphosyntactic omissions have been recognized in characterizations of grammatical deficits in aphasia since the term 'agrammatism' was coined. Errors from morphosyntactic insertions are described by some researchers as agrammatism while others describe them as reflecting paragrammatism. Deficits in grammatical structuring typify paragrammatism, but the nature of the syntactic malformation has not been spelled out. Patients with paragrammatism produce 'sentence monsters' (Kleist's term) that juxtapose phrases and clauses ungrammatically. Patients with agrammatism often produce reduced structures, although this is not universally considered definitional of the syndrome. Errors in lexical syntax do not currently define either agrammatism or paragrammatism, although a reduction in verbs is said to co-occur in agrammatism, while use of words in incorrect positions is said to co-occur in paragrammatism (Faroqi-Shah, 2013; Vigliocco et al., 2011). Inconsistent definitions of the grammatical syndromes may have contributed to lack of understanding of their nature.

Though characterizations of these syndromes remain ambiguous, they may best be understood in contrast to one another. Traditionally, definitions of (par)agrammatism have rested upon a distinction between omission and disordered use of grammatical structures, but distinctions of functional level of deficit and between comprehension and production deficits have also influenced (par)agrammatism definitions such that agrammatism is seen as relating to deficits through omission whereas paragrammatism is seen as relating to deficits through ungrammatical insertions (Stockbridge et al., 2020, 2021; Wilson et al., 2012). In addition, non-fluency[1] is still associated with agrammatic speech while fluency is associated with paragrammatic speech, even though non-fluency or fluency do not fully predict the presence or absence of (par)agrammatism. In other words, descriptions of these syndromes have remained inconsistent since their first characterizations. It is worth noting that these definitions refer to productive speech. There is currently no widely used term to describe deficits in grammatical comprehension, though deficits to grammatical comprehension have been described both separately and in conjunction with other syndromes (Caramazza & Zurif, 1976; Grodzinsky, 2000; Magnusdottir et al., 2013). The term "agrammatic comprehension" exists, however its use is problematic as many in the field tightly associate "agrammatism" with being an expressive deficit ((i.e.: overarching agrammatism; Matchin et al., 2023); referring to comprehension as agrammatic has the potential to confound discussion of expressive and receptive deficits and the potential for their dissociation.

The characterization of patients' speech as agrammatic or paragrammatic is typically completed informally using discourse samples (for which there is no standard metric) (see Bird & Franklin, 1996) or using a formal assessment, such as the Boston Diagnostic Aphasia Examination (BDAE) grammatical form measure (Goodglass, Kaplan, & Barresi, 2000), the Comprehensive Aphasia Test (CAT; Swinburn et al., 2004), the Northwestern Assessment of Verbs and Sentences (NAVS; Cho-Reyes & Thompson, 2012), and possibly subtests of the Psycholinguistic assessments of language processing in aphasia (PALPA; Kay et al., 1996), but these formal measures also lack clear diagnostic cutoffs. The traditional Kleist (1914) criteria for agrammatism and paragrammatism are neither sufficiently explicit nor reliant on theoretical positions for the

[1] Even the term "fluency" lacks consistent definitions across the literature, making its utility in the context of syntactic disorders unclear.

underlying causes of the syndromes for the creation of a modern evaluative tool for the classification of (par)agrammatism.

Recent work by Matchin et al. (2020) found a double dissociation of lesion location between patients categorized as agrammatic or paragrammatic. Casilio et al. (2024) recently reported a similar double dissociation between their perceptual-derived factors of 'Agrammatism' and 'Paraphasia', the latter of which included perceptual ratings of paragrammatic speech. The double dissociation notwithstanding, Matchin et al. reported lower interrater reliability for categorizations of paragrammatism than of agrammatism. They note that this may be due to their reliance on perceptual ratings of paragrammatism. The researchers in that study anlayzed the entire speech sample of each patient holistically, applying their categorization at the patient level, rather than analyzing each utterance separately. This was in part because they had not developed clear diagnostic criteria for paragrammatism at the individual utterance level, and they were concerned that some instances of functional element misuse would be incorrectly classified as agrammatism when taken in isolation from the rest of the patient's speech sample (p. 221).

Not only are definitions of agrammatism and paragrammatism inconsistent, they have not been closely tied to psycholinguistic models of language processing. In this work, we developed a novel classification of grammatical deficits which acts at the level of the utterance which draws inspiration from psycholinguistic models of sentence production. We will refer to this as the syntactic utterance-based analysis (SUBA) method. Rather than using traditional definitions of agrammatism or paragrammatism, we describe deficits in terms of underlying psycholinguistic components such as lexical syntax, linear ordering, morphosyntax, and hierarchical structuring, predicted to be affected in the speech of persons with aphasia (PWA). In this work, we are guided by the "consensus" two-stage psycholiguistic model of syntactic encoding proposed by by Bock and Levelt (1994; see also Ferreira & Slevc, 2007). Under the Bock & Levelt model, grammatical encoding occurs in two stages: *functional processing* and *positional processing*. Functional processing supports lemma processing, functional assignment, and hierarchical organization while positional processing underlies linear morphosyntactic processing. Following the logical conclusions of Matchin and Hickok (2020), we hypothesize that agrammatic deficits tend to align with errors reflecting *positional processing* and paragrammatic deficits tend to align to with errors reflecting *functional processing.*

Research Questions

The goals of this work are twofold: 1) to develop a modern evaluation tool for (par)agrammatism grounded in a well-established psycholinguistic framework, and 2) validate it against healthy controls as well as existing perceptual classification data in aphasia. To do so, we characterized isolated speech samples from transcriptions. This utterance-level and transcript-based approach was used to avoid potential perceptual confounds from fluency. To determine the normative values of these characteristics, we included the discourses of an equivalent group of matched, healthy control speakers in our analyses. Finally, we compared our results to those in Matchin et al. (2020) to see if our innovative diagnostic criteria better categorize discourse samples. Therefore, the present study presents innovative, clear diagnostic criteria for agrammatism and paragrammatism to inform extant psycho- and neurolinguistic theories of syntax, and to support future lesion-deficit correlation work.

Procedure

Novel diagnostic criteria for grammatical deficits in aphasic speech

Based on the Bock & Levelt model, we separate potential error types into two gross categories: Functional and Positional. Errors in functional processing occur at the level of hierarchical organization and would affect lemma processing/lexical syntax & functional assignment. Errors in lexical syntax could result from the inability to completely access lemma representations, the loss of (part of) lemma representations, represented by blends, (i.e., fusion of two words that are near-synonyms), subcategorization errors (e.g., regularization of verbs, incorrect grammatical gender) or a fusion error (i.e., a lexical item appears to be used as two different lexical syntactic categories simultaneously). Errors in hierarchical structuring could result from a failure to specify syntactic relations or incorrect matching of lexical items to syntactic relations. Errors in matching of lexical items to syntactic relations could present as improperly placed constituents (i.e., exchange errors) or subcategorization errors (i.e., constituents are not the correct type). Errors in syntactic relation specification could be represented by nodes (or even whole branches) unfilled by lexical content (because they were never selected in the first place).

Errors in positional/morphosyntactic processing would affect constituent assembly and inflection. Errors in constituent assembly could result from reduced ability to access sentence frames, resulting in the production of formulaic or canonical sentences only (which is not an error) or from a complete loss of the ability to order (i.e., linearize) syntactic relations, as morphosyntactic elements are inserted at the linearization level. Such a loss would result in telegraphic speech, since linearization occurs before inflection. Errors in inflection could result from reduced inflection inventory, resulting in omission of functional morphemes.

Some errors that occur in aphasic discourse are difficult to interpret as they may reflect a repair of errors generated at earlier processing levels, while others may reflect a combination of multiple errors at different processing levels. For example, stranding errors appear as errors in inflection but are caused by ‘repaired’ exchange errors during functional processing. Similarly, refilled nodes may be the result of repaired errors to lexical selection or functional processing. Agreement errors may represent errors at different processing levels, such as errors in defining syntactic relations, errors in selecting functional morpheme representations, errors in building sentence frames, or errors in inflectional integration after frame building is complete. Agreement errors due to the first two processes would be part of functional processing while the second two would be part of positional processing. We predict that English speaking patients with damaged functional processing (i.e., paragrammatic-like speech) would be more likely to produce agreement violations from the use of incorrect functional morphemes. On the other hand, English speaking patients with damaged positional processing (i.e., agrammatic-like speech) would be more likely to produce agreement violations of omissions; the inability to linearize elements is a precursor to the inflection insertion operation steps. Problems in either of these steps would result in morphemes not being inserted.

Our coding scheme creates new diagnostic criteria for the characterization of agrammatism and paragrammatism. For a full description of utterance codes, see Table 1, below. Our codes are based on a distinction of errors that we believe occur during functional, hierarchical processing (‘Hier’), and those we believe occur during positional, linear processing (‘Lin’). We also distinguished errors that we believe are part of grammatical encoding but were not clearly part of functional processing or positional processing: omission of a single morpheme (‘O’), and errors that may indicate breakdowns in both functional and positional processing (‘F+P’). Finally, we distinguished a number of other errors which we believe are part of phonological encoding or conceptual messaging and do not align with grammatical encoding, such as paraphasias (e.g., neologisms, ‘N’), semantic errors (e.g., semantic substitutions, ‘S’), and filled pauses (e.g., ‘H’). These codes were applied only to utterances that did not contain errors of grammatical encoding. Finally, utterances devoid of errors were coded as grammatical (‘G’). We are concerned primarily with the syntax but we note that work in this field often implicates phonological breakdowns in paragrammatism and even as being central to the syndrome. As such, we include syntactically well-formed sentences containing phonological and semantic errors as distinct categories in our analysis.

Utterance Codes	***Example***
Hier: deficits in **functional** processing; blends, subcategorization violation, fusion errors, exchange errors, unfilled nodes or branches (e.g., lexical items or phrases requirements by syntactic relations)	*"She was thinking about that things that she was for everything she was goings something"*
Lin: deficits in **positional** processing; word order violations, functional morpheme errors (e.g., inflection errors), telegraphic speech (evidence of defective frame building)	*"And he meet the prince at party"*
O: ungrammatical utterance (a/p indeterminate due to single **omission** of a functional morpheme)	*"And the first guy wasn't going them"*
H+L: combination of functional and positional errors	*"Anyway it's right at midnight she realize she's supposed to go back at midnight back"*
N: phonological errors **paraphasia**/neologism only	*"Well they ripped her outfit apout"*
S: semantic errors; grammatical but **semantically** incoherent	*"The slippers went off"*
FP: Filled Pauses/ hesitation	*"Okay, umm"*
G: **grammatical** utterance	*"And they were all trying on the shoes to see which of them could get into the shoes"*

Table 1. Utterance codes.

Discourse sampling and participant diagnostics.

Speech samples were collected as part of larger studies, with audio and visual recordings. Recordings were elicited according to AphasiaBank protocol (Macwhinney et al., 2011). Speech-pathology graduate students transcribed the samples. Two expert raters (authors DF, JY), blind to participants' aphasia types, psychological assessment scores, lesions, and ratings in Matchin et al. (2020), sequentially analyzed each utterance of each participant according to the novel rating scheme, described above. Consensus was reached between the initial raters for each utterance. Two additional expert raters (WM, BCS) rated a subset of samples in order to assess interrater reliability. The subset of samples was the same for each additional rater, and was selected based to represent a variety of aphasia types and evaluations. Raters were trained on the SUBA method, which they had not used previously. These expert raters were also blinded to participants' aphasia types, psychological assessment scores, lesions, and ratings in Matchin et al. Raters were experienced language scientists with backgrounds in aphasia research, psycholinguistics, and theoretical linguistics.

Participants.

We analyzed the connected speech samples from the 53 participants included in Matchin et al. (2020), plus an additional 35 participants drawn from a database of individuals with chronic post-stroke aphasia (n=88; 31 women) plus 53 age-matched healthy control participants (34 women).

The inclusion criteria for aphasic speech samples did not vary from Matchin et al. (2020): all participants had a single ischemic stroke to the left hemisphere at least six months prior to sample collection and were pre-morbidly right handed. Participants in this replication were 59.2 (SD = 11.8) years old at time of testing, were 48.5 (SD = 50.3) months post-stroke, had 15.6 (SD = 2.3) years of education and a Western Aphasia Battery - Revised [WAB-R] aphasia quotient [AQ] of 66.6 (SD = 20.2) (Kertesz, 2007). Participants were diagnosed with the following aphasia types: 21 with anomia, 32 with Broca's aphasia, 17 with conduction aphasia, 4 with global aphasia, 1 with transcortical motor aphasia, 1 with transcortical sensory aphasia, and 6 with Wernicke's aphasia. Six participants were classified as not aphasic according to the WAB-R. Participants with fewer than 5 utterances containing analyzable content in their Cinderella elicitation were excluded from this study. Control participants were age-matched healthy without a history of cardiac events or psychological disorders. Controls were 60.1 (SD = 14.0) years old at time of testing and had 15.1 (SD = 2.26) years of education. Control were gathered from the Wright (Wright, 2008), Richardson (Richardson, 2008), Capilouto (Capilouto, 2008), and MSU (Boyle, 2008) corpora on AphasiaBank (MacWhinney et al., 2011).

Analyses.

To probe our data for phenotypes related to functional vs. positional processing, we employed an unsupervised k-means clustering technique. For each participant (PWA and controls) a vector of 15 values was used to represent their distribution over error types. We counted up the total number of errors of each type (i.e., Hier, Lin, H+L, O, N, S) committed by each participant. We then used this to calculate the proportion of all utterances of each error type (e.g., number of P utterances divided by the total number of utterances from that participant), as well as the proportion of errorful utterances of each error type—number of errors of each type divided by total utterances that were not coded as being error-free. We also included the proportions of grammatical and ungrammatical utterances, as well as the total number of utterances produced. The use of proportions rather than counts of error types is motivated by the large variance in the total number of utterances produced (range: 5-178, mean = 42, SD = 33.1). Admittedly, using proportions overweights any given error from participants with few utterances, however the inverse (just counting utterances) would effective partition participants by the amount of utterances produced rather than by the kinds of errors they committed. We employed a Partitioning Around Medoids (PAM) variant of the k-means algorithm—as implemented in the factoextra package for R (Kassambara & Mundt, 2016)—which is less sensitive to outliers. Based on examination of the elbow plot, we set maximum k at 12. The partitioning around medoids algorithm then selected 10 clusters as the optimal number. The error distributions for all 141 participants (88 PWA, 53 HC) were included at the same time.

We completed qualitative syntactic analyses of selected individual utterances from these clusters. We herein provide additional analyses that examined specific features we expected to be affected from our interpretation of the Bock & Levelt model but that usage-based models (non-generativist models in which sentences are generated on the basis of statistics rather than hierarchically generated) predict would not be selectively impaired. These specific features were the lexical syntax (i.e., lemmas) of irregular words and aspect (with tense selectively preserved) (Bastiaanse et al., 2011; Marslen-Wilson et al., 1994; Martínez-Ferreiro et al., 2015; Martínez-Ferreiro & Bastiaanse, 2013). We then compare the results of our clustering analysis to the perceptual judgments presented in Matchin et al. (2020) for the participants who appear in both datasets. Finally, we compared the scores derived from our analysis with overall performance on the Western Aphasia Battery—Revised (WAB-AQ), the WAB fluency section, speech rate measured in words per minute (WPM) from their discourse sample, and mean length of utterance (MLU; calculated by the CLAN software) in both words and morphemes. We used Pearson's correlations to test for these relationships (using the cor.test function in R) with an alpha threshold of $p < 0.01$ given 86 degrees of freedom.

In order to determine rater agreement between the original consensus codes and the codes provided by the additional raters, Cohen's kappa was used. Overall, secondary raters agreed with the consensus codes on grammaticality in 79% of utterances, and exact code in 69% of utterances. There was "moderate" agreement between the raters and the original codes with regard to grammaticality (k=0.552, 95% CI: 0.489 – 0.615), and "moderate" agreement between raters with regard to specific codes (k=0.445, 95% CI: 0.385 – 0.503). The largest amount of disagreement surrounded whether an utterance was grammatical or had a functional processing error in it. Other sources of disagreement were between functional and mixed errors, or grammatical and single utterance omission errors (Fig. 1). The lack of agreement around what constitutes a paragrammatic error is consistent with the interrater reliability results from Matchin et al. (2020).

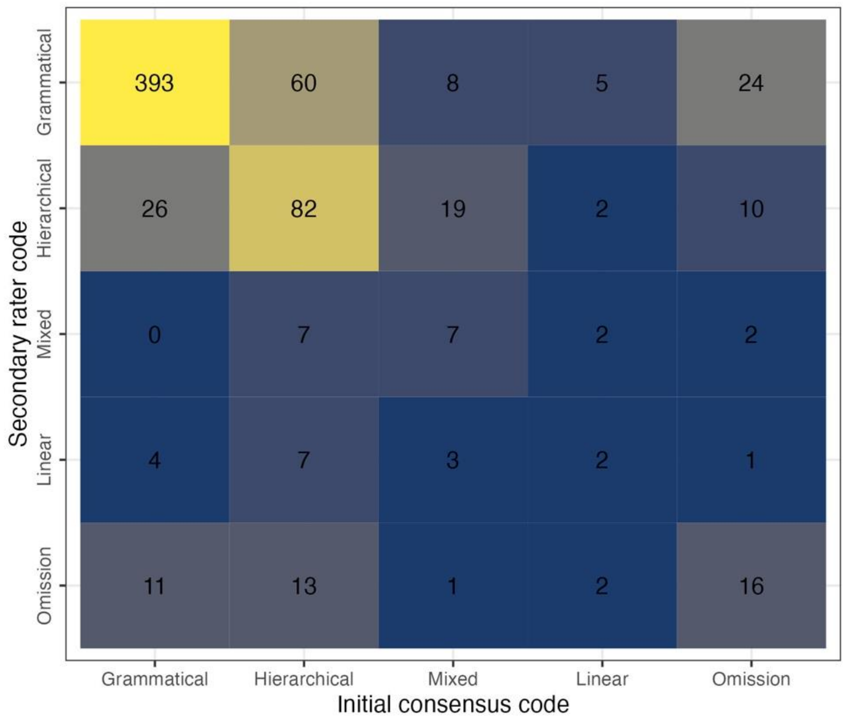

Figure 1. Agreement matrix for original consensus codes and secondary rater codes. The vast majority of codes fall along the diagonal, however there was some lack of alignment on functional processing vs mixed errors or grammatical utterances, as well as between single morpheme omissions and grammatical utterances.

Results.

Unsupervised k-means clustering of PWA and controls.

Results from our unsupervised k-means clustering of PWA and controls produced 10 clusters, visualized in Figures 2 and 3. We provide a brief summary of each cluster in Table 2.

Cluster ID	Description
1	High values of both proportion of utterances and proportion of errors showing errors in positional processing. Conversely, this cluster has very low numbers of both grammatical utterances, and errors in functional processing.

Cluster ID	Description
2	High values of both proportion of utterances and proportion of errors showing errors in functional processing. This cluster also produced more utterances on average, as well as some semantic paraphasias.
3	High values of mixed (P+F) errors. This cluster also showed a high value of errors of omission, as well as a relatively high value of errors in functional processing.
4	High values of errors of omission, as well as grammatical utterances containing phonological paraphasias.
5	Characterized by high values of phonological paraphasias but low omission rates.
6	Mixture of high numbers of semantic paraphasias and functional processing errors, as well as some omissions.
7	Characterized by errors of omission despite mostly grammatical output but low phonological paraphasia rates.
8	Predominantly grammatical utterances, where all errors are at the level of functional processing.
9	Predominantly grammatical utterances, with some semantic paraphasias.
10	Predominantly grammatical utterances with a small number of omissions, functional processing errors, and phonological paraphasias.

Table 2. Description of clusters.

In a nutshell, we interpret Cluster 1 as having difficulties in positional processing, while we interpret Clusters 2, 3 and 6 as having difficulties primarily in functional processing. We interpret Cluster 4 and 7 as having mild grammatical impairment, and Clusters 5 and 9 as having errors primarily unrelated to grammatical encoding. Finally, we interpret Clusters 8 and 10 as patterns of errors made by individuals without deficit.

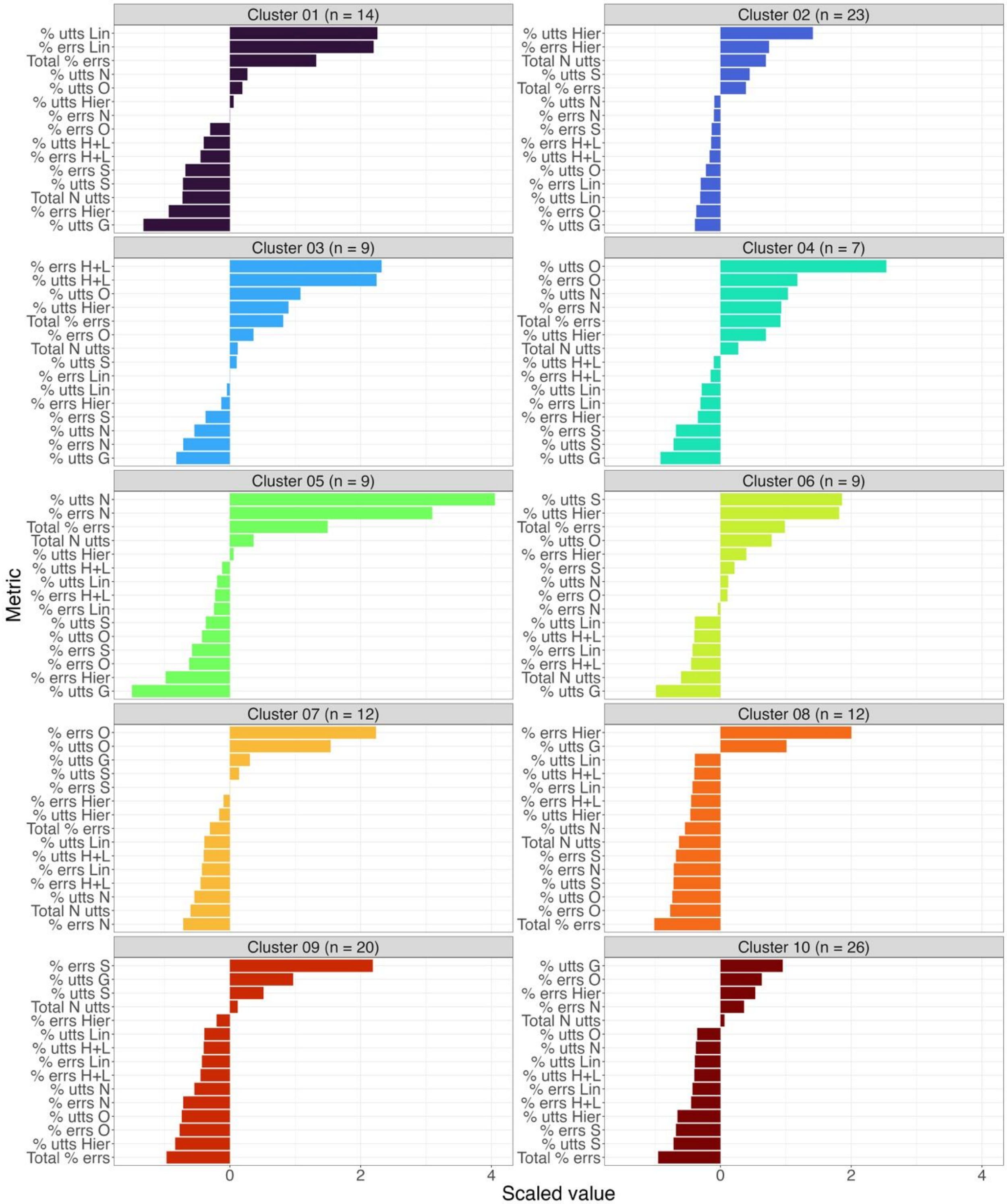

Figure 2. Clustering results: Cluster medoids. X-axis shows scaled value of medoid location for each metric: zero denotes average across all participants for that metric. Each y-axis is reordered with highest values (above average) at the top and lowest (below average) at the bottom. % utts α: Percentage of an individuals' utterances coded as α; % errs α: Percentage of an individuals' errorful utterances (total utterances less utterances coded as G) coded as α; Total N utts: Total number of utterances produced by an individual; Total % errs: Proportion of total utterances which were not coded as G. G: Utterances containing no syntactic, phonological, or semantic errors; L: errors in linear processing; H: errors in hierarchical processing; H+L: utterances containing hierarchical and linear errors; O: utterances missing a

single functional morpheme; S: syntactically well-formed but containing a semantic error; N: syntactically well-formed but containing a phonological error.

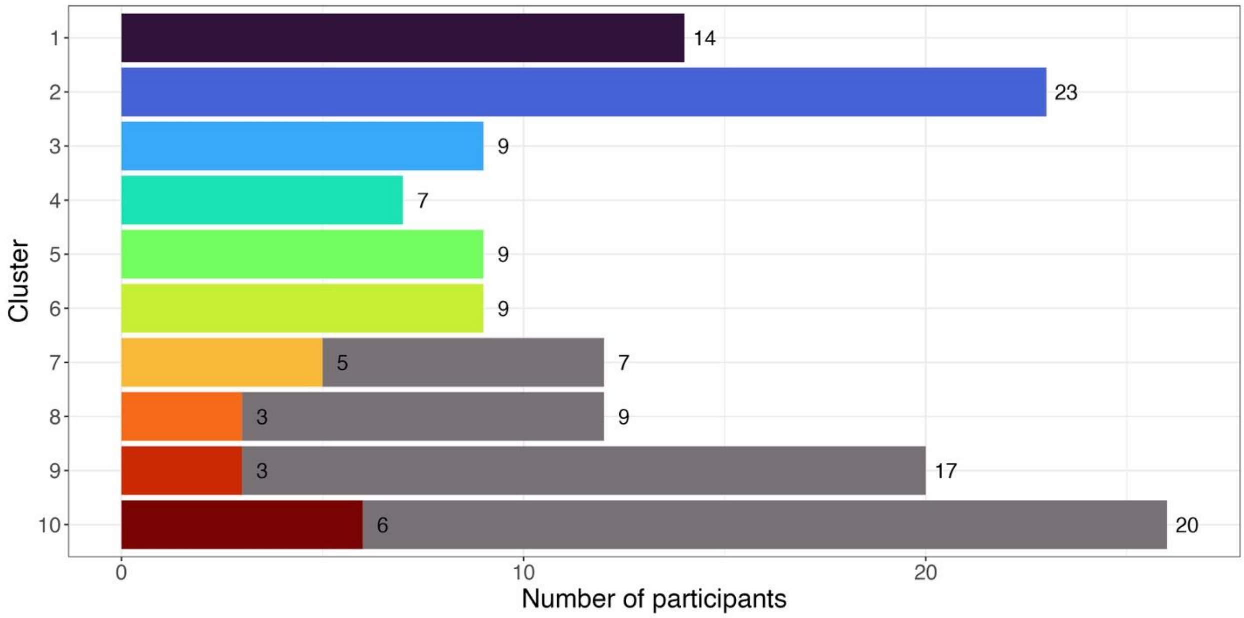

Figure 3. Clustering results. Cluster membership: Colored bars denote PWA, gray bars denote healthy controls. Note that clusters 7-10 contain both PWA and controls, while the remaining clusters are exclusive to PWA.

Qualitative analyses of individual utterances.

To reiterate, difficulties in functional processing should affect lemmas and hierarchical structuring. Errors at these levels should result in subcategorization errors, fusion errors, exchange errors, or nodes (or even whole branches) unfilled by lexical content. This maps to our description of Clusters 2 and 6. Individuals who have difficulties in functional processing may have errors in positional processing, as it occurs later in the cascade. Therefore, stranding errors and agreement violations (due to the use of incorrect functional morphemes) may be associated with difficulties in functional processing. This maps to our description of Cluster 3. Difficulties in positional processing should affect linearization (e.g., word order) then inflection, and result in formulaic or telegraphic speech then omission of functional morphemes. This maps to our description of Cluster 1. Omissions of inflections may result due to difficulties at either level, with some individuals only having mild grammatical impairment. This maps to our description of Clusters 4 and 7. To further our analyses, we now provide a qualitative description of individuals within these clusters, matching expected errors to patterns observed (or not) in Table 3.

Analyses of these examples largely match the predicted patterns, though some exceptions occur. Individuals in Clusters 2, 3 and 6 largely made errors in functional processing, but also made errors resulting from semantic substitutions (e.g., subcategorication and blending errors), and errors in positional processing (e.g., inflection errors). Individuals in Clusters 3 and 6 made functional processing errors and positional processing errors. There were no clear instances of stranding errors, though. Individuals in Cluster 1 largely had errors of positional processing, but some Cluster 1 individuals made errors by incorrectly inserting functional morphemes, rather than simply omitting them. Individuals in Cluster 4 made errors in a single grammatical node (i.e.: single morpheme/function word omissions), but also made more complex errors. Individuals in Cluster 7 largely made errors in a single grammatical node. Individuals in Cluster 8 were primilary controls but did have a higher proportion of functional errors compared to other control-dominated clusters.

Finally, individuals in Clusters 9 and 10 were not characterized largely by errors to functional or positional processing. Nevertheless, individuals in Clusters 7, 8, 9 and 10 did make errors in both processing stages. Individuals in Cluster 5 made a higher proportion of phonological paraphasias and neologisms which are common in aphasia, but not specifically related to grammatical processing.

Cluster	Participant	Utterance	Descriptive Analysis
1	M2002	"daughters angry"	telegraphic: only lexical items associated with proposition / missing many functional morphemes, but includes inflected noun
		"twelve o'clock ... Cinderella ... go to … home"	functional morpheme omission: missing prepositional head of temporal adjunct 'at', missing tense on noncanonical verb 'go', and missing determiner in
	M2025	"man Cinderella married"	word order error: both agentive nouns before the verb
	M2071	"dress … ball … and prince"	telegraphic: only lexical items associated with proposition
		"and Cinderella … one shoe … one shoe… shoes"	mixed error: unfilled node: missing verb; selection of incorrectly inflected form
	M2186	"sisters can't stand them"	single omission: would be grammatical with determiner
		"finally one o'clock have to go"	mixed error: unfilled node: missing subject pronoun 'she'; functional morpheme omission: missing preposition 'at', missing agreement on verb, but includes complementizer 'to'
		"the hurry up and go"	idiom error: idiomatic expression 'hurry up and go' used as a noun
		"and Cinderella the shoe fit"	word order error: agentive noun precedes patient noun (following canonical ordering)
	M41037	"and horses and cats and yeah"	telegraphic: lexical items associated with proposition, with filler phrase 'yeah'
		"and girl not good"	telegraphic: only lexical items associated with proposition
		"and ... but … little mouse yeah and Cindrella no but good yea"	telegraphic: lexical items associated with proposition, with filler phrase 'yeah'
		"and prince and Cinderella all the time thank you in castle"	telegraphic: lexical items associated with proposition, with some phrasal structure (e.g., 'all the time', 'in castle')
		"Cinderella not good"	single omission: would be grammatical with copula
	M4201	"the man slippers and look"	telegraphic: lexical items associated with proposition, with some functional morphemes 'the' and plural 's'
		"and slipper fit"	single omission: would be grammatical with determiner
2	M2031	"who they know that"	refilled node: complementizer 'that' fills node after movement of complementizer 'who'
		"and he thinks there's only two of it"	inflection error: singular pronoun 'it' selected
		"but anyway one of the person it leaves at home"	inflection error: singular noun of noncanonical 'person' selected
	M2061	"they were her… they were pretty up"	fusion error: 'pretty' used as adj. & adv.
	M2076	"anyway ... they been all these years"	unfilled node: missing auxiliary

Cluster	Participant	Utterance	Descriptive Analysis
	M2176	"and ... they was fit her"	filled node: agentive argument is used in object patient role (fem. pronoun), the noncanonical structure of unaccusative verb 'fit'is 'repaired' to a passivization
	M41048	"a kardsman [sic] came back well the three sisters didn't fit the ... slipper"	exchange error: noncanonical verb 'fit' subject and object exchanged
	M4171	"and then ... this was like who is thi[s] shoes"	functional morpheme error: complementizer 'what' replaced with complementizer 'who'
	M4175	"and so the father comes to … take care for the ... Stepmother"	blend: near-synonyms 'take care of' and 'care for'
		"she said well you know what I can be a …"	unfilled node: missing noun following determiner
		"you can't everything"	functional morpheme error: negative polarity error
		"that ... said oh ... I couldn't go"	functional morpheme error: nominative pronoun 'she' replaced with referential pronoun 'that'
		"and she hit"	single omission: would be grammatical with 3pSg 's'; alternatively, the past tense 'hit' was unsed with an unfilled node and/or mixing with the habitual form which cannot be past tense
3	M2046	"... shoe kit [got] caugh[t] ... too out ... and lost it too	subcategorization error: adverb 'too' modifying preposition 'out'
	M2182	"so later it's Cinderella"	functional morpheme error: incorrect usage of neoplastic pronoun
	M4214	"and she came back and it fell her [o]pen ... yeah"	word order error: ungrammatical dative shift of prepositional verb 'fell open' and pronoun 'her'
	M4218	"the wife and two brides is ugly"	functional morpheme error: copula number (and possible tense) error
		"but midnight is Cinderella's one shoe"	mixed error types: temporal concept 'midnight' is subject noun rather than adjunctive prepositional phrase; copula joins temporal concept and sentence proposition
4	M2006	"now another man came"	inflection error: aspectual agreement violation
	M4208	"she has with three girls"	subcategorization error: prepositional complement
		"an[d] then ... the girls, they was … it was … they was … they was [d]ancing"	attempt to repair functional morpheme error
		"woman said has ... something … was going back"	mixed error types: unfilled node (subject of 'has'), missing complementizer or improperly added clause
5	M2136	"when there [unintelligible] her seen her dance [unintelligible] boy"	improperly joined elements, but functional elements have proper case markings (e.g., 'there', 'her')

Cluster	Participant	Utterance	Descriptive Analysis
	M2181	"before that ... she want a have a different [unintelligible]"	mixed error: incorrectly inserted functional morpheme 'want *a* have'; missing functional morpheme for tense on 'want'
6	M2036	"and the whisk away"	fusion error: 'whisk' used as noun & verb
		"and … the girl is what can we do"	unfilled node: missing quotative 'like'
	M4134	"and she the lady that worked for everybody"	single omission: missing copula
	M4134	"before ... she was a turn back to a being a working lady"	fusion errors: 'turn' and 'being' used as nouns and verbs
7	M2102	"… the mouse were trying to do a dress for her ..."	Inflection error: verb number agreement
		"the woman both had daughters"	inflection error: singular noun of noncanonical 'woman' selected
	M2170	"and so Cinderella got to the ball and met the prince and started dancing with her"	functional morpheme error: potentially a reinterpretation of 'prince' as subject of start, rather than a gender mismatch on 'her'
8	M2139	"one upon a time"	indiom error: idiomatic expression 'once upon a time' numercal category incorrect
	M2139	"and … the twelve is dancing"	inflection error: verb agreement violation; alternatively, nominalized form of the temporal adverb replaced the nominative subject pronoun 'they'
	M2139	"and … they find it in"	unfilled node: missing required noun complement in prepositional phrase
	M2139	"and live happily ever after"	idiom error: idiomatic expression 'they lived happily every after' missing functional morphemes
9	M2069	"she's wasn't going"	refilled node: auxiliary verb node filled with non-complementary auxiliaries
10	M2103	"and he tries on the shoe on many women"	subcategorization/fusion: verb complement selection violation
	M2104	"not no one else"	functional morpheme error: mismatch negativity

Table 3. Qualitative examples of predicted functional and positional errors by cluster.

Beyond the types of errors characterizing the clusters, several other patterns of errors are notable. For one, individuals attempted idiomatic expressions often, particularly those associated with the Cinderella story. The functional morphemes and/or phrasal structure in these idioms were often affected. Another notable pattern of error affects noncanonical lexical items/structures. In particular, individuals producing this discourse sample often used the verb 'fit' which is noncanonical (e.g., ergative verb, irregular conjugation). Samples from PWA included errors on 'fit' 52% of the time, though matched controls similarly made errors on 'fit' 48% of the time. Finally, we observed that individuals in Cluster 1 omitted verbs more often than other lexical morphemes, but omission of verbs was not noticeable in other clusters.

Comparison of our results to the Matchin et al. (2020) study.

In comparison to the original Matchin et al. (2020) study, we predicted that participants in Cluster 1 who were included in the Matchin et al. (2020) study would have been classified as Agrammatic because Cluster 1 is weighted toward errors of positional processing. In the same way, we predicted that members of Cluster 2, which was weighted toward errors in functional processing, would have been classified as Paragrammatic by Matchin et al., and that members of Cluster 3 (weighted towards mixed P+F errors) would have been classified as either "both agrammatic & paragrammatic" or simply as "paragrammatic". Since Cluster 4 included a high proportion of utterances coded as having phonological errors ('N') and a high proportion of errors being 'N', we predicted that participants in this cluster who were included in the Matchin et al. (2020) study would be described as 'neither' (or 'paragrammatic' but with lower consensus). Since Cluster 5 included a high proportion of utterances coded as having semantic errors ('S'), functional processing errors ('F') and single omissions ('O'), we predicted that participants in this cluster who were included in the Matchin et al. (2020) study would also be described as 'paragrammatic'. Since Cluster 6 included a high proportion of utterances coded as having omissions ('O'), as well as a high proportion of errors as 'O', we predicted that participants in this cluster who were included in the Matchin et al. (2020) study would be described as 'neither' but may be coded as 'agrammatic' 'paragrammatic' or 'both'. Since Cluster 7 included a high proportion of errors coded as semantic errors ('S') but a high proportion of utterances coded as grammatical ('G'), we predicted that participants in this cluster who were included in the Matchin et al. (2020) study would also be described as 'neither' (or 'paragrammatic' with lower consensus scores). Finally, we predicted that participants in Clusters 8, 9 and 10 who were included in the Matchin et al. (2020) study would be described as 'neither' because they had a high proportion of grammatical utterances and these clusters primarily included healthy controls. It is important to note, however, that perceptual judgements were available for only the original 53 PWA included in the Matchin et al. study. Results of this comparison are shown in Figure 4. No such judgments were available for the additional 35 PWA, nor for the healthy controls.

We found that indeed most of the participants in Cluster 1 were classified as Agrammatic, however not all participants that were classified as Agrammatic were in Cluster 1. Participants classified as Paragrammatic by Matchin et al. appeared in every cluster except Cluster 9. On the other hand, the largest chunk of participants in Cluster 2 were classified as Paragrammatic, with the second largest bloc having been classified as Neither. Within Cluster 3, which seemed to be characterized by mixed errors, participants had been classified as some Agrammatic, some Paragrammatic, and some Both. Among participants in Cluster 5, a cluster characterized by phonological errors, some were classified as Paragrammatic, and some as Neither. In Cluster 6, which was characterized by semantic and functional processing errors, some participants were classified as Paragrammatic, some as Neither, and one each of Agrammatic and Both. Clusters 4, 7, 8, 9 and 10 were not populated with sufficient individuals from the original 53 PWA's samples to draw conclusions about those predictions.

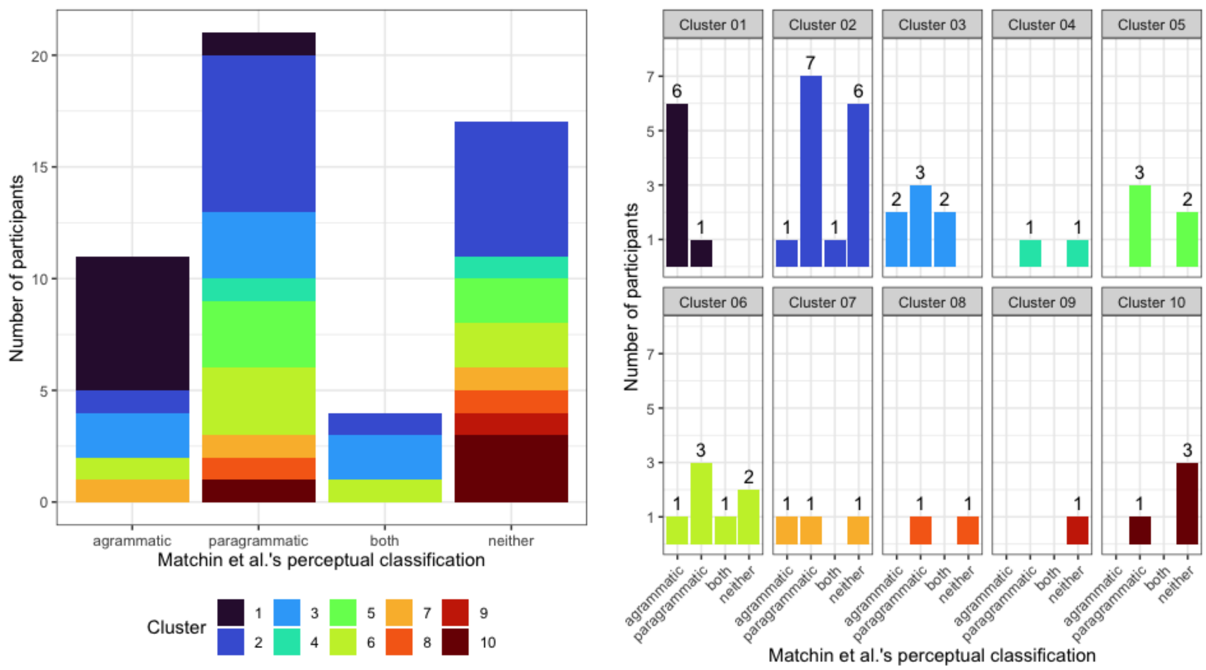

Figure 4. Cluster membership of participants included in Matchin et al. (2020). A: Cluster membership by perceptual classification. B: Perceptual classification by cluster.

Comparison of our results to WAB-R and fluency measures.

We found significant correlations between several error type distributions and behavioral scores from other standardized aphasia assessments (Table 4, Figure 5). Proportion of error-free (G) utterances was significantly positively correlated with WAB-AQ and with all of our fluency measures (WAB fluency, words per minute [WPM], and MLU in either words or morphemes). Proportion of utterances and proportion of overall errors with positional processing errors were negatively correlated with all of our fluency measures, but not significantly related to WAB-AQ. Proportion of overall errors (but not proportion of utterances) with functional processing errors was positively correlated with WAB fluency and MLU in either words or morphemes, but not with WAB-AQ. Proportion of overall errors which were grammatical but contained semantic errors was significantly positively correlated with WAB-AQ and the various fluency measures. Since potential bias from perception of speech rate was not possible with transcription analyses unlike the original analyses in Matchin et al. (2020), these parallel correlations suggest that there may be relationship between speech rate and grammatical processing that could aid in identifying how these syndromes dissociate.

Utterance type	Metric	Score	Degrees of freedom	Estimate (r)	P-value
P	Prop. Utterances	WAB Fluency	86	-.40	<.001
P	Prop. Utterances	WPM	86	-.40	<.001
P	Prop. Utterances	MLU Morphemes	86	-.32	.002

Utterance type	Metric	Score	Degrees of freedom	Estimate (r)	P-value
P	Prop. Utterances	MLU Words	86	-.32	.002
P	Prop. errors	WAB Fluency	86	-.42	<.001
P	Prop. Errors	WPM	86	-.41	<.001
P	Prop. Errors	MLU Morphemes	86	-.38	<.001
P	Prop. Errors	MLU Words	86	-.39	<.001
F	Prop. Errors	WAB Fluency	86	.38	<.001
F	Prop. Errors	MLU Morphemes	86	.33	.002
F	Prop. Errors	MLU Words	86	.32	.002
G	Prop. Utterances	WAB-AQ	86	.32	.003
G	Prop. Utterances	WAB Fluency	86	.47	<.001
G	Prop. Utterances	WPM	86	.45	<.001
G	Prop. Utterances	MLU Morphemes	86	.34	.001
G	Prop. Utterances	MLU Words	86	.32	.003
S	Prop. Errors	WAB-AQ	86	.31	.003
S	Prop. Errors	WAB Fluency	86	.36	<.001
S	Prop. Errors	WPM	86	.29	.005
S	Prop. Errors	MLU Morphemes	86	.32	.003
S	Prop. Errors	MLU Words	86	.32	.002

Table 4. Relationship between discourse scores and behavioral & fluency measures. Only significant correlations are shown at $p < 0.01$.

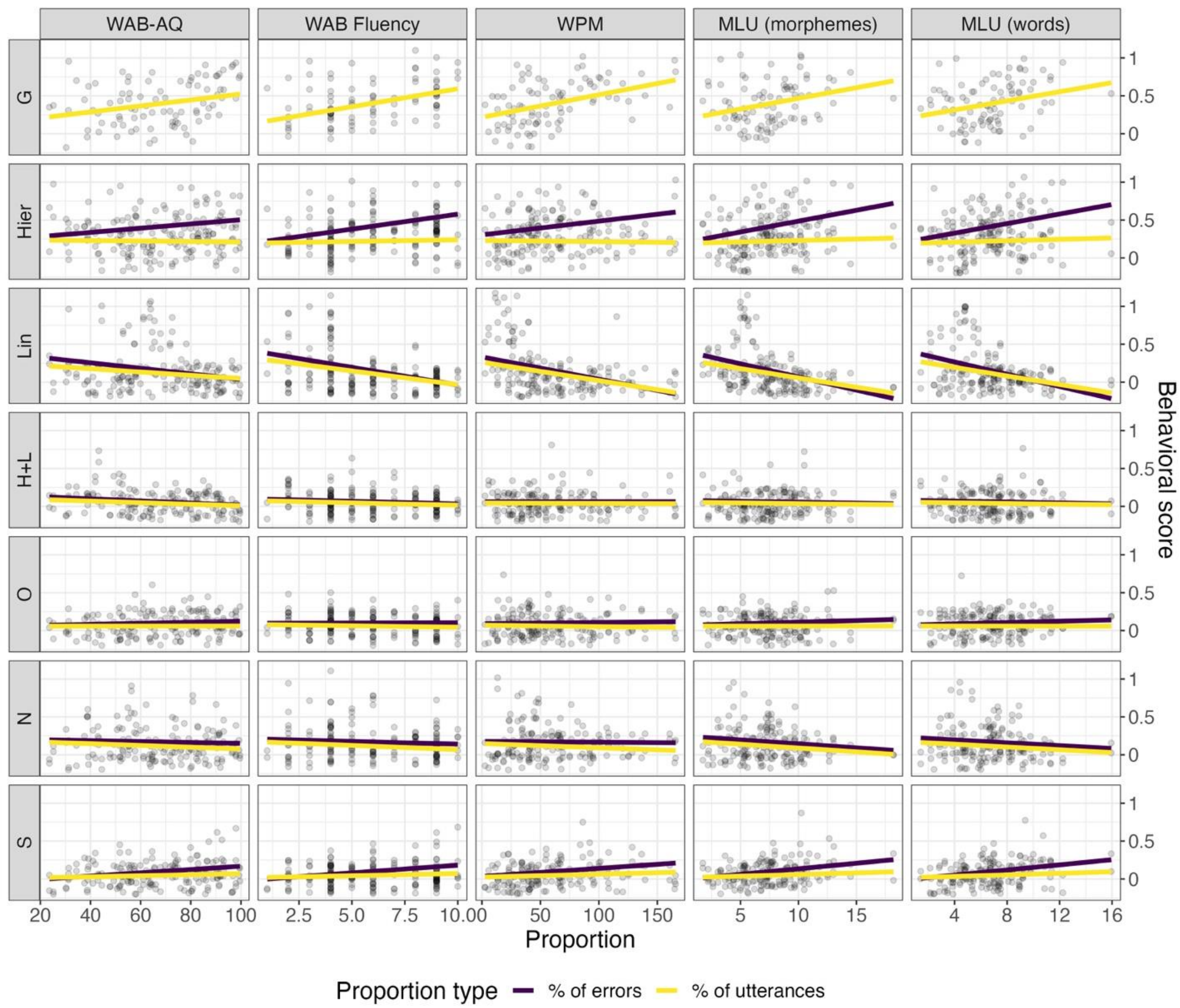

Figure 5. Correlations between standardized behavioral scores and proportion of utterances or proportion of errors of each code.

Discussion.

Based on assumptions about the underlying two-stage syntactic processing mechanisms proposed by Bock and Levelt (1994), our results largely support our predictions about the correspondence between agrammatism to positional processing and paragrammatism to functional processing. We predicted that disruptions to positional processing may result in word order or inflection errors, and should correspond to an agrammatic syndrome, while disruptions to functional processing may result in blends, subcategorization errors, fusion errors, exchange errors or stranding errors, and should correspond to a paragrammatic syndrome. Our k-means cluster analyses showed that individuals who made positional processing errors largely did not make functional processing errors, but individuals who made functional processing errors did make positional processing errors. This aligns to the cascade pattern of processing predicted by our interpretation of Bock and Levelt (1994) as well as dissociation of errors found by Matchin et al. (2020).

Historical difficulty in characterizing these syndromes or distinguishing them from one another closely parallels a debate over the nature of the underlying deficit and its neural correlates. Matchin and Hickok (2020) proposed that hierarchical syntactic processing is supported by the posterior middle temporal gyrus (pMTG) while linear morphosyntactic processing is supported by the inferior frontal gyrus pars triangularus (IFGtri). This theory was bolstered by findings from Matchin et al. (2020) which showed a double dissociation of lesion location between patients categorized as agrammatic or paragrammatic. Though there is a logical alignment of agrammatic deficits to errors in linear morphosyntactic processing and paragrammatic deficits to errors in hierarchical syntactic processing, supporting syntactic analyses have not been performed (Bastiaanse et al., 2011; Goodglass & Hunt, 1958; Grodzinsky, 1986, 2000; Grodzinsky & Friederici, 2006; Grodzinsky & Santi, 2008; Santi & Grodzinsky, 2007). Nonetheless, the Matchin & Hickok (2020) model's levels of "hierarchical" vs "linear" processing provide good conceptual alignment with the Bock & Levelt model. Furthermore, the spatial delineation described by Matchin & Hickok for these two processing levels provides a neurbiological basis for the presentation of the two grammatical production syndromes.

The definitional debate of grammatical production deficits has largely centered on 'agrammatism' and 'paragrammatism'. Our clusters of error patterns principally reflect our interpretations of these syndromes. While one large cluster of individuals seem to reflect the characterization of agrammatism as having difficulties with linearization and functional morphemes, multiple clusters reflect the characterization of paragrammatism as having difficulties with lexical syntactic processing and hierarchical syntactic processing. Our cluster of individuals with positional processing deficits (Cluster 1) seems comparable to historical definitions of 'agrammatism'. The production primarily of nouns matches Broadbent's (1872) characterization, while disturbances in grammatical morphemes aligns to Pick's (1913). One potential area of non-convergence from common descriptions is our exclusion of substitutions of grammatical morphemes as representative of 'agrammatism' (c.f. Bastiaanse et al., 2011). Our interpretation of morphosyntactic substitutions was that they are a carry-over effect from functional processing errors. While we believe that the evidence largely matches our theoretical interpretation, we must note that the sole use of English data may bias these findings. It is generally assumed that speakers of different languages store and process grammatical morphemes similarly (Tyler et al., 2011), so omissions of grammatical morphemes would hypothetically result in non-word stems in highly inflected languages. However, morphosyntactic errors by non-English speaking people with aphasia resulting in non-word forms is largely (though not entirely) unattested (Paradis, 2001). Determining whether morphosyntactic substitutions align to positional processing and agrammatism could be done by providing a cross-language comparision of errors by PWA using the SUBA method.

On the other hand, our results did not provide a single cluster of individuals with functional processing deficits correlating to our theory-driven analyses. Which is potentially to be expected as the algorithm was not forced to do so. The multiple clusters (Clusters 2, 3 and 6) corresponded to breakdowns in functional processing. One possible explanation for this inconsistency is that paragrammatism has distinct phenotypes, while agrammatism does not. Such phenotypes might correspond to errors in lexical selection and processing, cascading into functional processing, cascading into positional processing and/or structural checking mechanisms. This theoretical possibility aligns to these distinct clusters. For example, Cluster 2 was characterized by having a higher proportion of semantic paraphasias in addition to functional processing errors. Indeed, M4175's discourse included a grammatical error resulting from a blend. Therefore, blends, subcategorization errors, and fusion errors may result from disruptions beginning during conceptualization, prior to functional processing. Similar to the interpretation that exchange errors may be the result of functional processing errors reflected later in the processing cascade, blends, subcategorization errors, and fusion errors may result from a error beginning in lexical processing. Cluster 3 was also characterized by both functional and positional processing errors. We predicted that we should observe stranding errors as a result, but we did not did identify clearcut examples of stranding errors. Instead, we suggest that functional morpheme errors and refilled nodes may be the result of stranding errors. Through this lens, this type of error is not a failure to access a functional morpheme or a syntactic rule, but rather a disruption of a checking mechanism (Matchin & Hickok, 2020). It is conceivable that processing breakdowns may occur systematically at multiple steps, rather than just two, and there is theoretical concordance around the issue of a processing cascade (c.f. Dell et al., 1997; Levelt et al., 1999). This explanation may help resolve historical debates about the definition of 'paragrammatism'. For example, lexical selection resulting in grammatical errors parallels to "wrongly chosen" items described by Kleist (1914). Similarly, the sentences "swell[ing] to confused sentence monsters" described by Head (1920) may be reflective of ungainly sentences not prevented through the checking mechanism. In contrast, Broadbent (1872) described effects to formulaic speech (possibly idioms or noncanonical structures). Within examples of functional processing errors, we found several instances of these types of idiomatic and noncanonical structures being affected (e.g., 'happily ever after'; 'fit'). However, formulaic structures are not attached to the definitions of paragrammatism (or agrammatism as Broadbent was detailing) *per se*. Therefore, further investigations are needed to determine whether our phenotypes are psychologically valid, refining a new definition of paragrammatism. In particular, such research should specifically examine whether idioms or canonical structures are particularly affected.

The inclusion of a significant number of matched controls led unique clusters that were mostly 'healthy' speakers (Clusters 7, 8, 9 and 10). Were these speakers' patterns consistently similar, it would not be expected that they cluster separately. Indeed, these clusters did have unique qualities, including mild grammatical impairment (Cluster 7), semantic impairment (Cluster 9), and functional processing errors (Cluster 8). Mild functional processing errors may or may not be evidence of clinical findings. In fact, it could be representative of errors in checking mechanisms which have widely been noted in speech of healthy speakers (Levelt, 1989; Poulisse, 1999; Rogalsky & Hickok, 2009). While Clusters 9 and 10 were comprised of far more control samples, Clusters 7 and 8 had closer ratios. Such categories—particularly categories with similar numbers of control and aphasic samples—may indicate that some controls had processing difficulties. If these types of clusters are indicative of mild language impairment, particularly early language decline, they could be useful diagnostically. Currently, such decline is primarily assessed through cognitive assessments like the Mini-Mental State Exam (MMSE; Folstein et al., 1975) or Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005). Since language decline has been associated with early dementia diagnosis, tools for classification of this decline will be useful (c.f. Beltrami et al., 2018). Work examining brain age (c.f. Busby et al., 2023) may benefit from inclusion of measures such as those found by SUBA analyses.

Importantly, the recent definitions provided by Matchin et al. (2020) align well to our assignment of functional and positional processing errors to agrammatic and paragrammatic speech using the SUBA method and clustering analyses. Moreover, the perceptual classifications provided in Matchin et al. (2020) were decent at capturing the error types being produced by participants, but these judgments suffered from lack of agreement about what constitutes a paragrammatic error, and the fact that apart from severely agrammatic patients, participants with aphasia rarely produce only one single error type. The high inter-rater agreement for participants presenting with agrammatism in Matchin et al. may be due to their persistent production of a high number of agrammatic errors that we show here, but it is also possible that these judgments are bootstrapped and/or confounded by other factors—namely fluency variables—that tend to co-occur with agrammatic output. In a similar vein, it could be that the relatively low proportion of paragrammatic errors produced by any single participant (≤ 50% of utterances) made it difficult for Matchin et al. to extract that information from an assortment of either grammatical utterances or other error types and seemingly intact fluency. About two thirds of the 17 participants categorized by Matchin et al. as neither agrammatic nor paragrammatic patterned with the healthy controls or the clusters presenting with phonological and/or semantic but not syntactic deficitis in our clustering analysis (Fig. 4). However, six of them were grouped into the cluster characterized by the functional processing errors typical of paragrammatism in our analysis. It seems, thus, that differentiating at a macro scale between participants with no syntactic impairment and those with paragrammatism can be quite challenging. A question then arises about whether binning PWA as such is a useful approach. While it may be possible that a discrete number of categories or phenotypes could be found to describe the data, it seems that categorical judgment along two dimensions (i.e.: ±agrammatic, ±paragrammatic) is insufficient to capture the variance in discourse output among PWA. A classification into 'agrammatic' vs 'not-agrammatic' seems more promising, as there are specific intervention approaches that target the 'agrammatic' deficit (e.g., VNeST; Edmonds et al., 2009).

Following a theory that there are multiple phenotypes of paragrammatism (i.e., errors at different points in lexical and functional processing), we must consider the strength of SUBA method for categorizing potential new error types. The goal in developing this methodology was to provide a more nuanced approach to error identification using clear diagnostic criteria. We believe that SUBA was successful in distinguishing errors in functional and positional processing as was shown through the clustering analysis. For one, it provided improvement over fluency measures and normative assessments. From the comparison of our results with the WAB-R, none of the frequency of grammatical error types (F, P, F+P, O) were significantly related to overall aphasia severity. This points toward at least a partial dissociation between error frequency and overall aphasia severity: more severe aphasia does not in and of itself lead to a higher rate of functional or positional processing errors. Incidence of these errors does seem to correlate with some measures of fluency (cf. Table 3, Figure 5). In addition, the utterance-level analyses employed by SUBA may prevent prescriptive judgments of diverse varieties. Through SUBA, nonstandard constructions were not marked as errored. However, it is unclear the degree to which licit varieties across the dialect continuum can and should be coded as licit when the pre-morbid variety of speaker is unknown.

On the other hand, it is worth considering whether alternative available analyses could improve the linguistic analysis. A traditional type of analysis (c.f. MacWhinney et al., 2011) codes words and morphemes for structural errors from transcriptions *a la* SUBA, but does not include a mechanism for distinguishing structural errors or distinct types of dependencies, so would not necessarily offer an improvement over SUBA. A newer auditory-perceptual assessment system developed by Casilio and colleagues (Casilio et al., 2019, 2024) offers some improvements over those traditional methods. Casilio et al.'s auditory-perceptual framework distinguishes morphosyntactic errors from lexical misselection. Such a perceptual rating system from original recordings offers advantages over both traditional methods and SUBA because transcription followed by coding is time-intensive. Finally, it seems easier to train raters, including non-experts and students on this auditory-perceptual analysis. The ease of coding allowed good-to-excellent interrater reliability, compared to Matchin et al.'s (2020) good-to-very good interrater reliability and our moderate interrater reliability.

Though complicated, we believe that SUBA remains more subtle in the distinctions it draws between error types than other methodologies with similar goals. While we believe that SUBA represents an important step forward in characterizing expressive grammatical deficits, we do not by any means assert that the error categories presented here are the only ones possible. Future research should make adjustments to SUBA in order to distinguish errors at the phenotypic stages of functional processing. Additionally, future directions should incorporate lesion mapping analyses in order to crystalize the connection between psycholinguistic theory and well-supported brain models.

Conclusion

The goal in the present study was to connect neurolinguistic models and behavioral data from aphasic discourse samples to psycholinguistic models. The objective coding scheme of grammatical errors put forward here, informed by psycholinguistic models of underlying syntactic mechanisms, can provide a fruitful basis for revealing the neurobiology of syntax. We theorized that the processing model proposed by Bock and Levelt (1994) would provide a vital connection between the framework provided by Matchin and Hickok (2020) and post-stroke aphasia data. We used this framework to create the SUBA method, an initial step in analyzing discourse samples with fine-grained utterance level distinctions. After coding utterances in the SUBA method, we completed cluster analyses to determine whether these fine-grained classifications better defined the categorical syndromes of grammatical errors in aphasia based upon this framework. Overall, our analyses suggest that linear morphosyntactic errors are a ‘natural kind’ of grammatical defect that occur late in language processing. The isolated presence of linear morphosyntactic errors in a number of individuals with post-stroke aphasia motivates the existence of such a category of errors with a dissociable syndrome, corresponding to what is called agrammatism. Further, the presence of several distinct hierarchical processing errors overlapping with semantic errors or bleeding into morphosyntactic processing suggested multiple phenotypes of paragrammatism that should be considered in future work.

References

Bastiaanse, R., Bamyaci, E., Hsu, C. J., Lee, J., Duman, T. Y., & Thompson, C. K. (2011). Time reference in agrammatic aphasia: A cross-linguistic study. *Journal of Neurolinguistics*, *24*(6), 652. https://doi.org/10.1016/J.JNEUROLING.2011.07.001

Beltrami, D., Gagliardi, G., Favretti, R. R., Ghidoni, E., Tamburini, F., & Calzà, L. (2018). Speech analysis by natural language processing techniques: A possible tool for very early detection of cognitive decline? *Frontiers in Aging Neuroscience*, *10*, 414837. https://doi.org/10.3389/FNAGI.2018.00369/BIBTEX

Bird, H., & Franklin, S. (1996). Cinderella revisited: A comparison of fluent and non-fluent aphasic speech. *Journal of Neurolinguistics*, *9*(3), 187–206. https://doi.org/10.1016/0911-6044(96)00006-1

Blumstein, S. E., Cooper, W. E., Goodglass, H., Statlender, S., & Gottlieb, J. (1980). Production deficits in aphasia: A voice-onset time analysis. *Brain and Language*, *9*, 153–170.

Bock, K., & Levelt, W. J. M. (1994). Language production: grammatical encoding. In K. Bock, W. J. Levelt, & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 945–984).

Busby, N., Newman-Norlund, S., Sayers, S., Newman-Norlund, R., Wilmskoetter, J., Rorden, C., Nemati, S., Wilson, S., Riccardi, N., Roth, R., Johnson, L., den Ouden, D. B., Fridriksson, J., & Bonilha, L. (2023). Lower socioeconomic status is associated with premature brain aging. *Neurobiology of Aging*, *130*, 135–140. https://doi.org/10.1016/J.NEUROBIOLAGING.2023.06.012

Caplan, D., Waters, G., Kennedy, D., Alpert, N., Makris, N., Dede, G., Michaud, J., & Reddy, A. (2007). A study of syntactic processing in aphasia II: Neurological aspects. *Brain & Language*, *101*(2), 151–177. https://doi.org/10.1016/j.bandl.2006.06.226

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, *3*(4), 572–582. https://doi.org/10.1016/0093-934X(76)90048-1

Casilio, M., Kasdan, A. V, Bryan, K., Shibata, K., Schneck, S. M., Levy, D. F., Entrup, J. L., Onuscheck, C., de Riesthal, M., & Wilson, S. M. (2024). Four dimensions of naturalistic language production in aphasia after stroke. *Brain*. https://doi.org/10.1093/BRAIN/AWAE195

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

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.

Edmonds, L. A., Nadeau, S. E., & Kiran, S. (2009). Effect of Verb Network Strengthening Treatment (VNeST) on lexical retrieval of content words in sentences in persons with aphasia. *Aphasiology*, *23*(3), 402–424. https://doi.org/10.1080/02687030802291339

Fahey, D., Hickok, G., Fridriksson, J., & Matchin, W. (2024). Lesion-symptom mapping of acceptability judgments in chronic post-stroke aphasia reveals the neurobiological underpinnings of receptive syntax. *Journal of Cognitive Neuroscience*.

Faroqi-Shah, Y. (2013). Selective treatment of regular versus irregular verbs in agrammatic aphasia: Efficacy data. *Aphasiology*, *27*(6), 678–705. https://doi.org/10.1080/02687038.2012.751577

Gleason, J. B., Goodglass, H., Green, E., Ackerman, N., & Hyde, M. R. (1975). The retrieval of syntax in Broca’s aphasia. *Brain and Language*, *2*, 451–471.

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

Goodglass, H. (1997). Agrammatism in Aphasiology. *Clinical Neuroscience*, *4*, 51–56.

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
Goodglass, H., & Hunt, J. (1958). Grammatical Complexity and Aphasic Speech. In Y. Grodzinsky & K. Amunts (Eds.), *Broca's Region* (pp. 369–375). OUP.
Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences*, *23*(1), 1–71. https://doi.org/10.1017/S0140525X00002399
Iorga, M., Higgins, J., Caplan, D., Zinbarg, R., Kiran, S., Thompson, C. K., Rapp, B., & Parrish, T. B. (2021). Predicting Language Recovery in Post-Stroke Aphasia using Behavior and Functional MRI. *Scientific Reports*, *11*(8419), 1–14. https://doi.org/10.1038/s41598-021-88022-z
Kay, J., Lesser, R., & Coltheart, M. (1996). Psycholinguistic assessments of language processing in aphasia (PALPA): An introduction. *Aphasiology*, *10*(2), 159–180. https://doi.org/10.1080/02687039608248403
Kleist, K. (1914). Aphasie und geisteskrankheit. *Munchener Medizinische Wochenschrift*, *6*, 8–12.
Kussmaul, A. (1877). *Handbuch der speciellen Pathologie und Therapie: Die Storungen der Sprache versuch einer Pathologie der Sprache [Handbook of specific pathology and therapy. Disorders of language: an attempt of a pathology of language* (Vol. 12). FCW Vogel.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. The MIT Press.
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
Magnusdottir, S., Fillmore, P., Den Ouden, D.-B., Hjaltason, H., Rorden, C., Kjartansson, O., Bonilha, L., & Fridriksson, J. (2013). Damage to left anterior temporal cortex predicts impairment of complex syntactic processing: A lesion-symptom mapping study. *Human Brain Mapping*, *34*(10), 2715–2723. https://doi.org/10.1002/hbm.22096
Marslen-Wilson, W. D., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and Meaning in the English Mental Lexicon. *Psychological Review*, *101*(1), 3–33.
Martínez-Ferreiro, S., & Bastiaanse, R. (2013). Time reference in Spanish and Catalan non-fluent aphasia. *Lingua*, *137*, 88–105. https://doi.org/10.1016/j.lingua.2013.09.003
Martínez-Ferreiro, S., de Aguiar, V., & Rofes, A. (2015). Non-fluent aphasia in Ibero-Romance: A review of morphosyntactic deficits. *Aphasiology*, *29*(1), 101–126. https://doi.org/10.1080/02687038.2014.958915
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
Matchin, W., den Ouden, D.-B. B., Basilakos, A., Stark, B. C., Fridriksson, J., & Hickok, G. (2023). Grammatical parallelism in aphasia: A lesion-symptom mapping study. *Neurobiology of Language*, *4*(4), 550–574. https://doi.org/10.1162/nol_a_00117
Matchin, W., & Hickok, G. (2020). The cortical organization of syntax. *Cerebral Cortex*, *120*(3), 701–722. https://doi.org/10.31234/osf.io/6394f
Paradis, M. (Ed.). (2001). *Manifestations of Aphasia Symptoms in Different Languages*. Pergamon.
Poulisse, N. (1999). *Slips of the Tongue: Speech Errors in First and Second Language Production*. John Benjamins.
Rogalsky, C., & Hickok, G. (2009). Selective Attention to Semantic and Syntactic Features Modulates Sentence Processing Networks in Anterior Temporal Cortex. *Cerebral Cortex*, *19*(4), 786–796. https://doi.org/10.1093/cercor/bhn126
Stockbridge, M. D., Matchin, W., Walker, A., Breining, B. L., Fridriksson, J., Hickok, G., & Hillis,

A. E. (2021). One cat, two cats, red cat, blue cats: eliciting morphemes from individuals with primary progressive aphasia. *Aphasiology*, 1–12. https://doi.org/10.1080/02687038.2020.1852167

Stockbridge, M. D., Walker, A., Matchin, W., Breining, B. L., Fridriksson, J., Hillis, A. E., & Hickok, G. (2020). A double dissociation between plural and possessive “s”: Evidence from the Morphosyntactic Generation test. *Cognitive Neuropsychology*, *38*(1). https://doi.org/10.1080/02643294.2020.1833851

Swinburn, K., Porter, G., & Howard, D. (2004). *Comprehensive Aphasia Test (CAT)*. https://doi.org/10.1037/t13733-000

Thompson, C. K. (2019). Neurocognitive Recovery of Sentence Processing in Aphasia. *Journal of Speech, Language, and Hearing Research*, *62*(11), 3947.

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.

Turkstra, L. S., Thompson, C. K., Kreutzer, J., DeLuca, J., & Caplan, B. (2010). Agrammatism. In M.-L. Kean (Ed.), *Encyclopedia of Clinical Neuropsychology*. Academic Press. https://doi.org/10.1016/c2013-0-10949-x

Ullman, M. T., Pancheva, R., Love, T., Yee, E., Swinney, D., & Hickok, G. (2005). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language*, *93*(2), 185–238. https://doi.org/10.1016/j.bandl.2004.10.001

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). *Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies*. *35*(3), 407–426. https://doi.org/10.1016/J.NEUBIOREV.2010.04.007

Wilson, S. M., Galantucci, S., Tartaglia, M. C., & Gorno-Tempini, M. L. (2012). The neural basis of syntactic deficits in primary progressive aphasia. *Brain and Language*, *122*(3), 190–198. https://doi.org/10.1016/J.BANDL.2012.04.005