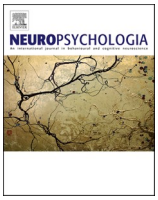
Neuropsychologia 151 (2021) 107713

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: http://www.elsevier.com/locate/neuropsychologia



Quantifying grammatical impairments in primary progressive aphasia: Structured language tests and narrative language production

Jennifer E. Mack [a,*], Elena Barbieri [a], Sandra Weintraub [b,c,d], M.-Marsel Mesulam [b,c], Cynthia K. Thompson [a,b,c]

[a] Roxelyn & Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, USA
[b] Mesulam Center for Cognitive Neurology and Alzheimer's Disease, Northwestern University, USA
[c] Ken & Ruth Davee Department of Neurology, Northwestern University, USA
[d] Department of Psychiatry and Behavioral Sciences, Northwestern University, USA

ARTICLE INFO

Keywords:
Primary progressive aphasia
Syntax
Narrative
Language production

ABSTRACT

Purpose: This study examined grammatical production impairments in primary progressive aphasia (PPA), as measured by structured tests and narrative samples. We aimed to quantify the strength of the relationship between grammatical measures across tasks, and identify factors that condition it. Three grammatical domains were investigated: overall sentence production, verb morphology, and verb-argument structure.

Methods: 77 participants with PPA (34 PPA-G, 16 PPA-L, 15 PPA-S and 12 other) completed a battery of grammatical tests and a narrative language sample was obtained. Accuracy scores were computed for the language tests and the narrative samples were analyzed for both accuracy of selected narrative variables as well as grammatical diversity across the three grammatical domains. Principal components analysis (PCA) and multiple regression were used to examine cross-task relationships for all measures.

Results: As expected on the basis of classification criteria, accuracy scores were lower for the PPA-G group as compared to the PPA-L and PPA-S participants for overall sentence production and verb morphology, but not argument structure. Grammatical accuracy in narratives strongly predicted overall language test performance in PPA-G, whereas grammatical diversity in narratives did so in PPA-L, and no significant correspondence between narrative and language test performance was found for PPA-S. For individuals with severe grammatical impairments only, error distribution for both morphology and argument structure was strongly associated in structured tasks and narratives.

Conclusions: Grammatical production in narrative language predicts accuracy elicited with structured language tests in PPA. However, unique narrative production patterns distinguish PPA by subtype: accuracy for PPA-G, and grammatical diversity for PPA-L. The impairment in PPA-G is likely to reflect a core impairment in grammar whereas that of PPA-L may be closely tied to the word retrieval and verbal working memory deficits that characterize this variant. This underscores the theoretical distinction between PPA-L and PPA-G, as well as the importance of including grammatical diversity measures in analyses of language production, especially for patients who do not display frank agrammatism. Further, the results suggest that measures of domain-specific language deficits (i.e., verb morphology vs. argument structure) are robust across tasks only in individuals with severe grammatical impairments.

1. Introduction

Agrammatism is a core feature of the nonfluent/agrammatic variant of PPA (PPA-G), which constitutes approximately 10–35% cases of PPA (Botha et al., 2015; Harris et al., 2013; Mesulam et al., 2012; Sajjadi, Patterson, Arnold, Watson and Nestor, 2012a; Wicklund et al., 2014) and is associated with distinct patterns of neural degeneration (i.e., atrophy of left inferior frontal regions and dorsal language tracts (see Thompson and Mack, 2014; Wilson et al., 2012, for reviews)). Agrammatism is also prominent in a subset of "mixed" cases of PPA, which are additionally characterized by lexical-semantic deficits (Mesulam and Weintraub, 2014). In contrast, individuals with the logopenic variant of

* Corresponding author. Department of Communication Disorders, University of Massachusetts-Amherst, 358 N. Pleasant St, Amherst, MA, 01003, USA.
E-mail address: jemack@umass.edu (J.E. Mack).

https://doi.org/10.1016/j.neuropsychologia.2020.107713
Received 20 April 2020; Received in revised form 6 November 2020; Accepted 3 December 2020
Available online 5 December 2020
0028-3932/

PPA (PPA-L), characterized by phonological errors and repetition difficulty, and the semantic variant of PPA (PPA-S), associated with impaired word comprehension and naming (Gorno-Tempini et al., 2011), in general, show preserved grammatical ability. Notably, however, several studies have found a greater rate of grammatical errors (e.g., verb morphology errors) (Ash et al., 2013; Wilson et al., 2010), production of fewer syntactically-complex structures (e.g., embedded clauses, left-branching structures, and passives) (Fraser et al., 2014; Knibb et al., 2009; Marcotte et al., 2017; Sajjadi, Patterson, Tomek and Nestor, 2012b), and production of shorter utterances than unimpaired controls in these two groups (Thompson et al., 2012a; Wilson et al., 2010).

For evaluation of grammatical ability in PPA, both free narrative language tasks and structured language measures have been used. Narrative tasks include story telling or re-telling, picture description, interviews, and conversation, eliciting a sequence of related utterances. These samples are then analyzed for both lexical features (e.g., word class usage) and morphosyntax (e.g., grammatical accuracy) (Ash et al., 2013; Fraser et al., 2014; Knibb et al., 2009; Marcotte et al., 2017; Thompson et al., 1997; Thompson et al., 2012a; Thompson et al., 2013; Wilson et al., 2010; see Boschi et al., 2017, for a review). In contrast, structured tests of grammatical production, including anagram (word-ordering) tasks, production priming, and sentence completion, elicit responses to specific syntactic forms (e.g., Cupit et al., 2016; DeLeon et al., 2012; Thompson et al., 2013; Weintraub et al., 2009; Wilson et al., 2014a).

Although both methods have been instrumental in delineating patterns of impaired grammatical production in PPA, they raise an important question: do structured tests and connected language production provide similar pictures of grammatical deficits? Little research has examined this question. Studies on stroke-induced aphasia have shown a high degree of correspondence between the two (Goodglass et al., 1993) and one study with PPA patients found a similar pattern (Billette et al., 2015), however, the latter found differences when comparing language elicitation methods (e.g., picture description vs. semi-structured interviews) (Sajjadi et al., 2012b). These results suggest that, in general, the two tasks measure the same underlying linguistic phenomena, even though the demands of structured tests and connected language production tasks are quite different. By nature, unconstrained, free language production requires all stages of linguistic processing, including conceptual planning, discourse organization, word retrieval and grammatical processing. Structured language tests, in contrast, are designed to evaluate specific linguistic processes, such as word order or verb inflection, and to minimize the influence of other processes. In addition, structured language tests obligate the production of a variety of grammatical structures, some of which are rarely used in free narrative tasks and thus cannot be evaluated using this method. Thus, differences in production accuracy, which is typically the dependent measure in both types of task, reflect the structures that are analyzed (in free narratives) and tested in structured measures.

Another important dimension of language production, in addition to accuracy, concerns the complexity/diversity of structures used. Complexity and diversity are related notions that reflect the distribution of structures produced. With regard to grammatical ability, complexity can be evaluated in free narratives by computing the proportion of complex versus simple structures (e.g., sentence types, verb types) (e.g., Fraser et al., 2014; Knibb et al., 2009; Marcotte et al., 2017; Sajjadi et al., 2012b). Complexity can also be evaluated in structured language tasks if the measures used are designed to measure it. Some tests, for example, elicit production of both simple and complex sentences as well as verbs based on the number of arguments required. Diversity reflects the *range* of structures produced. For example, a high verb-argument diversity score reflects production of a range of argument structures, both simple and complex (e.g., Malyutina, Richardson, & den Ouden, 2016). An individual who has difficulty with grammar may produce an atypically high rate of simple structures, resulting in low complexity and diversity scores. Complexity/diversity in language production may reveal finer-grained differences in grammatical accessibility across PPA subtypes, particularly between PPA-G and PPA-L.

From the perspective of understanding language impairments, it is important to determine whether grammatical impairments are stable across assessment methods; if so, this would suggest that the derived production patterns are not artefacts of task constraints. This is also important for clinical practice. In clinical settings, there typically is not sufficient time to administer structured tests of grammar, and while there may be time to collect a connected language sample, the clinician will likely have limited time to code it. Although automated approaches to coding aphasic language production have been developed (e.g., Fraser et al., 2014; MacWhinney et al., 2010), much grammatical coding is still done manually, especially for complex phenomena such as verb-argument structure (Hsu and Thompson, 2018; but see Fromm et al., 2020). Furthermore, complexity/diversity coding is especially labor-intensive, because it requires coding the grammatical type of each structure in the sample (see Malyutina et al., 2016, for discussion). In contrast, accuracy coding can be done relatively quickly as the researcher or clinician needs to only record the total number of structures and the number of errors. Thus, it is useful to determine whether complexity/diversity measures contribute significantly to the characterization of grammatical impairments, and are therefore worth the time investment it takes to code them.

In addition, a complete understanding of language impairments in PPA requires investigation of multiple grammatical *domains* (i.e., types of grammatical processes). Although PPA-G (aka *nonfluent variant* PPA) is defined by the presence of "agrammatism" (Gorno-Tempini et al., 2011), it is well-established that agrammatism is associated with a constellation of linguistic features that often co-occur, including word order deficits associated with verbs and verb-argument structure (i.e., failure to produce sentences that satisfy the verb's syntactic and semantic combinatory requirements) and grammatical morpheme errors (particularly for verb morphology (e.g., tense and agreement)) (Berndt and Caramazza, 1980; Goodglass et al., 1972; Goodglass and Menn, 1985; Menn et al., 1990; Milman et al., 2008). Accurate word order production requires sensitivity to local and long-distance dependencies (i.e., constraints on the syntactic categories of adjacent words within phrases, and relationships between words that are separated by other linguistic material such as structures with displacement, respectively). In the domain of verb morphology, the morphological form used must be compatible with the intended meaning (i.e., the semantic notion of time reference: the temporal context of an event in the past, present or future (e.g., *Alex skated* vs. *Alex skates*)) and syntactic context (i.e., different verb forms require unique syntactic environments, e.g., *Alex loves to skate* and *Alex loves skating*, although they are more-or-less equivalent in meaning).

Notably, these grammatical domains can be dissociated in individual patients (Bastiaanse &Thompson, 2012; Druks, 2017; Miceli et al., 1983; Saffran et al., 1989; and others; also see Kussmaul (1877) who first described such dissociations and proposed different nomenclature for the two: *aktaphasie*, associated with impairments of word order, and *agrammatismus*, referring to impairments of grammatical morphology). This suggests that *distinct* processes may be impaired in agrammatism, and that the two may be engaged at different stages of language production. The influential model of Bock and Levelt (1994) suggests that assignment of thematic and grammatical roles (important for word order) occurs on the *functional level*, which precedes the *positional* level, at which verb morphology is selected. In the present study, we included measures of both grammatical domains, in order to tease out the contributions of common vs. distinct processes to grammatical impairments in PPA.

The present study used narrative language samples and a battery of structured tests to examine impairments of grammar (overall sentence production, verb morphology, and verb argument structure) in patients with PPA. Accuracy and diversity measures were derived from language samples. We chose to use diversity rather than complexity measures,

because diversity measures provide a closer parallel to structured language tests, in that they emphasize the production of structures of varying complexity.

Our research questions were as follows. First, which grammatical domains are impaired in the major subtypes of PPA, and do impairments manifest themselves as grammatical errors or grammatical simplification? Our criteria for subtype diagnoses included impairment on measures of noncanonical sentence production (impaired in PPA-G; relatively intact in PPA-L and in PPA-S), but verb morphology and argument structure did not contribute to subtype diagnoses. Following previous research, we expected to see deficits in PPA-G with respect to verb morphology and verb argument structure, manifested primarily in grammatical errors. In PPA-L, we expected to find less pronounced grammatical errors and possible grammatical simplification. In PPA-S, we expected to find mild, if any, grammatical impairments, relating to impaired lexical semantics (e.g., impaired production of irregular morphology; Auclair-Ouellet, 2015; Wilson et al., 2014a). Second, how strong is the relationship between grammatical measures derived from structured tests and narratives? We expected to find at least a moderately strong correspondence, reflecting core impairments of the grammatical system evident across tasks. We also investigated key linguistic and individual factors that might condition this relationship, such as the type of narrative variable (grammatical accuracy vs. diversity), grammatical domain, PPA subtype, and severity of grammatical impairments. We also anticipated that the results of the study would have the potential to inform clinical assessment procedures, e.g., how patient- and clinician-time can best be used to provide a robust and detailed picture of grammatical abilities.

2. Material and methods

2.1. Participants

The participants with PPA were identified from a group of 127 patients consecutively enrolled in the Language in PPA project at Northwestern University. Participants were excluded due to incomplete data sets (n = 47; defined as one or more missing grammatical measures or a Cinderella narrative containing fewer than 10 verbs), prominent cognitive impairments as a consequence of advanced disease in addition to aphasia (n = 2), or bilingualism (n = 1), with a total of 77 patients completing the study. We refer to these 77 individuals as the All-PPA group.

In addition, cognitively healthy older adults provided data for each of the primary dependent measures of the study. Partially overlapping groups of healthy controls contributed data for each task: the Northwestern Assessment of Verbs and Sentences (Thompson, 2011; n = 26), Northwestern Anagram Test (Thompson, Weintraub and Mesulam, 2012b; n = 40), Northwestern Assessment of Verb Inflection (Lee and Thompson, 2017; n = 30), and a Cinderella narrative sample (n = 25). All were native English speakers with no history of speech, language, learning, or neurological deficits, and scored within the normal range on a neuropsychological test battery. The study was approved by the Institutional Review Board at Northwestern University and all participants provided informed consent.

The diagnosis of PPA was based on each participant's medical history, neurological examination (conducted by one of the authors, M.-M. M.), and neuropsychological and language test profiles (derived from administration of associated tests by S.W., C.K.T. and colleagues). PPA subtypes were based on word comprehension, grammatical production, and repetition ability (after Mesulam et al., 2012) and were consistent with the 2011 consensus criteria for PPA classification (Gorno-Tempini et al., 2011). Single word comprehension was assessed using a subset of moderately challenging items from the Peabody Picture Vocabulary Test (PPVT; Dunn and Dunn, 2006); grammatical impairments were derived from performance on noncanonical items on the Sentence Production Priming Test (SPPT) of the Northwestern Assessment of Verbs and Sentences (Thompson, 2011) and the Northwestern Anagram Test (Thompson, Weintraub and Mesulam, 2012b), and sentence repetition was assessed using a subset of difficult items from the Repetition Subtest of the Western Aphasia Battery-Revised (WAB-R; Kertesz, 2006). The PPA subtypes were assigned as follows: PPA-G (n = 34): impaired grammar with intact word comprehension; PPA-L (n = 16): relatively intact grammar and word comprehension but impaired repetition (compared to controls); PPA-S (n = 15): impaired word comprehension but relatively intact grammar and repetition. There were also 12 people included in the study whose language profiles fell outside the major PPA subtypes: 5 with mixed PPA (PPA-M, characterized by impaired word comprehension and grammar); 3 with PPA-GSp (predominant motor speech impairments with variable agrammatism), 3 with language profiles that were not clearly differentiated between PPA-G and PPA-L, and 1 unclassifiable participant whose primary presenting symptom was anomia.

Demographic information is provided in Table 1. The All-PPA group did not differ from any of the task-specific control groups with respect to age, years of education, handedness (measured using the Edinburgh Handedness Inventory; Oldfield, 1971) (Mann-Whitney U, $p > 0.05$ (FDR)) or gender (Fisher's Test, $p > 0.05$ (FDR)). In addition, the PPA-G, PPA-L, and PPA-S groups did not differ from each other or from any of the task-specific control groups on these measures. The approximate duration of symptoms of aphasia also did not differ between PPA-G, PPA-L, and PPA-S (Mann-Whitney U, $p > 0.05$ (FDR)).

Each PPA participant was administered a battery of cognitive and language measures, which indicated that aphasia was the primary presenting cognitive symptom. Table 2 provides WAB-R Aphasia Quotients (AQs) as well as performance patterns derived from the language measures administered. Overall aphasia severity was greater in PPA-G and PPA-S than in PPA-L (WAB-R AQ). No significant differences were found between PPA-G and PPA-L with respect to repetition ability (WAB-R Repetition subtest), confrontation naming (Boston Naming Test (BNT); Kaplan et al., 1983), word comprehension (PPVT), and semantic association (Pyramids and Palm Trees Test (pictures version); Howard and Patterson, 1992). In contrast, repetition performance was better in PPA-S than in PPA-G and PPA-L, whereas confrontation naming, word comprehension, and semantic association were more impaired in PPA-S. The All-PPA, PPA-S, and PPA-G groups were significantly impaired relative to age-matched controls on these measures and the PPA-L group was significantly impaired relative to controls only on repetition and naming.

As a direct consequence of our classification criteria, the PPA-G group was significantly impaired on grammatical production (noncanonical items from the SPPT and NAT) as compared to PPA-L, PPA-S, and controls. Grammatical production, though relatively preserved in PPA-L and PPA-S, was significantly impaired relative to controls.

2.2. Procedure

Participants were administered four structured tests of grammatical production, and produced a connected language sample (Cinderella story re-tell).

2.2.1. Structured tests of grammar

Two tests were administered to examine overall sentence production: the Sentence Production Priming Test (SPPT) of the Northwestern Assessment of Verbs and Sentences (NAVS; Cho-Reyes and Thompson, 2012; Thompson, 2011) and the Northwestern Anagram Test (NAT; Thompson et al., 2012b; Weintraub et al., 2009). As previously noted, the noncanonical items on these tests were used as measures of grammatical ability for PPA subtype classification.

The SPPT tests the ability to verbally produce sentences of varying complexity. In each trial, the participant was provided with a pair of semantically reversible action pictures side-by-side (e.g., left picture: a dog chasing a cat; right picture: a cat chasing a dog). The experimenter

Table 1
Participant variables by PPA subtype and for healthy controls.

Group		Age (Years)		Symptom Duration (months)		Education (Years)		Gender		Handedness	
	N	M	SD	M	SD	M	SD	F	M	M	SD
Agrammatic	34	65.8	6.9	39.2	21.7	15.8	2.2	16	18	94.4	9.8
Logopenic	16	68.4	6	39.5	17	17.4	1.5	8	8	94.1	9.3
Semantic	15	65	5.9	45.3	21.6	15.9	3.0	5	10	93.7	9.0
All-PPA	77	66.6	6.9	40.6	19.4	16.1	2.4	34	43	93.8	10
Controls: NAVS	26	63.8	7.9	NA	NA	15.8	2.4	12	14	93.3	10
Controls: NAT	40	64.6	7.9	NA	NA	15.9	2.3	18	22	94.5	8.6
Controls: NAVI	30	63.3	7.7	NA	NA	15.8	2.5	14	16	94.2	9
Controls: Narratives	25	63.6	7.8	NA	NA	15.8	2.3	12	13	93.6	9.4

Note: NAVS = Northwestern Assessment of Verbs and Sentences (Thompson, 2011); NAT = Northwestern Anagram Test (Thompson et al., 2012b; Weintraub et al., 2009); NAVI = Northwestern Assessment of Verb Inflection (Lee and Thompson, 2017); F = female; M = male.

Table 2
Scores from language measures for PPA and control participants.

Measure (maximum score)	WAB-R AQ (100)		WAB-R Repetition (subset) (66)		BNT (60)		PPVT (36)		PPT Pictures (52)		SPPT/NAT Noncanonical Items (100%)	
Group	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Agrammatic	84.8^{L***}	8.0	$49.1^{S**, C***}$	11.8	44.1^{C***}	12.7	33.5^{C***}	2.2	49.2^{C***}	2.5	$54.8^{L***, S***, C***}$	23.5
Logopenic	92.9	3.7	$53.3^{S**, C***}$	7.8	51.6^{C***}	8.3	34.6	1.4	49.9	2.1	83.7^{C***}	11.7
Semantic	82.6^{L**}	13.7	59.2^{C***}	4.1	$9.7^{G***, L***, C***}$	5.3	$20.6^{G***, L***, C***}$	7.1	$43.6^{G***, L***, C***}$	5.1	90.7^{C***}	10.1
All-PPA	86.0	9.4	51.7^{C***}	10.4	37.7^{C***}	19.0	30.5^{C***}	7.0	48.1^{C***}	4.1	70.1^{C***}	24.1
Controls	NA	NA	65.1	1.6	58	1.8	35.2	1.2	50.9	1.1	98.3	3.9

Note: WAB-R = Western Aphasia Battery-Revised; AQ = Aphasia Quotient; BNT = Boston Naming Test; PPVT = Peabody Picture Vocabulary Test (items 157–192); PPT = Pyramids and Palm Trees Test; SPPT = Sentence Production Priming Test from the Northwestern Assessment of Verbs and Sentences; NAT = Northwestern Anagram Test. $^{C/G/L/S}$ = impaired relative to Controls/PPA-G/PPA-L/PPA-S at the FDR-adjusted significance level of $p < 0.05$ (*), $p < 0.01$ (**), or $p < 0.001$ (***).

produced a prime sentence to describe the picture on the left (e.g., *Pete saw the cat who the dog chased*) and the participant was asked to produce a sentence of the same structure for the picture on the right (e.g., *Pete saw the dog who the cat chased*). There were a total of 30 items, five each for the following structures: active, passive, subject *wh*-question, object *wh*-question, subject-relative clause, and object-relative clause. Responses were scored as correct when they exactly matched the target sentence, or did so apart from (1) phonological paraphasias, (2) semantically-appropriate word substitutions, (3) grammatically-appropriate substitutions of verb tense, and/or (4) grammatically-appropriate substitution or omission of the relative pronoun (in relative clauses). Correct passive sentences (e.g., *The dog was chased by the cat*) require at least two of three passive morpho-syntactic indicators (auxiliary verb, main verb in past-participle form, agentive by-phrase). In previous studies, greater deficits in SPPT performance have been observed in participants with agrammatic vs. anomic stroke-induced aphasia (Cho-Reyes and Thompson, 2012), and in participants with PPA-G as compared to PPA-L, particularly for more complex (non-canonical) forms (Thompson et al., 2013).

The NAT examines the ability to arrange words into a grammatical sentence, without requiring spoken language production, making it particularly appropriate for participants with severe motor speech impairments. Participants were presented with a picture (e.g., a boy pulling a girl) and the base form of the verb and the event participants (arguments) labelled (e.g., *pull, boy, girl*). They were also presented with a set of word cards, each containing one printed word from the target sentence. The experimenter selected words to begin the sentence (e.g., passive: The girl) and the participant was asked to arrange the remaining word cards to produce a grammatical sentence matching the event picture. There were 30 items, five for each of the following structures: active, passive, subject *wh*-question, object *wh*-question, subject-cleft, and object-cleft. Previous research on PPA has shown that performance on the NAT and SPPT is highly correlated (Weintraub et al., 2009).

The Northwestern Assessment of Verb Inflection (NAVI; Lee and Thompson, 2017) tests production of verb morphology. Participants were presented with a picture (e.g., a woman pouring a glass of water), the base form of the verb (e.g., *pour*), and a sentence template with the verb missing (e.g., *Yesterday the lady __ the water*). Participants were asked to produce the full sentence with the correct form of the verb on 60 trials, 10 from each of the following conditions: past-tense, present-tense third-person singular, present-tense third-person plural, infinitive, present-progressive, and future. For the present study, the future condition was excluded, because it required production of a modal and a main verb (e.g., *will eat*), whereas all other conditions required only production of a main verb. Responses were scored as correct if the participant produced the correct verb form.

The Argument Structure Production Test (ASPT) of the NAVS tests the ability to use verb argument structure information to arrange words into a grammatical sentence. Participants were presented with an action picture (e.g., a dog biting a cat), with the base form of the verb and each of the arguments labelled (e.g., *bite, cat, dog*). The participant was asked to spontaneously produce a sentence to describe the picture. There were 32 items: 5 one-place verbs, 15 two-place verbs, and 12 three-place verbs. Responses are scored as correct when the verb and arguments were produced in the correct order. A previous study demonstrated greater impairments on the ASPT (particularly for three-argument verbs) for individuals with agrammatic vs. anomic stroke-induced aphasia (Cho-Reyes and Thompson, 2012).

2.2.2. *Cinderella narrative*

Participants were presented with a wordless picture book of the Cinderella story to examine. Then, the book was put away and the participant narrated the story. The resulting language sample was recorded and transcribed verbatim, including paraphasias and neologisms. The sample was segmented into utterances, using semantic, syntactic, and prosodic criteria. The following were transcribed but excluded from the analysis: conjunctions (*and, but*) at the beginning of

an utterance, meta-linguistic or meta-cognitive comments (e.g., *I can't remember her name*), and disruptions to fluency such as filled pauses, false starts, and word-level repetitions.

Then, the narrative samples were coded according to the NNLA (Hsu and Thompson, 2018; Thompson et al., 2012a). The lexical and syntactic properties of each utterance were coded on five levels: the *utterance level*, with utterances coded as ungrammatical when they contained a grammatical or logical-semantic error; the *sentence level*, with main clauses coded as active or passive, and embedded clauses coded as relative, complement, or adjunct clauses; the *lexical level* which categorized each word produced by syntactic categories; the *verb morphology level*, which coded the presence of inflectional morphemes and errors in verb morphology and a verb morphology index (VMI), reflecting the density of the verb phrase produced; and the *verb argument structure level*, which coded verbs by type (e.g., copula, one-place, two-place, three-place, sentential). In the present study, all transcription and coding procedures were performed by at least two researchers, with disagreements resolved by consensus.

To quantify overall grammatical productivity in each group, we compared the mean number of clauses produced between the PPA groups and controls using the Mann-Whitney *U* Test with FDR correction for multiple comparisons. Control participants produced a significantly greater number of clauses (Mean (standard deviation) = 45.4 (15.4)) than the All-PPA group (34.8 (21.7); $p < 0.01$), PPA-G group (34.5 (21.8); $p < 0.01$), and PPA-S group (25.4 (15.0); $p < 0.01$). In addition, the participants with PPA-L produced, on average, more clauses than those with PPA-S ($p < 0.05$). No other significant group differences were found.

Following linguistic coding, the percentage of grammatical sentences was computed (as a measure of overall sentence production accuracy) and the diversity of sentential forms produced was quantified, resulting in a Clause Diversity Index, reflecting the relative proportions of active matrix clauses, passive matrix clauses, relative clauses, complement clauses, and adjunct clauses. Diversity was computed using a Shannon H index (Shannon and Weaver, 1949); a higher value indicated a relatively even distribution of the five clause types, thus greater diversity. For verb morphology, two measures were computed. The accuracy measure computed the rate of grammatical morpheme errors per verb, including errors of verb inflection, auxiliaries, modals, and infinitival "to"; the error rate was then subtracted from 1, and the resulting value multiplied by 100, to keep the direction and scale of this value consistent with other accuracy measures. The verb morphology Diversity Index was the Shannon H value quantifying the distribution of VMI values for correctly-inflected verbs (1, 2, 3, and 4+). Similarly, for argument structure, the percentage of verbs with correct argument structure was computed and a Diversity Index was generated (i.e., the Shannon H value quantifying the distribution of correctly-produced copula, one-place, two-place, three-place, and sentential verbs; cf. Malyutina et al., 2016). These measures are summarized in Table 3.

Table 3
Grammatical measures from language tests and narratives included in the present study.

Measure Type	Sentence Production	Verb Morphology	Verb-Argument Structure
Language Tests: Accuracy (% Correct)	SPPT Total Score NAT Total Score	NAVI Total Score	ASPT Total Score
Narratives: Accuracy	% Grammatical Sentences	Accuracy ((1-error rate)*100)	% Verbs with Correct Verb Arguments
Narratives: Diversity	Clause Diversity Index	Diversity Index	Diversity Index

Note: SPPT = Sentence Production Priming Test (Northwestern Assessment of Verbs and Sentences (Thompson, 2011)); NAT = Northwestern Anagram Test (Thompson et al., 2012b; Weintraub et al., 2009); NAVI = Northwestern Assessment of Verb Inflection (Lee and Thompson, 2017); ASPT = Argument Structure Production Task (Northwestern Assessment of Verbs and Sentences).

2.3. Statistical analyses

Statistical analyses were conducted using R (R Core Team, 2020). For the dependent measures listed in Table 3, we tested for differences between PPA subtypes, and between controls and PPA (PPA-G, PPA-L, PPA-S, and All-PPA), using pairwise Mann-Whitney U Tests, with p-values adjusted for multiple comparisons using the FDR method. Significant differences between the PPA-G and PPA-L/PPA-S groups for the SPPT and NAT measures were expected, given that performance on the noncanonical items on these tests contributed to subtype classification.

To examine cross-task relationships between grammatical measures, pairwise Spearman correlations between the grammatical accuracy measures from structured tests and narrative samples were conducted within the All-PPA group, with FDR-adjusted p-values. Interpretations of correlation strength were based on the criteria proposed by Cohen (1988) and Rosenthal (1996): moderate if $r > 0.3$; strong if $r > 0.5$, and very strong if $r > 0.7$.

To quantify common and distinct impairments across grammatical domains, principal components analyses (PCAs) were conducted for each of the following sets of grammatical measures: accuracy on structured tests, narrative accuracy, and narrative diversity. The *princomp* function in R was used, which employs spectral decomposition to compute unrotated principal components. These computations were performed on the correlation matrix in order to scale the variables. The variance accounted for by the first two principal components (PCs), as well as the loadings of each measure on the PCs were computed. An absolute value threshold of 0.4 or greater was used to identify meaningful factor loadings. When two or more measures with strong loadings have the same sign (e.g., 0.5 for measure 1 and 0.7 for measure 2), this suggests that the PC reflects performance on both measures. Differences in sign between two strongly-loading measures (e.g., 0.5 for measure 1 and -0.7 for measure 2) indicates that the PC reflects the difference in performance between those measures.

The components identified in the PCAs were entered into multiple linear regression analyses, in which narrative measures were used to predict structured test measures. The first set of regression analyses tested whether overall narrative accuracy and/or diversity predict overall test scores. For these analyses, the first principal components from the narratives (Narrative Accuracy PC1 and Narrative Diversity PC1) were used to predict the first principal component of test performance (Structured Test PC1). The second set of regression analyses tested whether the specific pattern of errors identified in the narratives (i.e., the relative distribution of verb morphology and argument structure errors, as reflected by Narrative Accuracy PC2) predicted the specific pattern of performance in structured tests (Structured Test PC2). To test for the potential effects of severity of agrammatism, Narrative Accuracy PC1 and its interaction with Narrative Accuracy PC2 were included in these models. The analyses were conducted first for the All-PPA group and then for just the participants with a PPA-G, PPA-L, or PPA-S diagnosis. In the latter analyses, we tested for interactions between PPA subtype and narrative variables in predicting structured test scores. In all multiple regression analyses, continuous variables were centered and categorical variables were simple-coded (i.e., each level of the factor was compared to a reference level, with the intercept being the grand mean). A step-wise backwards model comparison procedure (in which minimally-different nested models were compared using the ANOVA test, with a threshold of $p < 0.1$) was used to select the best-fitting model for each data set. The strength of each model (adjusted R^2) was interpreted by adapting the criteria of Cohen (1988) and Rosenthal (1996): moderate if $R^2 > 0.09$; strong if $R^2 > 0.25$, and very strong if $R^2 > 0.49$.

3. Results

3.1. Group differences in grammatical production measures

Table 4 provides group summary scores and statistics for the grammatical measures from structured tests and narratives. We note that for the structured tests of overall sentence production (SPPT and NAT), the finding of greater impairments in PPA-G vs. PPA-L, PPA-S and controls is a direct consequence of the subtype classification criteria. For structured test and narrative accuracy measures, the PPA groups (PPA-G, PPA-L, PPA-S, All-PPA) were significantly impaired relative to controls on all measures, with the exception of the SPPT and verb morphological accuracy in narratives for PPA-S. For narrative diversity, the PPA-G and All-PPA groups were impaired relative to controls on Clause Diversity. As compared to the PPA-L and PPA-S groups, the PPA-G group showed greater impairments on structured test and narrative accuracy measures for overall sentence production and verb morphology, but not verb argument structure. In addition, the PPA-G group showed reduced Clause Diversity (this comparison was fully significant for PPA-S and marginally so for PPA-L). No other significant group differences were observed.

3.2. Correlations between grammatical measures

Correlations between grammatical accuracy measures from structured tasks and narratives, in the All-PPA group, appear in Table 5. Significant correlations were found between the SPPT, NAT, and NAVI and two measures of narrative accuracy: the proportion of grammatical sentences and the proportion of verbs with correct argument structure. Of these, the correlations between the proportion of grammatical sentences in narratives and the NAVS and NAT were strong ($r \geq 0.5$), whereas the rest were of moderate strength ($r > 0.3$). There was also a moderately strong ($r > 0.3$) correlation between the NAT and verb morphology accuracy in narratives.

Table 5
Correlations between grammatical accuracy measures from structured tests and narrative samples.

		Structured Tests			
		SPPT	NAT	NAVI	ASPT
Narrative Accuracy Measure	Grammatical Sentences	**0.53**	**0.50**	**0.38**	0.15
	Verb Morphology	0.21	**0.33**	0.21	−0.13
	Verb Argument Structure	**0.35**	**0.34**	**0.38**	0.15

Note: Spearman R values are presented. **BOLD** = $p < 0.05$ (FDR). SPPT = Sentence Production Priming Test of the NAVS; NAT = Northwestern Anagram Test; NAVI = Northwestern Assessment of Verb Inflection; ASPT = Argument Structure Production Task of the NAVS.

3.3. Principal components analyses

The results of the principal components analyses appear in Table 6. For the structured test measures, the first principal component (Structured Test PC1) accounted for 60% of the variance and loaded on all four variables. The second component (Structured Test PC2), accounting for 22% of the variance, loaded on the difference between ASPT (argument structure) and NAVI (verb morphology) scores. For narrative accuracy, the first component (Narrative Accuracy PC1) loaded on all three accuracy scores, and reflected 74% of the total variance. The second component (Narrative Accuracy PC2) reflected 19% of the total variance and loaded on the difference between verb morphology and verb argument structure accuracy. For narrative diversity, the first component (Narrative Diversity PC1) loaded on all three measures and comprised 50% of the variance. The second component (Narrative Diversity PC2) loaded on the difference between diversity in the domains of morphology and argument structure, and encompassed 31% of the variance.

Table 4
Grammatical measures: Summary scores and statistical comparisons by group.

Structured Tests (% Correct)								
	SPPT		NAT		NAVI		ASPT	
Group	M	SD	M	SD	M	SD	M	SD
Agrammatic	$70.8^{C***, L*, S***}$	27.9	$67.3^{C***, L***, S***}$	18.6	$79.0^{C***, L*, S*}$	16.0	93.3^{C***}	14.2
Logopenic	$89.0^{C***, S+}$	16.4	89.8^{C***}	7.4	89.6^{C**}	10.3	96.9^{C***}	4.3
Semantic	96.4^{C+}	6.1	92.7^{C**}	10.6	90.5^{C***}	6.6	96.3^{C***}	2.9
All-PPA	80.9^{C***}	23.7	78.3^{C***}	18.9	84.2^{C***}	13.9	94.3^{C***}	12.7
Control	99.6	1.4	98.3	3.9	95.5	11.8	99.9	0.6
Narrative Accuracy								
	GS				VM Accuracy		Correct VAS	
Group	M	SD			M	SD	M	SD
Agrammatic	$61.0^{C***, L**, S*}$	20.7			$91.7^{C***, L*, S**}$	8.6	93.8^{C***}	6.9
Logopenic	79.1^{C***}	9.3			97.1^{C**}	3.1	97.1^{C***}	3.3
Semantic	77.1^{C***}	11.4			98.1	3.5	96.3^{C**}	5.7
All-PPA	69.8^{C***}	18.0			94.9^{C***}	6.8	95.3^{C***}	5.7
Control	93.7	4.7			99.6	0.8	99.8	0.7
Narrative Diversity (Shannon H)								
	Clause				VM		VAS	
Group	M	SD			M	SD	M	SD
Agrammatic	$0.40^{C**, L+, S**}$	0.19			0.82	0.12	1.33	0.20
Logopenic	0.50	0.20			0.82	0.13	1.40	0.12
Semantic	0.53	0.10			0.88	0.15	1.32	0.20
All-PPA	0.46^{C*}	0.18			0.83	0.13	1.34	0.17
Control	0.59	0.19			0.83	0.09	1.39	0.09

Note: SPPT = Sentence Production Priming Test; NAT = Northwestern Anagram Test; NAVI = Northwestern Assessment of Verb Inflection; ASPT = Argument Structure Production Task; GS = grammatical sentences; VM = verb morphology; VAS = verb-argument structure; $^{C/L/S}$ = impaired relative to Controls/PPA-L/PPA-S at the FDR-adjusted significance level of $p < 0.1$ ($^{+}$), $p < 0.05$ (*) , $p < 0.01$ (**), or $p < 0.001$ (***).

Table 6
Results of principal components analyses.

Structured Tests			Narrative Accuracy			Narrative Diversity		
Component	PC1	PC2		PC1	PC2		PC1	PC2
Proportion of variance	0.60	0.22	Proportion of variance	0.74	0.19	Proportion of variance	0.50	0.31
Cumulative variance	0.60	0.81	Cumulative variance	0.74	0.93	Cumulative variance	0.50	0.81
Loadings			Loadings			Loadings		
SPPT	**0.55**	−0.24	GS	**0.63**		Clause	**0.68**	
NAT	**0.55**	0.27	VM	**0.55**	**0.71**	VM	**0.44**	**−0.84**
NAVI	**0.46**	**0.62**	VAS	**0.55**	**−0.71**	VAS	**0.59**	**0.54**
ASPT	**0.44**	**−0.69**						

Note: SPPT = Sentence Production Priming Test from NAVS (Thompson, 2011); NAT = Northwestern Anagram Test (Thompson et al., 2012b); NAVI = Northwestern Assessment of Verb Inflection (Lee and Thompson, 2017); ASPT = Argument Structure Production Task from NAVS; GS = grammatical sentences; VM = verb morphology; VAS = verb-argument structure. BOLD = relatively strong loadings (absolute value > 0.4).

3.4. Multiple regression analyses

We now turn to the multiple regression models in which the principal components of the narrative accuracy and diversity measures were used to predict those from structured tests. Table 7 summarizes the results of the regression model in which Narrative Accuracy PC1 and Narrative Diversity PC1 (overall grammatical accuracy and diversity, respectively) were used to predict Structured Test PC1 (overall accuracy). For the All-PPA group, the best-fitting model had an adjusted R^2 of 0.269 ($p < 0.001$), indicating a strong model fit. It contained Narrative Accuracy PC1 as a significant predictor ($p < 0.001$) but not Narrative Diversity PC1. This indicates that in the All-PPA group, better accuracy in narratives predicted better performance in language tests, but greater diversity in narratives did not. For the group containing only the participants from the three major PPA subtypes (PPA-G, PPA-L, and PPA-S), the best-fitting model had an adjusted R^2 of 0.476 ($p < 0.001$), indicating a strong model fit. It contained main effects of Narrative Accuracy PC1 ($p < 0.001$), Subtype (PPA-S > PPA-G ($p = 0.001$); PPA-L > PPA-G ($p < 0.05$)), and a significant interaction between Narrative Diversity PC1 and Subtype for the comparison of PPA-G vs. PPA-L ($p < 0.05$). Follow-up model comparisons conducted within each subtype indicated that only Narrative Accuracy PC1 significantly predicted test performance in PPA-G ($p = 0.001$; adjusted $R^2 = 0.288$; a strong model fit) and only Narrative Diversity PC1 did so in PPA-L ($p = 0.001$; adjusted $R^2 = 0.506$; a very strong model fit). In contrast, neither narrative component significantly predicted test performance in PPA-S. Scatterplots illustrating the significant relationships found between narrative variables and Structured Test PC1 in PPA-G and PPA-L are provided in Fig. 1.

Table 7
Multiple regression results: Narrative Accuracy and Diversity PC1s as predictors of Structured Test PC1.

All-PPA	β	SE	t	p
Intercept	<0.001	0.151	<0.001	1.000
Accuracy PC1	0.546	0.101	5.383	<0.001
Diversity PC1				
PPA-G, PPA-L, & PPA-S				
Intercept	0.225	0.152	1.478	0.145
Accuracy PC1	0.456	0.107	4.274	<0.001
Diversity PC1	0.082	0.130	0.633	0.529
Subtype (G vs. L)	0.845	0.359	2.355	0.022
Subtype (G vs. S)	1.271	0.372	3.416	0.001
Subtype (L vs. S)	0.426	0.406	1.048	0.299
Accuracy PC1 * Subtype				
Diversity PC1 * Subtype (G vs. L)	0.529	0.258	2.046	0.045
Diversity PC1 * Subtype (G vs. S)	0.104	0.333	0.311	0.757
Diversity PC1 * Subtype (L vs. S)	−0.425	0.352	−1.207	0.232
PPA-G				
Intercept	−0.415	0.248	−1.671	0.104
Accuracy PC1	0.489	0.129	3.787	0.001
Diversity PC1				
PPA-L				
Intercept	0.663	0.148	4.486	0.001
Accuracy PC1				
Diversity PC1	0.432	0.107	4.043	0.001
PPA-S				
Intercept	1.059	0.121	8.770	<0.001
Accuracy PC1				
Diversity PC1				

Note: Blank cells indicate predictors that were excluded from the best-fitting model. β = beta-weight (estimate of co-efficient); SE = standard error. For comparisons between two levels of a factor (e.g., G vs. L), the reference level is listed first and the comparison level second.

Table 8 summarizes the results of multiple regression analyses in which narrative accuracy and diversity PCs were used to predict Structured Test PC2, which measured the relative degree of impairment for verb argument structure (ASPT) vs. verb morphology (NAVI). In the All-PPA group, the best fitting model (adjusted $R^2 = 0.205$; $p < 0.01$; a moderately strong model fit) contained a significant interaction between Narrative Accuracy PC1 and PC2 ($p < 0.01$). In follow-up analyses, participants were split into two groups based on their overall degree of agrammatism in narratives. We defined a "severe agrammatism" group consisting of the 13 individuals whose Narrative Accuracy PC1 scores were less than −1.2. This group consisted of 11 people with PPA-G, one with PPA-S, and one with mixed PPA. In this group, Narrative Accuracy PC2 significantly predicted Structured Test PC2 (adjusted $R^2 = 0.607$; $p < 0.01$; a very strong model fit), whereas for those with milder or no grammatical impairments (Narrative Accuracy PC1 > −1.2; n = 64), these variables were unrelated (adjusted $R^2 = -0.015$; $p > 0.1$; a weak model fit). Fig. 2 illustrates the relationship between Narrative Accuracy PC2 and Structured Test PC2 in these two groups of participants. There were no other significant effects, including interactions with PPA subtype.

4. Discussion

The aims of this study were to quantify the performance of individuals with PPA in multiple domains of grammatical production, and to determine the extent to which structured language tests and narrative production provide a similar picture of grammatical impairments. Seventy-seven individuals with PPA completed a battery of structured language tests that probed overall sentence production ability, verb morphology, and verb-argument structure. Narrative language samples were collected and analyzed for the accuracy and diversity of grammatical production in the same domains. Consistent with the subtype classification criteria, the results indicated impairments in grammatical production across domains in PPA-G, with milder impairments observed in individuals with PPA-L and PPA-S. Grammatical production measures from language tests and narratives were generally in close correspondence, although the nature of the relationship was conditioned by several linguistic and individual factors. These results provide preliminary insights into how grammatical impairments can be assessed most effectively within the limited time afforded in clinical settings.

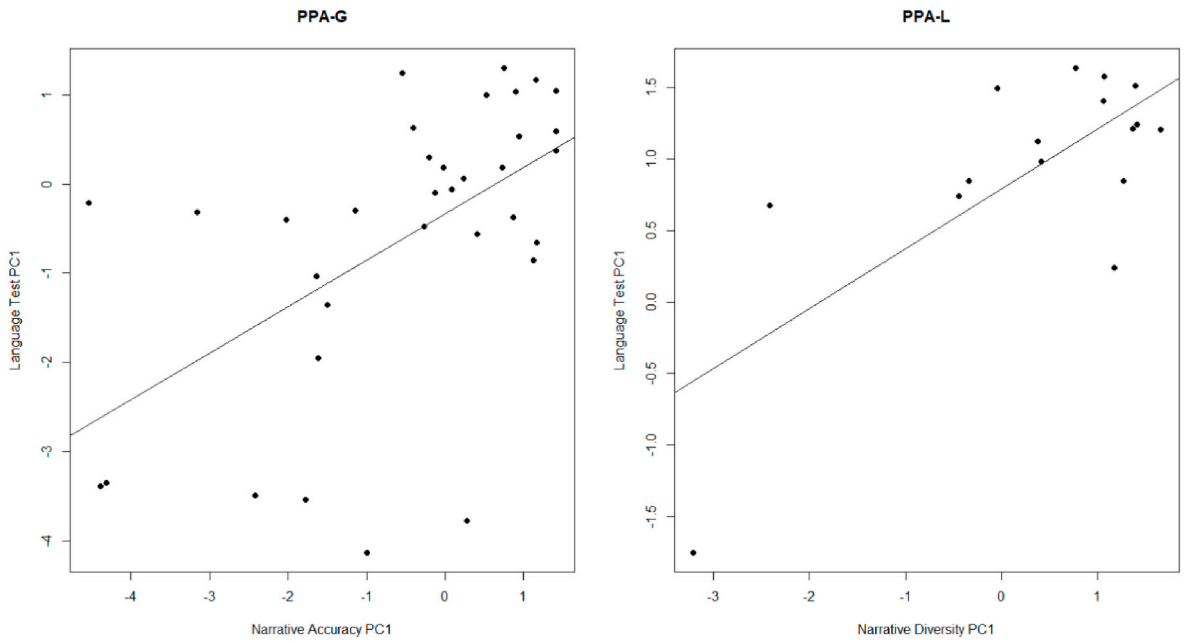


Fig. 1. Multiple regression analyses. The left panel displays the significant positive relationship between Narrative Accuracy PC1 and Structured Test PC1 in PPA-G. The right panel displays the significant positive relationship between Narrative Diversity PC1 and Structured Test PC1 in PPA-L.

Table 8
Multiple regression results. In these analyses, narrative accuracy and diversity PCs were used to predict Structured Test PC2, which reflected differences in performance between argument structure and verb morphology.

All PPA	β	SE	t	p
Intercept	0.000	0.096	<0.001	1.000
Narr. Accuracy PC2	−0.003	0.179	−0.018	0.986
Narr. Accuracy PC1	0.122	0.073	1.673	0.099
Narr. Diversity PC2	0.026	0.112	0.236	0.814
Narr. Diversity PC1	0.068	0.087	0.781	0.438
Narr. Accuracy PC2 * Narr. Accuracy PC1	−0.247	0.085	−2.896	0.005
Narr. Diversity PC2 * Narr. Diversity PC1	0.148	0.079	1.875	0.065
13 PPA with severe agrammatism (Narr. Accuracy PC1 < −1.2)				
Intercept	−0.589	0.201	−2.936	0.014
Narr. Accuracy PC2	0.691	0.156	4.419	0.001
64 PPA with mild or no grammatical impairments (Narr. Accuracy PC1 > −1.2)				
Intercept	0.107	0.108	0.988	0.327
Narr. Accuracy PC2	0.039	0.182	0.215	0.831

Note: β = beta-weight (estimate of co-efficient); SE = standard error.

4.1. Grammatical performance across PPA subtypes

In the present study, the PPA-G group was classified on the basis of impaired performance on structured tests of overall sentence production (NAT and NAVS SPPT noncanonical items), and the PPA-L and PPA-S groups on the basis of relatively preserved performance on these measures. As compared to healthy age-matched controls, individuals with PPA-G also showed impaired accuracy in verb morphology and argument structure, as well as reduced diversity of sentence form (Clause Diversity). These findings are consistent with previous research that investigated these grammatical domains in PPA-G (Ash et al., 2013; DeLeon et al., 2012; Knibb et al., 2009; Sajjadi et al., 2012b; Thompson et al., 2012a, 2013; Wilson et al., 2010). As compared to participants with PPA-L and PPA-S, those with PPA-G showed greater impairment in verb morphology accuracy; previous studies comparing verb morphology across these groups have yielded mixed results (e.g., Thompson et al., 2013; Wilson et al., 2014a). In addition, Clause Diversity was lower in PPA-G than in PPA-S (and marginally lower than in PPA-L), indicating that speakers with PPA-G show a greater degree of syntactic simplification (cf. Ash et al., 2013; Wilson et al., 2010).

Although the classification criteria for PPA-L and PPA-S include relatively preserved grammar, the present study revealed mild impairments in grammatical performance. As compared to unimpaired controls, individuals with PPA-L and PPA-S showed impairments of overall sentence production and verb morphology. In addition, the PPA-L and PPA-S groups showed impairments of verb argument structure that did not differ significantly in severity from PPA-G, in contrast with one previous study which found greater impairments in PPA-G (Thompson et al., 2012a).

These results converge with previous studies which have found evidence for mild impairments in grammatical accuracy in PPA-L, even though these patients are not "frankly agrammatic" (Ash et al., 2013; Wilson et al., 2010, 2014a). That is, although grammatical impairment is not a clinically salient feature of language production in PPA-L, in contrast with word-retrieval difficulty, it is nevertheless evident in some quantitative measures. Similarly, Teichmann et al. (2013) diagnosed PPA-L on the basis of language samples in which word-finding pauses were at least four times as common as syntactic errors; nevertheless, 40% of patients had syntactic impairments. These impairments may arise through damage to left temporo-parietal gray matter and adjacent dorsal white matter pathways, which is characteristic of PPA-L (Galantucci et al., 2011; Gorno-Tempini et al., 2004; Teichmann et al., 2013). Interestingly, speakers with PPA-L did not differ significantly from controls with respect to measures of grammatical complexity in narratives, in contrast with the clinical impression of language in PPA-L as "syntactically simple" (Wilson et al., 2012). These findings demonstrate that production of syntactically simple language is not a consistent phenotype of the subtype.

Grammatical production profiles in PPA-S were very similar to PPA-L, indicating mild impairments at the group level. Historically, much of the research on grammatical impairments in PPA-S (and the related condition of semantic dementia) has focused on irregular verb inflection, which relies on lexical-semantic knowledge (see Auclair-Ouellet, 2015 for review). However, there is also evidence for impaired inflection of *regular* verbs, possibly due to impaired concepts for time reference (Auclair-Ouellet et al., 2016), as well as syntactic impairments affecting word order (Cupit et al., 2016). The present study indicated that, on average, individuals with PPA-S show mild grammatical impairments across domains. Notably, there was substantial individual variability

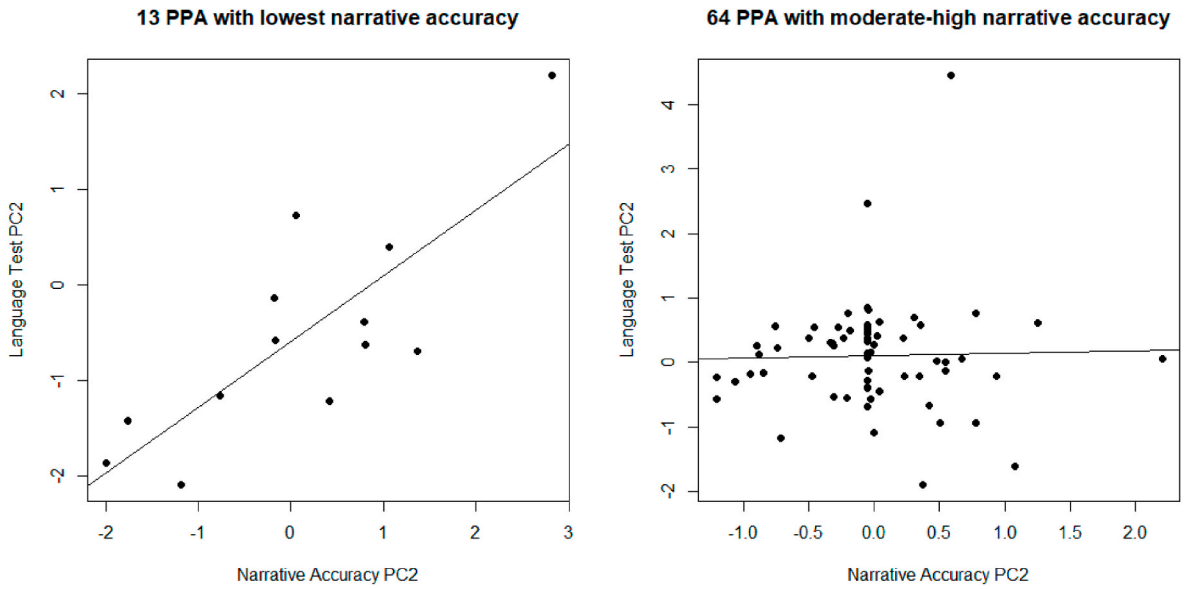


Fig. 2. Multiple regression results. The left panel shows the significant positive relationship between Narrative Accuracy PC2 and Structured Test PC2 in the 13 PPA patients with severe agrammatism (Narrative Accuracy PC2 < -1.2). The right panel shows the absence of a significant relationship between Narrative Accuracy PC2 and Structured Test PC2 in participants with milder or no grammatical impairments (Narrative Accuracy PC2 > -1.2).

across participants. Although most participants showed mild or no impairments, one was included within our "severely agrammatic" group and made errors of both verb morphology and verb-argument structure in narratives and structured tests. Nevertheless, lexical-semantic impairment was the most clinically salient feature of their language production.

4.2. The relationship between structured test and connected production measures

Correlational analyses examined the strength of the relationship between structured test performance and grammatical accuracy in narrative samples. With respect to measures of overall sentence production, performance on structured tests (NAT, SPPT) was strongly correlated with the proportion of grammatical sentences in narratives. However, there were no significant cross-task correlations in the domains of verb morphology and argument structure. Interestingly, however, several moderate correlations were found across grammatical domains: performance on the NAT was correlated with morphology accuracy in narratives, whereas performance on the NAT, SPPT, and NAVI correlated with argument structure accuracy in narratives.

These findings suggest, first, that measures of overall sentence production ability provide the greatest degree of convergence between structured tests and narratives. This is probably because overall sentence production measures encompass a wide range of linguistic phenomena, and therefore capture language impairments most robustly. In addition, overall sentence production measures had the greatest degree of inter-subject variability in the present study. For example, the mean percentage of grammatical sentences in PPA patients in the present study was 70% (SD = 18%), whereas for verb morphology accuracy the mean was 95% (SD = 7%) and for argument structure accuracy, 95% (SD = 6%). These scores suggest the possibility of ceiling effects in the latter measures, which may have contributed to the lack of cross-task correlations observed within these domains.

In addition, the existence of correlations across grammatical domains suggests that there may be common processes supporting language production abilities across domains. These findings were also supported by the results of principal components analyses (PCAs). These analyses quantified the amount of shared variance within each of the three sets of measures: structured tests, narrative accuracy, and narrative diversity. The three PCAs each contained a first PC that loaded substantially on all measures, including measures of overall sentence production, verb morphology and argument structure. These PCs thus reflect the overall degree of grammatical impairment. In each case, there was also a second PC reflecting the difference in performance between distinct linguistic processes, namely verb morphology vs. argument structure.

Multiple regression analyses were used to relate the principal components across tasks. These analyses indicated that the severity of grammatical impairments predicted key aspects of the correspondence between structured test and narrative measures. Specifically, participants with severe grammatical impairments showed consistent impairment patterns across tasks; those who had relatively impaired verb argument structure (as compared to verb morphology) in structured tests also did so in narratives, and vice versa. These findings indicate that individuals with more severe aphasia had true differences in language deficit patterns that were robust to task differences.

However, impairment patterns were inconsistent across tasks for patients with milder impairments. We suggest two possible explanations for this finding. First, as mentioned above, the argument structure and verb morphology measures showed signs of ceiling effects and PPA patients with mild grammatical impairments made very few errors on these measures. Thus, these measures may lack sensitivity for detecting more subtle impairments. In future work, online methods such as eye-tracking may be used to test for such subtle impairments. Alternatively, individuals with mild grammatical impairments may truly lack domain-specific impairments, and instead may show a mild, undifferentiated impairment in computing grammatical dependencies. More research is needed to tease out these possibilities.

In addition, these analyses revealed different relationships between structured test and narrative performance. Specifically, in PPA-G, overall grammatical accuracy in narratives (Narrative Accuracy PC1) strongly predicted overall structured test performance (Structured Test PC1), but grammatical diversity in narratives (Narrative Diversity PC1) did not. In PPA-L, the pattern was reversed: grammatical diversity, but not accuracy, in narratives predicted structured test performance. In contrast, grammatical production in narratives did not significantly predict structured test performance in PPA-S.

4.3. Neurocognitive sources of grammatical impairments in PPA

These results suggest that grammatical impairments likely have different sources across PPA subtypes. In PPA-G, grammatical impairments may relate to a core impairment in computing grammatical dependencies, resulting in errors that are observable across tasks and domains. In particular, verb morphology is differentially impaired relative to PPA-L and PPA-S. This pattern, corresponding to Kussmaul's (1877) notion of *agrammatismus,* may index difficulty with morpho-semantic dependencies (time reference/tense) and/or local morpho-syntactic dependencies (e.g., subject-verb agreement). In the language production model of Bock and Levelt (1994), impairments of these sort arise at the positional level. The severity of these deficits varies across participants, ranging from the 11 participants in our "severely agrammatic" group to the 23 with milder deficits. In terms of neural substrates, impaired sentence production in PPA has been associated with damage to the left inferior frontal gyrus and dorsal language tracts, such as the superior longitudinal fasciculus/arcuate fasciculus (DeLeon et al., 2012; Rogalski et al., 2011; Wilson et al., 2011), which tend to show greater atrophy in PPA-G vs. PPA-L (Galantucci et al., 2011). The same regions have been implicated in the production and processing of verb morphology in both neurotypical adults and adults with aphasia (Jakuszeit et al., 2013; Kielar et al., 2011; Miozzo et al., 2010; Regel et al., 2017; Wilson et al., 2014a).

In contrast, compromised grammatical production in individuals with PPA-L may relate to impaired verbal working memory and/or word retrieval impairments, rather than a core deficit in grammatical dependency formation. There is ample evidence for impairments of verbal working memory in PPA-L that affect sentence comprehension and repetition (Amici et al., 2007; Gorno-Tempini et al., 2008; Teichmann et al., 2013; Wilson et al., 2012). In the context of sentence production, Wilson and colleagues (Wilson et al., 2010) found a strong correlation in PPA-L (but not other PPA subtypes) between a composite measure of syntactic production and re-tracings, i.e., verbal reformulations that may reflect working memory failure. We suggest that impaired verbal working memory may limit planning scope and push speakers with PPA-L to favor a "word-by-word" sentence planning approach. In this approach, words are retrieved one-by-one, immediately before they are produced, thus reducing demands on verbal working memory, but increasing demands on online grammatical computations (e.g., Lee et al., 2015). If speakers with PPA-L rely on the word-by-word approach, this may result in a tendency to produce simple and accurate structures in narratives, but to produce errors as the production of complex structures is obligated in structured tests. These hypotheses could be tested in future research using eye-tracking to probe sentence production planning in PPA-L (as has been done in stroke-induced aphasia: Lee and Thompson, 2011a, 2011b; Lee et al., 2015; Mack et al., 2017).

Alternatively, as a core feature of PPA-L is anomia, this deficit could interfere with grammatical constructions apart from working memory failure. In particular, verb-argument structure production (which was comparably impaired in the three major PPA subtypes in the present study) requires retrieval of both a verb and one or more nouns, and therefore anomia would naturally tax argument structure production. In the Bock and Levelt (1994) model, lexical retrieval and argument structure are intricately linked, both occurring on the functional level. Left temporo-parietal regions play a key role in processing verb argument structures (Thompson and Meltzer-Asscher, 2014), and thus damage to these regions may account for the argument structure deficits seen in PPA-L. Similarly, lexical-semantic and conceptual deficits likely contribute to impaired verb-argument structure production in PPA-S, caused by atrophy in anterior temporal regions that contribute to lexical-semantic representations and combinatorial semantic processing (Wilson et al., 2014b).

4.4. Implications for clinical assessment procedures

We now turn to how the present results might bear on clinical assessment of grammatical impairments. Although administering both structured tests and a narrative production protocol provides the greatest degree of information regarding grammatical impairments, this may not be practical in clinical settings, given limitations on face-to-face opportunities between the clinician and patient for data collection, and constraints on the clinician's time for data analysis. This raises the question: can a narrative sample substitute for a structured test battery, in order to reduce patient testing time? The present results, which reveal a generally strong correspondence between grammatical measures across tasks, suggest that this would be a reasonable course of action. Consistent with this, other studies have reported successful assessment of grammatical abilities in PPA using only connected language samples (e.g., Ash et al., 2013).

In addition, the present results motivate some preliminary recommendations for a "decision tree" for coding of narrative samples that makes best use of limited data analysis time. First, the clinician computes the proportion of grammatical sentences in the narrative sample. This measure is relatively quick to compute, and in the present study showed especially high correspondence with structured test results. If this measure indicates relatively severe grammatical impairments, the clinician could perform a more in-depth coding of errors of verb argument structure and morphology. This coding procedure, though more time-consuming, is expected to yield reliable measures of the specific deficit pattern for more severe patients, based on the present results. In contrast, coding grammatical diversity may not be necessary for these patients, based on our finding that grammatical diversity does not predict structured test performance in those with PPA-G. In contrast, for patients with mild (or no) grammatical impairments, the present results suggest that coding grammatical diversity is critical, given its status as a strong predictor of structured test performance in PPA-L in the present study. Specifically, we suggest coding Clause Diversity, as it showed the greatest sensitivity to impairments in PPA, and was the strongest factor in the first principal component in the PCA. Following this protocol may yield pertinent information about grammatical production impairments with relatively little time investment, and this information may guide subsequent decisions about clinical management.

Further, the development of automated coding tools may considerably improve the efficiency of this process. Although we coded grammatical production manually in the present study, Fromm et al. (2020) recently described new additions to CLAN (MacWhinney, 2000) for coding grammatical variables quantified using the Northwestern Narrative Language Analysis (NNLA) system (i.e., the Computerized NNLA (C-NNLA)). These automated coding procedures have been validated and we recommend their use.

4.5. Limitations

We acknowledge that this study had several limitations. First, mean performance on verb morphology and verb-argument structure measures was fairly high in PPA, which may have contributed to ceiling effects. Second, although we used a comprehensive language test battery to assess grammatical production, we used only a single narrative task (Cinderella story re-tell). This limits our ability to generalize our findings, especially given that different connected speech tasks (e.g., picture description, conversation) have elicited distinct patterns of grammatical production in previous research (Sajjadi et al., 2012b). Relatedly, including a single narrative task means that the amount of data per participant is limited. Although previous studies have successfully quantified patterns of grammatical production using a single, brief language sample (Ash et al., 2013), obtaining more data would likely support identification of specific performance patterns (i.e., differential impairment of verb morphology vs. verb-argument structure). Participants with PPA-S produced the shortest language samples, and this

factor may have combined with ceiling effects to produce a weak correspondence between narrative and structured test performance in this group. Finally, we acknowledge that the PPA groups were unequal in size, with the PPA-G group being over twice as large as the PPA-L and PPA-S groups.

4.6. Concluding remarks and future directions

The evidence base that underpins clinical assessment procedures for acquired grammatical impairments continues to grow, and the present study has contributed new information about the relationship between structured test and narrative assessment methods in PPA. However, there are still many aspects of assessment in this domain that require further research. First, it is unknown how grammatical impairments are most effectively assessed in individuals with concomitant lexical-semantic impairments (e.g., in "mixed" PPA or PPA-M). Second, although we know that some individuals without "frank agrammatism" produce grammatically simplified sentences, we have not yet identified the source of this profile. Online methods such as eye-tracking may help to distinguish potential linguistic and cognitive explanations. Finally, we have made preliminary suggestions regarding a "decision tree" for coding connected language samples in order to maximize useful information while minimizing data analysis time. However, future research is needed to test whether this, or other simplified coding procedures, are effective in delineating grammatical impairments in individual participants. More generally, there is a need to identify the assessment methods which result in clinically-relevant information, for example, predicting clinical trajectories over time (e.g., Cupit et al., 2016; Thompson et al., 1997) and identifying patients that will benefit from grammatically-based interventions (Henry et al., 2018).

Author statement

Jennifer E. Mack: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Visualization, Elena Barbieri: Conceptualization, Methodology, Investigation, Writing – review & editing, Sandra Weintraub: Conceptualization, Investigation, Writing – review & editing, M.-Marsel Mesulam: Conceptualization, Investigation, Writing – review & editing, Supervision, Funding acquisition, Cynthia K. Thompson: Conceptualization, Methodology, Investigation, Writing – review & editing, Supervision, Funding acquisition

Acknowledgements

This research was supported by the National Institutes of Health: R01-DC08552 (Mesulam), P50-DC012283 (Thompson), and R01-DC01948 (Thompson). The authors would like to thank their colleagues at the Mesulam Center for Cognitive Neurology and Alzheimer's Disease and the Aphasia and Neurolinguistics Research Laboratory at Northwestern University, in particular Sarah Chandler and Benjamin Rader for assistance with data collection, management, and analysis. We also thank the research participants, their families and caregivers for their contributions to this work.

References

Amici, S., Brambati, S.M., Wilkins, D.P., Ogar, J., Dronkers, N.L., Miller, B.L., Gorno-Tempini, M.L., 2007. Anatomical correlates of sentence comprehension and verbal working memory in neurodegenerative disease. J. Neurosci. 27 (23), 6282–6290.

Ash, S., Evans, E., O'Shea, J., Powers, J., Boller, A., Weinberg, D., Grossman, M., 2013. Differentiating primary progressive aphasias in a brief sample of connected speech. Neurology 81 (4), 329–336.

Auclair-Ouellet, N., 2015. Inflectional morphology in primary progressive aphasia and Alzheimer's disease: a systematic review. J. Neurolinguistics 34, 41–64.

Auclair-Oullet, N., Macoir, J., Laforce, R., Bier, N., Fossard, M., 2016. Regularity and beyond: impaired production and comprehension of inflectional morphology in semantic dementia. Brain Lang. 155–156, 1–11.

Bastiaanse, R., Thompson, C.K., 2012. Perspectives on Agrammatism. Psychology Press, Sussex, UK.

Berndt, R.S., Caramazza, A., 1980. A redefinition of the syndrome of Broca's aphasia: implications for a neuropsychological model of language. Appl. Psycholinguist. 1 (3), 225–278.

Billette, O.V., Sajjadi, S.A., Patterson, K., Nestor, P.J., 2015. SECT and MAST: new tests to assess grammatical abilities in primary progressive aphasia. Aphasiology 29 (10), 1135–1151.

Bock, K., Levelt, W.J.M., 1994. Language production: grammatical encoding. In: Gernsbacher, M.A. (Ed.), Handbook of Psycholinguistics. Academic Press, San Diego, pp. 945–984.

Boschi, V., Catricala, E., Consonni, M., Chesi, C., Moro, A., Cappa, S.F., 2017. Connected speech in neurodegenerative language disorders: a review. Front. Psychol. 8, 269.

Botha, H., Duffy, J.R., Whitwell, J.L., Strand, E.A., Machulda, M.M., Schwarz, C.G., Josephs, K.A., 2015. Classification and clinicoradiologic features of primary progressive aphasia (PPA) and apraxia of speech. Cortex 69, 220–236.

Cho-Reyes, S., Thompson, C.K., 2012. Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences. Aphasiology 26 (10), 1250–1277.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences, Second ed. Lawrence Erlbaum, Hillsdale.

Cupit, J., Graham, N.L., Leonard, C., Tang-Wai, D., Black, S.E., Rochon, E., 2016. Wh-questions and passive sentences in non-fluent variant PPA and semantic variant PPA: longitudinal findings of an anagram production task. Cogn. Neuropsychol. 33 (5–6), 329–342.

DeLeon, J., Gesierich, B., Besbris, M., Ogar, J., Henry, M.L., Miller, B.L., et al., 2012. Elicitation of specific syntactic structures in primary progressive aphasia. Brain Lang 123 (3), 183–190.

Druks, J., 2017. Contemporary and Emergent Theories of Agrammatism. Routledge, New York, NY.

Dunn, L.A., Dunn, L.M., 2006. Peabody Picture Vocabulary Test, Fourth ed. Pearson, San Antonio, Texas.

Fraser, K.C., Meltzer, J.A., Graham, N.L., Leonard, C., Hirst, G., Black, S.E., Rochon, E., 2014. Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex 55, 43–60.

Fromm, D., MacWhinney, B., Thompson, C.K., 2020. Automation of the Northwestern Narrative Language Analysis system. J. Speech Lang. Hear. Res. 63 (6), 1835–1844.

Galantucci, S., Tartaglia, M.C., Wilson, S.M., Henry, M.L., Filippi, M., Agosta, F., Gorno-Tempini, M.L., 2011. White matter damage in primary progressive aphasias: a diffusion tensor tractography study. Brain 134 (Pt 10), 3011–3029.

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.

Goodglass, H., Gleason, J.B., Bernholtz, N.A., Hyde, M.R., 1972. Some linguistic structures in the speech of a Broca's aphasic. Cortex 8 (2), 191–212.

Goodglass, H., Menn, L., 1985. Is agrammatism a unitary phenomenon? In: Kean, M.-L. (Ed.), Agrammatism. Academic Press, New York.

Gorno-Tempini, M.L., Brambati, S.M., Ginex, V., Ogar, J., Dronkers, N.F., Marcone, A., et al., 2008. The logopenic/phonological variant of primary progressive aphasia. Neurology 71 (16), 1227–1234.

Gorno-Tempini, M.L., Dronkers, N.F., Rankin, K.P., Ogar, J.M., Phengrasamy, L., Rosen, H.J., et al., 2004. Cognition and anatomy in three variants of primary progressive aphasia. Ann. Neurol 55 (3), 335–346.

Gorno-Tempini, M.L., Hillis, A.E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S.F., et al., 2011. Classification of primary progressive aphasia and its variants. Neurology 76 (11), 1006–1014.

Harris, J.M., Gall, C., Thompson, J.C., Richardson, A.M., Neary, D., du Plessis, D., et al., 2013. Classification and pathology of primary progressive aphasia. Neurology 81 (21), 1832–1839.

Henry, M.L., Hubbard, H.I., Grasso, S.M., Mandelli, M.L., Wilson, S.M., Sathishkumar, M. T., et al., 2018. Retraining speech production and fluency in non-fluent/agrammatic primary progressive aphasia. Brain 141 (6), 1799–1814.

Howard, D., Patterson, K., 1992. The Pyramids and Palm Trees Test: A Test of Semantic Access from Words and Pictures. Thames Valley Test Company, Bury St. Edmunds, UK.

Hsu, C.H., Thompson, C.K., 2018. Manual vs. automated narrative analysis of agrammatic production patterns: the Northwestern Narrative Language Analysis and Computerized Language Analysis. J. Speech Lang. Hear. Res. 61, 373–385.

Jakuszeit, M., Kotz, S.A., Hasting, A.S., 2013. Generating predictions: lesion evidence on the role of left inferior frontal cortex in rapid syntactic analysis. Cortex 49 (10), 2861–2874.

Kaplan, E., Goodglass, H., Weintraub, S., 1983. The Boston Naming Test. Lea & Febiger, Philadelphia.

Kertesz, A., 2006. Western Aphasia Battery-Revised (WAB-R). Pearson, San Antonio, TX.

Kielar, A., Milman, L., Bonakdarpour, B., Thompson, C.K., 2011. Neural correlates of covert and overt production of tense and agreement morphology: evidence from fMRI. J. Neurolinguistics 24 (2), 183–201.

Knibb, J.A., Woollams, A.M., Hodges, J.R., Patterson, K., 2009. Making sense of progressive non-fluent aphasia: an analysis of conversational speech. Brain 132 (Pt 10), 2734–2746.

Kussmaul, A., 1877. Die störungen der Sprache. Ziemssen's Handbuch der speciellen Pathologie und Therapie 12, 1–300.

Lee, J., Thompson, C.K., 2011a. Real-time production of arguments and adjuncts in normal and agrammatic speakers. Lang. Cognit. Process. 26 (8), 985–1021.

Lee, J., Thompson, C.K., 2011b. Real-time production of unergative and unaccusative sentences in normal and agrammatic speakers: an eyetracking study. Aphasiology 25 (6–7), 813–825.

Lee, J., Thompson, C.K., 2017. Northwestern Assessment of Verb Inflection (NAVI). Northwestern University, Evanston, IL.
Lee, J., Yoshida, M., Thompson, C.K., 2015. Grammatical planning units during real-time sentence production in agrammatic aphasia and healthy speakers. J. Speech Lang. Hear. Res. 58, 1182–1194.
Mack, J.E., Nerantzini, M., Thompson, C.K., 2017. Recovery of sentence production processes following language treatment in aphasia: evidence from eyetracking. Front. Hum. Neurosci. 11 article 101.
MacWhinney, B., 2000. The CHILDES Project: Tools for Analyzing Talk, Third ed. Lawrence Erlbaum Associates, Mahwah, NJ.
MacWhinney, B., Fromm, D., Holland, A., Forbes, M., Wright, H., 2010. Automated analysis of the Cinderella story. Aphasiology 24 (6–8), 856–868.
Malyutina, S., Richardson, J.D., den Ouden, D.B., 2016. Verb argument structure in narrative speech: mining AphasiaBank. Semin. Speech Lang. 37 (1), 34–47.
Marcotte, K., Graham, N.L., Fraser, K.C., Meltzer, J.A., Tang-Wai, D.F., Chow, T.W., et al., 2017. White Matter Disruption and Connected Speech in Non-Fluent and Semantic Variants of Primary Progressive Aphasia. Dement. Geriatr. Cognit. Disord. Extra 7 (1), 52–73.
Menn, L., Obler, L., Miceli, G., 1990. Agrammatic Aphasia: A Cross-Language Narrative Sourcebook. In: , vol. 2. John Benjamins, Amsterdam/Philadelphia.
Mesulam, M.M., Weintraub, S., 2014. Is it time to revisit the classification guidelines for primary progressive aphasia? Neurology 82 (13), 1108–1109.
Mesulam, M.M., Wieneke, C., Thompson, C., Rogalski, E., Weintraub, S., 2012. Quantitative classification of primary progressive aphasia at early and mild impairment stages. Brain 135 (Pt 5), 1537–1553.
Miceli, G., Mazzucchi, A., Menn, L., Goodglass, H., 1983. Contrasting cases of Italian agrammatic aphasia without comprehension disorder. Brain Lang. 19 (1), 65–97.
Milman, L.H., Dickey, M.W., Thompson, C.K., 2008. A psychometric analysis of functional category production in English agrammatic narratives. Brain Lang. 105 (1), 18–31.
Miozzo, M., Fischer-Baum, S., Postman, J., 2010. A selective deficit for inflection production. Neuropsychologia 48 (9), 2427–2436.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9 (1), 97–113.
R Core Team, 2020. R: A Language and Environment for Statistical Computing. Retrieved from. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
Regel, S., Kotz, S.A., Henseler, I., Friederici, A.D., 2017. Left inferior frontal gyrus mediates morphosyntax: ERP evidence from verb processing in left-hemisphere damaged patients. Cortex 86, 156–171.
Rogalski, E., Cobia, D., Harrison, T.M., Wieneke, C., Thompson, C.K., Weintraub, S., Mesulam, M.M., 2011. Anatomy of language impairments in primary progressive aphasia. J. Neurosci. 31 (9), 3344–3350.
Rosenthal, J.A., 1996. Qualitative descriptions or strength of association and effect size. J. Soc. Serv. Res. 21 (4), 37–59.
Saffran, E.M., Berndt, R.S., Schwartz, M.F., 1989. The quantitative analysis of agrammatic production: procedure and data. Brain Lang. 37 (3), 440–479.
Sajjadi, S.A., Patterson, K., Arnold, R.J., Watson, P.C., Nestor, P.J., 2012a. Primary progressive aphasia: a tale of two syndromes and the rest. Neurology 78 (21), 1670–1677.
Sajjadi, S.A., Patterson, K., Tomek, M., Nestor, P.J., 2012b. Abnormalities of connected speech in the non-semantic variants of primary progressive aphasia. Aphasiology 26 (10), 1219–1237.
Shannon, C.D., Weaver, M., 1949. The Mathematical Theory of Communication. University of Illinois Press.
Teichmann, M., Kas, A., Boutet, C., Ferrieux, S., Nogues, M., Samri, D., et al., 2013. Deciphering logopenic primary progressive aphasia: a clinical, imaging and biomarker investigation. Brain 136 (Pt 11), 3474–3488.
Thompson, C.K., 2011. Northwestern Assessment of Verbs and Sentences (NAVS). Evanston, IL.
Thompson, C.K., Ballard, K.J., Tait, M.E., Weintraub, S., Mesulam, M., 1997. Patterns of language decline in non-fluent primary progressive aphasia. Aphasiology 11 (4/5), 297–321.
Thompson, C.K., Cho, S., Hsu, C.J., Wieneke, C., Rademaker, A., Weitner, B.B., et al., 2012a. Dissociations between fluency and agrammatism in primary progressive aphasia. Aphasiology 26 (1), 20–43.
Thompson, C.K., Mack, J.E., 2014. Grammatical impairments in PPA. Aphasiology 28 (8–9), 1018–1037.
Thompson, C.K., Meltzer-Asscher, A., 2014. Neurocognitive mechanisms of verb argument structure processing. In: Bachrach, A., Roy, I., Stockall, L. (Eds.), Structuring the Argument: Multidisciplinary Research on Verb Argument Structure. John Benjamins, Amsterdam, pp. 141–168.
Thompson, C.K., Meltzer-Asscher, A., Cho, S., Lee, J., Wieneke, C., Weintraub, S., Mesulam, M.M., 2013. Syntactic and morphosyntactic processing in stroke-induced and primary progressive aphasia. Behav. Neurol. 26 (1–2), 35–54.
Thompson, C.K., Weintraub, S., Mesulam, M., 2012b. Northwestern Anagram Test (NAT). Evanston, IL.
Weintraub, S., Mesulam, M.M., Wieneke, C., Rademaker, A., Rogalski, E.J., Thompson, C. K., 2009. The Northwestern Anagram Test: measuring sentence production in primary progressive aphasia. Am. J. Alzheimer's Dis. Other Dementias 24 (5), 408–416.
Wicklund, M.R., Duffy, J.R., Strand, E.A., Machulda, M.M., Whitwell, J.L., Josephs, K.A., 2014. Quantitative application of the primary progressive aphasia consensus criteria. Neurology 82 (13), 1119–1126.
Wilson, S.M., Brandt, T.H., Henry, M.L., Babiak, M., Ogar, J.M., Salli, C., et al., 2014a. Inflectional morphology in primary progressive aphasia: an elicited production study. Brain Lang. 136, 58–68.
Wilson, S.M., DeMarco, A.T., Henry, M.L., Gesierich, B., Babiak, M., Mandelli, M.L., et al., 2014b. What role does the anterior temporal lobe play in sentence-level processing? Neural correlates of syntactic processing in semantic variant primary progressive aphasia. J. Cognit. Neurosci 26 (5), 970–985.
Wilson, S.M., Galantucci, S., Tartaglia, M.C., Gorno-Tempini, M.L., 2012. The neural basis of syntactic deficits in primary progressive aphasia. Brain Lang. 122 (3), 190–198.
Wilson, S.M., Galantucci, S., Tartaglia, M.C., Rising, K., Patterson, D.K., Henry, M.L., et al., 2011. Syntactic processing depends on dorsal language tracts. Neuron 72 (2), 397–403.
Wilson, S.M., Henry, M.L., Besbris, M., Ogar, J.M., Dronkers, N.F., Jarrold, W., et al., 2010. Connected speech production in three variants of primary progressive aphasia. Brain 133 (Pt 7), 2069–2088.