HHS Public Access

Author manuscript

J Commun Disord. Author manuscript; available in PMC 2021 May 01.

Published in final edited form as:

J Commun Disord. 2020 ; 85: 105994. doi:10.1016/j.jcomdis.2020.105994.

Using narratives in differential diagnosis of neurodegenerative syndromes

Yasmeen Faroqi-Shah, PhD,
University of Maryland, Department of Hearing and Speech Sciences

Ashlyn Treanor, MA,
University of Maryland, Department of Hearing and Speech Sciences

Nan Bernstein Ratner, PhD,
University of Maryland, Department of Hearing and Speech Sciences

Bronte Ficek, BS,
Johns Hopkins University, Department of Neurology

Kimberly Webster, MA,
Johns Hopkins University, Department of Neurology

Kyrana Tsapkini, PhD
Johns Hopkins University, Department of Neurology

Abstract

Purpose.—Language decline has been associated with healthy aging and with various neurodegenerative conditions, making it challenging to differentiate among these conditions. This study examined the utility of linguistic measures derived from a short narrative language sample for 1) identifying language characteristics and cut-off scores to differentiate between healthy aging, Primary Progressive Aphasia (PPA), Mild Cognitive Impairment (MCI), and Alzheimer's dementia (AD); and 2) differentiating among PPA variants in which language is the primary impairment.

Method.—Participants were 25 neurologically healthy English speakers, 20 individuals with MCI, 20 with AD, and 26 with PPA (non-fluent/agrammatic N=10, logopenic N=9, semantic

[*]**Corresponding author:** Department of Hearing and Speech Sciences, University of Maryland 7251, Prienkert Drive, College Park, MD 20742, 301-405-4229 (phone); 301-314-2417 (fax); yfshah@umd.edu.

Author Contributions:
Conceptualization: Yasmeen Faroqi-Shah
Data Curation: Ashlyn Treanor
Formal Analysis: Ashlyn Treanor
Funding Acquisition: Kyrana Tsapkini
Methodology: Yasmeen Faroqi-Shah, Ashlyn Treanor, Nan Ratner
Project Administration: Kyrana Tsapkini
Resources: Nan Ratner
Supervision: Yasmeen Faroqi-Shah
Writing original draft: Yasmeen Faroqi-Shah, Ashlyn Treanor

N=7). Narrative language samples of the Cookie Theft Picture of persons with healthy aging, MCI and AD were retrospectively obtained from the DementiaBank database (https://talkbank.org/DementiaBank/) and PPA samples were obtained from an ongoing research study. The language samples were analyzed for fluency, word retrieval success, grammatical accuracy, and errors using automated and manual analysis methods. The sensitivity and specificity of various language measures was computed.

Results.—Participants with PPA scored lower than neurologically healthy and MCI groups on fluency (words per minute and disfluencies), word retrieval (Correct Information Units and number of errors), and sentence grammaticality. PPA and AD groups did not differ on language measures. Agrammatic PPA participants scored lower than logopenic and semantic PPA groups on several measures, while logopenic and semantic PPA did not differ on any measures.

Conclusion.—Measures derived from brief language samples and analyzed using mostly automated methods are clinically useful in differentiating PPA from healthy aging and MCI, and agrammatic PPA from other variants. The sensitivity and specificity of these measures is modest and can be improved when coupled with clinical presentation.

Keywords

Alzheimers disease; Cookie theft picture; Fluency; Mild Cognitive Impairment; Narrative language; Primary progressive aphasia

INTRODUCTION

Language abilities, such as word choice and grammatical complexity, are known indicators of brain reserve and neurodegeneration in later life (Riley et al., 2005; Snowdon, Greiner, & Markesbery, 2000). Language decline has not only been associated with healthy aging, but is also documented for various neurodegenerative conditions, particularly Alzheimer's disease (AD) and Primary Progressive Aphasia (PPA) (Mesulam, 2001; Mesulam, Wieneke, Thompson, Rogalski, & Weintraub, 2012). The goals of this study were to identify narrative language measures that would differentiate between persons with healthy aging, Primary Progressive Aphasia (PPA), Mild Cognitive Impairment (MCI), and Alzheimer's dementia (AD), and would differentiate among PPA variants.

Current diagnostic criteria for probable AD include progressively worsening cognitive or behavioral symptoms in two of following five domains: remembering new information, poor reasoning/judgement, visuospatial impairments, impaired language, and changes in personality or behavior (McKhann et al., 2011). Therefore, language deficits can be present in the early stages of the disease, in addition to other cognitive difficulties. Mild Cognitive Impairment (MCI) is a diagnostic label for patients with performance between that of healthy aging individuals and people with dementia, with evidence of cognitive deterioration, and impaired activities of daily living (Winblad et al., 2004). Some individuals with MCI may transition to AD, and others with MCI may end up with another condition, such as PPA (Petersen, 2004). Language impairments can be seen early in the disease course, but this does not always occur due to the heterogeneity of deficits in MCI (Taler & Phillips, 2008). Identifying patterns of language change is crucial for diagnosing PPA, which

is primarily a language-based neurodegenerative condition marked by a gradual decline in language production, object naming, syntax, or word comprehension abilities (Gorno-Tempini et al., 2011; Westbury & Bub, 1997).

The three core criteria for diagnosing PPA are: a language disorder that is not caused by motor or perceptual deficits, that the language impairment is the most salient deficit affecting activities of daily living, and that the underlying disease pathology is progressive in nature (Mesulam, 1982, 2001; Mesulam et al., 2012). Given that the main positive indicator of PPA, language decline affecting activities of daily living, is likely not evident until the disease has progressed significantly, early stage PPA can be difficult to differentiate from language changes of healthy aging (Ash et al., 2013; Fraser et al, 2014; Wilson et al., 2010), and neurodegenerative conditions such as MCI and AD (Choi, 2009; Taler & Phillips, 2011). Further, cognitive abilities remain generally intact in the early years of PPA. Hence commonly used measures that differentiate healthy aging individuals from those who have other neurodegenerative diseases such as Alzheimer's Dementia may not be sensitive to the diagnosis of PPA. For example, according to the criteria on healthy aging established by Tyas and colleagues (2007) it would be expected for a healthy aging individual to obtain a score of 24 or above on the Mini Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975), which is a measure of cognitive impairment that is scored out of 30 possible points. However, even approximately four years past the initial diagnosis of PPA, studies have reported that participants obtained scores on the MMSE that were within normal limits (Ash et al., 2006; Bird, Lambdon-Ralph, Patterson, & Hodges, 2000; Fraser et al., 2014; Mack et al., 2015; Meteyard & Patterson, 2009; Wilson et al., 2010). Thus, for the early differential diagnosis of PPA, cognitive measures such as the MMSE are insufficient, although they may aid clinicians in ruling out cognitive impairments in patients. Additionally, the heterogeneous language profiles of PPA variants increase the challenge of characterizing core language impairments in PPA. The predominant features of agrammatic PPA (PPA-G) are non-fluent speech and impaired syntax, of semantic PPA (PPA-S) are word retrieval and comprehension deficits, and of logopenic PPA (PPA-L) are word retrieval and repetition deficits (Gorno-Tempini et al., 2011). When patients present with characteristics of multiple PPA variants, they are classified as having a "mixed" PPA (Gorno-Tempini et al., 2011). That is, early confirmatory diagnosis of PPA is still an under-developed science.

Although clinicians can use a combination of cognitive and language tests to differentiate among persons with healthy aging, PPA, MCI, and AD (Mesulam et al., 2012), these measures are time-consuming and there is little uniformity in measures selected. There is a need for further research on identifying a brief, reliable test with simple administration procedures to aid in differentiating among language changes associated with healthy aging and neurodegenerative conditions, particularly for early diagnosis of PPA. Ideally, this short test can be scored with ease. In outpatient clinical settings, speech-language pathologists as well as physicians often elicit a short narrative language sample to informally gauge speech-language abilities. A narrative language sample provides an ecologically valid view of how an individual speaks and captures different language domains, such as lexical, morphosyntactic and socio-pragmatic competence. However, fine-grained analysis of narratives is typically not conducted in outpatient clinical settings due to limited time and expertise. Further, there is little consensus on which measures are most sensitive in their

discriminatory power and how to obtain these measures (Ash et al., 2013; Graham, Patterson, & Hodges, 2004; Mack et al., 2015; Meteyard & Patterson, 2009; Thompson et al., 2012; Wilson et al., 2010). These challenges are addressed in the present study by analyzing a short, commonly used narrative sample (the Cookie theft picture, Goodglass & Kaplan, 1983) and using a semi-automated analysis protocol. The purpose of this study was to investigate the utility of a short language sample and semi-automated analyses to 1) identify characteristics of narrative language unique to PPA vs. healthy aging, MCI, and AD; and 2) determine if PPA variants can be differentiated.

Language in healthy aging, AD, MCI and PPA

Prior to describing the study, we present an overview of what is currently known about language differences in healthy aging, AD, MCI, PPA and its variants. This compilation of selected findings enabled us to identify language measures that have the greatest potential of differentiating among clinical conditions. The findings have been organized into the following categories: fluency, word retrieval, and syntax, and are summarized in Supplementary Tables S1 to S4.

Fluency.—Fluency refers to an individual's facility of speech and language performance; individuals who speak fluently produce speech that is unbroken and flowing. Fluent speech requires the synchronization of communication between the speaker and the listener (Clark, 2002; Clark & Fox Tree, 2002). Fluency reflects the cumulative endpoint of speech and language production processes from language formulation to speech production. Common disfluencies include fillers (e.g., interjections such as "um" or "uh"), prolonged syllables, non-reduced vowels, restoration of continuity, and many others (Clark, 2002). Some disfluencies are normal; however, a large quantity of normal-type disfluencies may reflect deficits in the production of speech and language. Studies that have investigated fluency in healthy aging and neurodegeneration have used the following measures: speech rate (i.e., words per minute, WPM), fluency disruptions per 100 words, and number of false starts, unfilled pauses, filled pauses, and repaired sequences (Ash et al., 2006; Ash et al., 2013; Bird et al., 2000; Fraser et al., 2014; Garrard & Forsyth, 2010; Graham et al., 2004; Knibb et al.; Mack et al., 2015; Sajjadi et al., 2012; Thompson et al., 2012; Wilson et al., 2010).

Compared to younger adults, aging healthy adults typically have a slower speech rate, are slower to initiate speech and produce more disfluencies, especially with increased language load and prior to words that are less frequent and have no consistent label (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Horton, Spieler & Shriberg, 2010; Spieler & Griffin, 2006). Individuals with AD often have fluent but empty speech (Tsantali, Economidids, & Tsolaki, 2013). They may produce hesitations; however, these hesitations are often related to impaired semantic systems, with most of their word errors consisting of semantic paraphasias (Croot et al., 2000; Forbes, McKay, & Venneri, 2005; McKhann et al., 2011). Research on fluency in MCI is limited, with one study showing no difference from healthy aging (Roark, Mitchell, Hosom, Hollingshead & Kaye, 2011).

Individuals with PPA, like healthy aging, experience declines in fluency (Ash et al., 2013; Bird et al., 2000; Fraser et al., 2014; Kemper et al., 2001; Knibb, Woollams, Hodges, &

Patterson, 2009; Mack et al., 2015; Spieler & Griffin, 2006; Wilson et al., 2010). PPA variants differ in the extent to which their fluency is impacted, and these studies are summarized in Supplementary Table S2. Out of the eleven studies that compared fluency, only three studies compared the performance all three variants of PPA (Ash et al., 2013; Thompson et al., 2012; Wilson et al., 2010). In these studies, participants with PPA-G consistently had the lowest rates of speech, measured in words per minute. Participants with PPA-L also showed reduced rates of speech, in addition to a high number of false starts, pauses and repaired sequences compared to controls (Mack et al., 2015). The PPA-L group produced the most pauses prior to producing nouns, with effects persisting even after accounting for lexical frequency (Mack et al., 2015). Although a reduced rate of speech is often used to categorize participants with PPA-G, studies often do not report or analyze the types of disfluencies in PPA variants. Further research that could better capture specific types of disfluencies, such as unfilled pauses, filled pauses, word repetitions, phrase repetitions, and phrase revisions, may be beneficial in better understanding if aspects of fluency are impaired across all subtypes of PPA, potentially aiding in the differential diagnosis of PPA from healthy aging and other neurodegenerative conditions.

Word retrieval.—In speech production, word retrieval is a complex process that relies on the integrity and ease of access to semantic and phonological representations, and post-lexical phonemic planning (Butterworth, 1992; Levelt, 1999; Ferreira, 2010). Therefore, difficulties with word retrieval can occur at many levels in the process of selecting and producing a target word. In connected speech, these errors can have cascading effects on an individual's ability to produce fluent and grammatically correct speech. For example, inaccessible verbs and verb arguments will impact an individual's ability to form syntactically correct sentences. Studies examining word retrieval typically use tasks such as confrontation naming, verbal fluency or connected speech, and measures include diversity of words (type-token ratio), occurrence of word classes (open:closed class ratio, noun:verb ratio) and number of words per idea (lexical efficiency, correct information units or idea density). Access to the phonological form of the word is often measured using frequency effects, word length effects and analysis of phonemic paraphasias and neologisms (Bose & Buchanan, 2007; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997).

In aging, although a decline in word retrieval abilities is observed, this is a subtle change, approximately 2% per decade, with most changes occurring after 70 years of age (Connor, Spiro, Obler, & Albert, 2004). It has also been found that healthy aging adults use more words to convey the same number of ideas as younger adults (i.e., lower idea density; Kemper, Greiner, Marquis, Prenovost, & Mitzner, 2001). Individuals with AD may present difficulties with low frequency words, and produce semantic paraphasias and circumlocutions (Croot et al., 2000; Emery, 2000; Sajjadi et al., 2012). Limited research in MCI suggests that lexical-semantic deficits typically occur early in the course of the disease with fewer concepts (Choi, 2009; Taler & Phillips, 2008) and possibly reduced idea density (Roark et al., 2011). Research suggests that naming tests alone may be insufficient to differentiate persons with MCI from AD (Bschor et al., 2001). In PPA, word retrieval deficits are noticed across all variants although there are differences. The proportion of higher frequency words increases with disease progression (Ash et al., 2013; Bird et al.,

2000). Individuals with PPA present with a higher number of word substitutions, circumlocutions, and speech-sound errors than healthy aging adults (Ash et al., 2013; Sajjadi et al., 2012). The findings on word retrieval impairments in the different variants of PPA are summarized in Supplementary Table S3. Four of the eleven studies in Table S3 compared word retrieval abilities across all subtypes of PPA (Ash et al., 2013; Mack et al., 2015; Riello et al., 2018; Thompson et al., 2012; Wilson et al., 2010). These studies found that overall, participants with PPA-S used the least amount of nouns and had impaired open: closed class word ratios. Additionally, the findings from these studies showed that individuals with PPA-L had noun deficits compared to control participants, although not as significant as the noun deficits seen in PPA-S groups, and that individuals with PPA-G had verb deficits.

Morphosyntax.—Deficits in syntax are characterized by verb deficits, the use of fewer grammatical morphemes (e.g., articles, pronouns, auxiliaries, inflections), the presence of grammatical morpheme errors (e.g., verb inflection errors), and an overall reduction in syntactic complexity (Marshall, 2011). Measures that have been used to quantify syntactic impairments have focused on the length of utterances (mean length of utterance: MLU), proportion of grammatically accurate utterances, grammatical complexity, and number of verbs per utterance. While higher MLUs generally reflect more syntactically complex language, these could also reflect verbosity with just more words per utterance. Healthy aging has been associated with a graduate decline in syntactic complexity (Kemper et al., 2001). Simplified syntax is also reported in several studies of PPA across all three variants (e.g., Ash et al., 2006, 2013; Graham et al., 2004; Thompson et al., 2012; Wilson et al., 2010) and occasionally in persons with AD (Emery, 2000). Syntactic complexity in persons with MCI is reported to be similar to healthy controls and persons with AD (Choi, 2009; Drummond et al., 2015; Roark et al., 2011). In terms of outright syntactic errors, these are a characteristic feature of the PPA-G variant. Supplementary Table S4 summarizes findings of syntactic production in persons with PPA. Despite reporting decreased grammaticality of sentences, studies often do not report the types of grammatical errors made by participants with PPA. Further research is needed to not only quantify the grammatical errors of these participants, but also to analyze grammatical error types to identify patterns that could be useful in differentiating participants with PPA from healthy aging individuals and individuals with other neurodegenerative diseases.

Currently, there is little uniformity in which measures are used to quantify and differentiate among language in healthy aging and age-related neurodegenerative conditions, and this makes it difficult to compare findings across studies and populations. Nevertheless, comparing across language domains and participant groups, the following patterns emerge: Healthy aging is characterized by increase in disfluencies, decrease in syntactic complexity and retrieval of less frequent and less codable words (Connor et al., 2004; Kemper et al., 2001; Spieler & Griffin, 2006). The primary domain of language affected in persons with MCI and AD is lexical-semantics, in the absence of deficits in fluency, phonology, and syntax. Individuals with MCI present with a milder impairment of semantics than individuals with AD, but the errors made in spontaneous speech are similar (e.g., semantic paraphasias and circumlocutions) (Choi, 2009; Drummond et al., 2015; Jarrold et al., 2010; Roark et al., 2011; Taler & Philips, 2008). Variables that may be impaired across all three

PPA variants are speech rate, lexical diversity, MLU, and % of grammatical sentences (Ash et al., 2006, 2013; Connor et al., 2004; Fraser et al., 2014; Graham et al., 2004; Kavé et al., 2007; Knibb et al., 2009; Mack et al., 2015; Meteyard & Patterson, 2009; Sajjadi, Patterson, Tomek, & Nestor, 2012; Speiler & Griffin, 2006; Thompson et al., 2012; Wilson et al., 2010). One key characteristic of individuals with PPA is that they not only experience a decline in language abilities like those experienced by healthy aging adults (e.g., declines in fluency, difficulty producing low frequency words, decreased complexity of syntax), but they also show *deviations* in the form of errors in fluency (e.g., false starts, pauses), semantics (e.g., semantic paraphasias), phonology (e.g., phonemic paraphasias), and syntax (e.g., ungrammatical sentences). Past studies of PPA have primarily focused on differentiating between PPA variants, rather than finding commonalities across PPA variants that can be used to differentiate PPA from healthy aging and other related conditions. Further research is needed with this specific goal, especially using a simple narrative task and with semi-automatic analyses procedures.

While there is considerable variability in measures and methods used for narrative analysis, the Cookie Theft picture description task from the Boston Diagnostic Aphasia Examination (BDAE; Goodglass & Kaplan, 1983) is widely used. This task has a standardized administration procedure, provides a clear focus, contains a familiar scene (kitchen) and can be administered quickly, which would make it an efficient and informative measure in a clinical setting. One concern is that describing the Cookie Theft picture yields a rather short narrative. However, language measures elicited from the Cookie Theft picture description task correlate strongly with the longer “Frog, Where Are You?” (Mayer, 1969) narratives in persons with PPA (Ash et al. (2013). Further, a 150-word sample is found to be sufficient in identifying key language measures in persons with semantic dementia (i.e.., PPA-S) and mild Alzheimer’s disease (Sajjadi and colleagues, 2012). The time-consuming process of extraction language measures (e.g., Mean Length of Utterance) from transcribed language samples has been considerably facilitated by a variety of automated language analyses, including automatic part-of-speech (POS) tagging (Toutanova, Klein, Manning, & Singer, 2003), computing syntactic complexity (Lu, 2010), and calculating idea density (CPIDR, Kemper et al., 2001; Kemper, Thompson, & Marquis, 2001; Riley et al., 2005; Snowdon et al., 2000). These and other measures have been incorporated into CLAN (MacWhinney, 2000). Recently the ability of automated and semi-automated computer analyses to differentiate between neurodegenerative conditions is being tested (Fraser et al., 2014; Peintner et al., 2008; Themistocleous et al., under review). To our knowledge, the sensitivity of the Cookie Theft picture description task for differentially diagnosing between persons with healthy aging, MCI, AD, PPA and its variants have not been examined.

The Present Study

The main goal of this study is to identify characteristics of narrative language that are unique to persons with healthy aging, MCI, mild AD and PPA that can be extracted from a semi-automated analysis of a brief language sample. This question will not only isolate which language measures are compromised in each condition, but also the sensitivity and predictive value of different language measures in the differential diagnosis. Based on the

existing literature (summarized in Supplementary Tables S1–S4), we proposed the following hypotheses:

H1. Individuals with PPA, MCI, and AD will present with lexical-semantic deficits compared to control participants, thus differentiating these groups from healthy aging adults.

H2. Participants with MCI and AD will not be significantly impaired in measures of fluency, phonology, syntax, and overall number of errors, thus differentiating participants with MCI and AD from individuals with PPA.

H3. Participants with PPA will show impairments in measures of fluency and syntax when compared to healthy aging adults, MCI and AD. Additionally, individuals with PPA will produce a higher total number of errors, as their language abilities not only decline, but also deviate from the expected performance of healthy aging individuals and individuals with related neurodegenerative diseases.

The second goal of this study is to determine which measures of fluency, word retrieval, and syntax derived from narrative language samples best differentiate among PPA subtypes. While it is expected that all individuals with PPA will have reduced speech rates, MLUs, and % grammatical utterances, additional linguistic measures that have not been used in PPA research will be assessed for their utility in differentiating PPA variants. We hypothesized that:

H4. Individuals with PPA-G will show impairment based on measures of fluency (i.e., lowest speech rate) and syntax (i.e., low performance on all measures of syntax and highest number of morphological errors).

H5. Individuals with PPA-S will show impairment on measures of word retrieval (i.e., lowest performance on all measures and highest number of semantic paraphasias), and H6. Individuals with PPA-L will show impairment on measures of fluency (e.g., highest number of disfluencies) and phonology (i.e., highest number of phonemic paraphasias).

METHODS

This study was a retrospective analysis of data. Language samples of persons with healthy aging, MCI and AD data were obtained from the DementiaBank database (https://talkbank.org/DementiaBank/; Becker, Boiler, Lopez, Saxton, & McGonigle, 1994) and PPA were obtained from an ongoing research study at Johns Hopkins University. All language samples were collected from participants after obtaining informed consent as per the Declaration of Helsinki and the data collection was approved by the Institutional Review Boards at the institutions where it was collected.

Participants

Twenty-four neurologically healthy participants (10 males, 14 females), 20 participants with Mild Cognitive Impairment (MCI; 10 males, 10 females), and 20 participants with Alzheimer's Dementia (AD; 10 males, 10 females) were selected from the DementiaBank database (Becker, et al., 1994). According to the protocol administered by the Alzheimer and

related Dementias Study at the University of Pittsburgh School of Medicine (Becker et al., 1994), control participants had no history of a neurological condition, a cognitively deteriorating condition, or depression. For participants with MCI and AD, inclusion required that all participants be above 44 years of age, have at least 7 years of education, no history of nervous system disorders, have a Mini-Mental State Exam (MMSE, Folstein et al., 1975) score of 10 or greater (out of a maximum of 30), and be able to give informed consent. These participants received neuropsychological and physical assessments, and in 1992 each patient was given a final diagnosis based on their clinical record and additional information available (e.g., autopsy; Becker et al., 1994).

Twenty-six people (14 males, 12 females) with PPA (10 with PPA-G, 9 with PPA-L, 7 with PPA-S) were selected from a Primary Progressive Aphasia study data pool at Johns Hopkins University. The diagnosis and subtyping of PPA was made by neurologists and was based on neurological evaluation, language and cognitive assessments and neuroimaging markers as defined in Gorno-Tempini et al. (2011) criteria. These criteria include a primary language deficit that impacts activities of daily living, and absence of other neurodegenerative, psychiatric, behavioral and medical diagnoses. These criteria do not use MMSE (Folstein et al., 1975) scores, unlike the DementiaBank database (Becker et al., 1994). All participants with PPA were right-handed English-speakers who had received at least a 12^{th} grade education. No participants with a history of stroke or other neurological disorders, language-based learning disorders, or a quotient score of less than 30 on the Western Aphasia Battery-Revised (Kertesz, 2006) were included. The healthy aging and clinical groups did not differ significantly on age, $F(1,3) = .511$, $p = .676$ and years of education, $F(1,3) = 1.48$, $p = .226$ (Table 1).

The data collection was approved by the Human Subjects Ethics Board at University of Pittsburgh School of Medicine (Becker et al., 1994) for Dementia bank and at Johns Hopkins University (Tsapkini et al., 2016) for Primary Progressive Aphasia.

Language Analysis

Several methods of assessing the fluency, lexical-semantic/phonological, and syntactic abilities of individuals based on narrative language have been suggested, both in language development and aphasia research (Templin, 1957; Lee, 1966; Malvern & Richards, 1997; MacWhinney, Fromm, Holland, Forbes, & Wright, 2010; Thorne & Faroqi-Shah, 2016). A detailed discussion of the pros and cons of these measures is beyond the scope of this paper. After careful consideration of prior research, the measures that will be used in the present study along with rationale and our predictions are summarized in Table 2.

This study used a language sample from each participant elicited by the Cookie Theft picture description from the Boston Diagnostic Aphasia Examination (BDAE; Goodglass & Kaplan, 1983). This was the only language sample available for all four populations. Narrative samples from participants with PPA were transcribed using CHAT conventions (MacWhinney, 2000). Narrative samples from the DementiaBank database (Becker et al., 1994) were previously transcribed in CHAT (MacWhinney, 2000) by the researchers who collected the recordings. Each transcript was double-checked for accuracy using the original audio files. Task clarification questions were not coded and excluded from further analysis

(e.g., “Can I look at it and tell you?”). Various programs from CLAN (MacWhinney, 2000) were used to extract various language measures, as outlined in Table 2. Codes from the CLAN manual (http://childes.psy.cmu.edu/manuals/Clin-CLAN.pdf) and Coding Cheat Sheet on AphasiaBank (http://aphasia.talkbank.org/) were added to capture errors such as disfluencies (e.g., fillers, whole word repetitions, part word repetitions), word retrieval errors (e.g., semantic paraphasias, phonemic paraphasias, neologisms, etc.), syntactic errors (e.g., the omission of function words, verb inflection errors), and morphosyntactic errors (e.g., verb inflection errors). Syntactic errors and morphosyntactic errors were both coded separately since sentences containing grammatical errors were used to calculate the proportion of grammatical utterances, a more global measure of syntax, and morphosyntactic errors were used to investigate if verb inflection is specifically impaired in PPA. Additional codes were created to capture other potential deficits in persons with PPA. The [* sp:m] code was created to capture mixed paraphasias (i.e., containing both semantic and phonological errors), the [+ sce] code was created to mark sentences that were semantically incorrect (e.g., “The boy is in the cookie jar.”), and the [+ comp] code was created to mark sentences with more complex construction (i.e., sentences containing embedded clauses, qualifiers, and passive construction). A list of all codes used in the study can be found in Appendix A.

Two measures, speech rate and CIUs, were calculated manually outside of CLAN. In order to calculate rate of speech, the duration of the audio sample for 5 consecutive utterances was measured using the Praat program (Boersma, 2001). The total number of intelligible, whole words (i.e., no phonological fragments or fillers) was counted manually. Correct information units were calculated by manually counting the total number of words in a sample, and then counting the number of CIUs based on criteria outlined by Nicholas & Brookshire (1993). The proportion of CIUs was calculated by dividing the number of CIUs by the number of words in the sample.

Inter-rater reliability—A subset of the narratives produced by each participant group (10–20% of the transcripts per group) was randomly chosen to be independently transcribed and coded by an additional researcher. To compare the reliability of transcriptions, intra-class correlations (ICC) were calculated. ICCs were used rather than Kappa statistics because, unlike Cohen’s kappa which quantifies reliability based on complete agreement, ICCs use the magnitude of disagreement to calculate reliability estimates (Hallgren, 2012). A high degree of reliability was found between researchers. The average ICC measure for the total number of utterances was .94 (95% confidence interval (CI) .84 to .97, $F(2, 18) = 15.81$, $p = .001$). The average ICC measure for coding reliability (i.e., total number of codes used per sample) was .98 (95% CI .95 to .99, $F(2, 8) = 77.5$, $p = .001$). The random sample of narratives was also checked for CIU reliability. The average ICC measure for CIUs was .99 (95% CI .995 to .999, $F(2,18) = 473.4$, $p = .0001$). Any disagreements were resolved by re-listening to samples and reaching a consensus between transcribers.

Statistical Analysis

Statistical analyses were performed with SPSS version 24. For each measure, the four groups (i.e., Healthy aging, PPA, MCI, and AD) were analyzed using one-way ANOVAs,

with a Bonferroni-corrected *p* value for significance set at $p < .003$ (.05/14 language measures); and significant comparisons were followed with Tamhane's T2 post-hoc tests with a threshold criterion set at $p < .050$. The Levene's test indicated that the subgroups of PPA (i.e., PPA-G, PPA-L, PPA-S) did not meet the requirement of homogeneity of variance for parametric statistical tests, so the Kruskal-Wallis non-parametric test was used to for the comparison of linguistic measures across PPA subgroups with a Bonferroni-corrected *p* value for significance set at $p < .003$ (.05/14 language measures). Significant comparisons were followed up with pairwise comparisons using the Mann-Whitney U test with a significance threshold set at $p < .050$.

For measures that were statistically different between the healthy group and any of the neurodegenerative groups, cut-off scores for the diagnosis were determined based on the 95% confidence interval (CI) of the mean difference between the two groups. For measures in which a neurodegenerative group had higher values than the control group, the upper bound of the 95% CI was added to the healthy group's mean score. For measures in which the neurodegenerative group had lower values than the healthy group, the lower bound of the 95% CI was subtracted from the control group's mean score. For example, the pairwise comparison between persons with PPA and healthy aging for words per minute (WPM) yielded a mean difference of 64.8, SE=10.9, 95% CI of difference = 43.1 to 86.5. Given that persons with PPA have lower WPM than healthy adults, the lower bound of the 95% CI of the difference (43.1) was subtracted from the healthy group's mean WPM (141.7, see Table 3), resulting in a cut-off WPM value of 98.6 (see Table 4) for differentiating between persons with healthy aging and PPA based on WPM. Sensitivity values (i.e., how well the cut-off score detects a true positive) were then calculated by counting the number of participants with a clinical diagnosis (e.g., PPA) whose scores fell within the abnormal range (true positives) and within the normal range (false negatives) using the formula: true positives/(true positives + false negatives) × 100. For example, for WPM, 17 and 7 PPA participants had WPM values below and above the 98.6 cut-off respectively. This gave a sensitivity value of 17/(17+9) × 100 = 65.3%. Similarly, specificity (i.e., how well the cut-off score excludes a true negative) was calculated as true negatives/(true negatives + false positives) × 100. For WPM, 19 healthy participants' WPM was higher than the 98.6 cut-off value, yielding a specificity value of 19/(19+5) × 100 = 79.2%. Additionally, Z-scores were calculated for each group, based on the control group's performance on each language measure.

RESULTS

1. (a) Differentiating among Healthy Aging, MCI, PPA, and AD

The mean values of the linguistic measures and the statistical results of the ANOVA are summarized in Table 3. The findings of the post-hoc tests are given in the following paragraphs.

Fluency measures.—Speech rate showed a main effect of group. Pairwise post-hoc comparisons showed that it was reduced in participants with PPA compared to healthy aging, $p = .001$, and MCI, $p = .001$. The total number of disfluencies produced was also statistically

significant. Participants with PPA produced a greater number of disfluencies than healthy aging, p =.001, and MCI, p = .002. A post-hoc analysis of disfluency types revealed that participants with PPA produced more prolongations, p = .001, word repetitions, p = .008, and filled pauses, p = .004 than healthy aging participants. Additionally, participants with PPA produced more prolongations, p= .001 than participants with MCI, and more filled pauses that participants with AD, p = .001.

Word retrieval measures.—The following three measures showed a main effect of group: proportion of CIUs per sample, number of phonological errors, and total number of word retrieval errors. Post-hoc tests showed that compared to healthy aging, participants with PPA had lower CIU scores, p = .001, produced more phonological errors, p = .010, and more word retrieval errors, p = .001. Participants with PPA also had lower CIU scores, p = .001, more phonological errors, p = .009, and more total errors, p = .001 compared to participants with MCI.

Syntactic measures.—The proportion of grammatical utterances was the only syntactic measure to show a main effect of group; participants with PPA produced fewer grammatically well-formed sentences than healthy aging, p = .001, and MCI, p = .001.

General measures.—The total number of errors in a sample showed a main effect of group with participants with PPA produced more errors per narrative (i.e., disfluencies, word retrieval errors, and grammatical errors) than healthy aging, p < .001, and MCI, p = .001.

To summarize, the following linguistic measures were found to be statistically significant in differentiating participants with PPA from healthy controls and participants with MCI: speech rate, total number of disfluencies, proportion of CIU's, number of phonological errors, total number of word retrieval errors, proportion of grammatical utterances, and total number of errors in the sample. No measures were found to be statistically significant in differentiating participants with PPA from AD.

1. (b) Sensitivity, and specificity of language measures

The sensitivity and specificity values for differentiating between persons with healthy aging versus PPA, and persons with MCI versus PPA are given in Table 4. PPA versus AD values are not listed because these two groups did not differ statistically (see Table 3). The measures with the highest sensitivity scores were words per minute and the proportion of CIUs. Specificity values were higher than sensitivity values for all measures (over 75% accurate in detecting a true negative), suggesting that these cut-off scores are better at excluding true negatives (i.e., identifying persons with healthy aging or MCI) than they are at detecting true positives (i.e., identifying a participant with PPA). The highest specificity values were found for total disfluencies, total retrieval errors, and total errors. Z-scores for each measure also showed that the performance of participants with PPA is farthest from that of control participants on the measures of total disfluencies, total retrieval errors, and total errors.

2. Differentiating among PPA variants

The results are summarized in Table 5. Three language measures were statistically significant: WPM, MLU based on morphemes, and verbs/utterance. Note that idea density approached significance (p =.004), but did not reach the stringent significance threshold of p=.003. Follow-up pairwise comparisons revealed that compared to both PPA-L and PPA-S groups, participants with PPA-G had decreased WPM (PPA-G versus PPA-L, p = .006; PPA-G versus PPA-S, p = .001), lower MLU (PPA-G versus PPA-L, p = .001; PPA-G versus PPA-S p = .007), and fewer verbs/utterance (PPA-G versus PPA-L, p = .001; PPA-G vs. PPA-S, p = .001). No measures were significantly different between the PPA-L and PPA-S groups.

DISCUSSION

The primary goal of this study was to determine if any specific measures derived from narrative language samples could differentiate between persons with healthy aging, MCI, AD and PPA. Participants with PPA produced lower fluency (speech rate and disfluencies), lower word retrieval (proportion of Correct Information Units, number of errors), and a lower proportion of grammatical utterances than participants with healthy aging and MCI. There were no differences between persons with PPA and AD in language measures. The second goal of this study was to determine which language measures best differentiate PPA subtypes from each other. Participants with PPA-G scored lower than persons with PPA-L and PPA-S in measures of fluency (speech rate) and syntax (MLU and verbs/utterance), while the PPA-L and PPA-S groups did not differ from each other. These findings will be discussed in the following sections.

Differentiating among Healthy Aging, MCI, PPA, and AD

We had hypothesized that all three neurodegenerative groups would score lower than persons with healthy aging in lexical-semantic measures (H1), and that the PPA group would score lower than persons with healthy aging, MCI and AD on measures of fluency and syntax (H2, H3). The first hypothesis was supported to the extent that the PPA groups scored lower than the healthy aging group on CIUs and the total number of word retrieval errors. The findings of the present study support previous research showing that participants with some subtypes of PPA have word retrieval impairments relative to healthy controls (Ash et al., 2006; Kavé et al., 2007; Mack et al., 2015; Meteyard & Patterson, 2009; Sajjadi et al., 2012; Thompson et al., 2012; Wilson et al., 2010). The Cookie theft picture descriptions of the persons with MCI and AD did not differ from persons with healthy aging for word retrieval measures, which is consistent with some research (Croot et al., 2000, Emery, 2000; Roark et al., 2011) but not with other studies (Supplementary Table S1; Choi, 2009; Forbes et al., 2005; Gitit & Dassa, 2018; Jarrold et al., 2010; McKhann et al., 2011; Riley et al., 2005; Sajjadi et al, 2012; Taler & Phillips, 2008). It is likely that the different findings across studies may be due to differences in the language task and language measures used. For instance, Gitit and Dassa (2018) found that persons with AD and healthy aging differed in Cookie Theft picture descriptions in measures that were not used in the current study, including word frequency, and proportions of words, content words, pronouns, and prepositions.

The results supported our second hypothesis in that the PPA group had lower fluency and grammatical accuracy compared to healthy aging and MCI groups. Although few previous studies have compared speech rates and grammatical accuracy in persons with PPA as an entire group to healthy aging adults, the present study supports research showing reduced speech rate (Ash et al., 2006, 2013; Fraser et al., 2014; Graham et al., 2004; Mack et al., 2015; Sajjadi et al., 2012; Thompson et al., 2012; Wilson et al., 2010) and grammatical accuracy (Ash et al., 2013; Mack et al., 2015; Thompson et al., 2012; Wilson et al., 2010) in specific subtypes of PPA relative to healthy aging adults.

Interestingly, although individuals with AD have been reported to produce fluent speech (Tsantali et al., 2013), this study did not find a significant difference in fluency between persons with PPA and AD. It is likely that fluency is compromised in persons with AD primarily due to word retrieval difficulties as evidenced by the empty content of their speech (Tsantali et al., 2013), while in PPA syntactic, motoric, and word retrieval problems might impact fluency. The failure to find language differences between AD and PPA is not entirely surprising given that recent studies have found overlap in left temporal regions of neural atrophy in persons with AD and PPA (Leyton et al., 2017, 2019). Our results suggest that a combination of the Cookie Theft picture description task and the language measures used in the present study does not aid in differentiating persons with PPA from AD. This is not of significant clinical concern because the diagnosis of AD is seldom made based on narrative language measures alone (McKhann et al., 2011).

Overall, the findings of this study support the use of narrative language sampling in differentiating persons with PPA from healthy aging and MCI. We identified six language markers that could differentiate persons with PPA from healthy aging and MCI using the Cookie Theft picture language sample (Table 4): speech rate lower than about 100 WPM, more than 20 disfluencies, fewer than 70% CIUs, more than .6 word retrieval errors, fewer than 78% grammatical utterances and more than 21 errors total (including semantic and phonemic paraphasias, disfluencies, grammatical errors etc.). While all six measures had relatively high specificity, the sensitivity was modest, ranging from 55% to 65%. WPM and proportion CIU had the highest sensitivity (at 65%) in differentiating persons with PPA from healthy aging and MCI. Although the sensitivity can be raised by using different cut-off scores, this would lower the specificity. This trade-off between sensitivity and specificity can be addressed by considering the pitfalls of overidentification (with low specificity) versus under-diagnosis (with low sensitivity). Given that neurodegenerative diagnoses made on the basis of multiple criteria that include patient history, cognitive tests and neuroimaging findings (Gorno-Tempini et al., 2011), and narrative sample would not be used as the sole basis for a diagnosis, the modest sensitivity of narrative language measures is not necessarily problematic. Rather, this study suggests that narrative language samples help inform a diagnosis, and it is expected that information derived from narrative samples would be used in conjunction with a combination of language and cognitive tests in order to confirm an individual’s diagnosis.

Differentiating among PPA variants

Based on prior research (Supplementary Tables S2–S4), we had hypothesized that among the PPA subtypes, persons with (1) PPA-G would score lowest on fluency (WPM) and syntax, (2) PPA-S would score lowest on word retrieval measures, and (3) PPA-L would show the most disfluencies and phonemic paraphasias. Only the first prediction was confirmed, replicating previous findings that individuals with PPA-G are nonfluent (Ash et al., 2006, 2013; Fraser et al., 2014; Graham et al., 2004; Mack et al., 2015; Thompson et al., 2012; Wilson et al., 2012) and have syntactic impairments among all of the PPA variants (Ash et al., 2006; Ash et al., 2013; Fraser et al., 2014; Knibb et al., 2009; Mack et al., 2015; Thompson et al., 2012; Wilson et al., 2012). It is important to note the specific grammatical measures that were significantly different: while the PPA group as whole produced *fewer grammatically accurate utterances* than healthy and MCI groups, persons with PPA-G produced *shorter utterances* and *fewer verbs per utterance* than other PPA variants. This provides guidelines for selecting appropriate language measures depending on whether the clinical purpose is to identify persons with PPA or to differentiate among PPA variants.

In the present study, none of the word retrieval measures differentiated these groups from each other or from PPA-G. It is not entirely surprising that persons with PPA-S and PPA-L patterned together, as both variants are characterized by word retrieval deficits (Ash et al., 2013; Mack et al., 2015; Thompson et al., 2012; Wilson et al., 2010). And some authors have argued that PPA-L is an earlier manifestation of PPA-S in that, as the neurogenerative changes of PPA progress, word form information degenerates (i.e., PPA-L) before semantic representations of words are lost (i.e., PPA-S) (Mesulam et al., 2014). Table 5 shows that persons with PPA-S trended towards a higher number of semantic paraphasias, as one would have expected. This trend might not have reached significance due to the short length of the language sample. Idea density, which is often taken as a word retrieval measure, was not significantly different for PPA-S compared to the other PPA groups. As previous studies have suggested, it is possible that idea density does not purely measure semantic ability since it counts “propositions” as non-noun entities (Brown et al., 2008; Thorne & Faroqi-Shah, 2016). Participants with PPA-S have been shown to use fewer nouns than participants with other subtypes of PPA, and it has also been suggested that participants with PPA-L have noun deficits (Ash et al., 2013; Mack et al., 2015; Thompson et al., 2012; Wilson et al., 2010).

To summarize, the following linguistic measures were useful in differentiating persons with PPA-G from the other variants of PPA: WPM, MLU, and verbs per utterance. The word retrieval measures used in the present study were either not sensitive enough to detect the word retrieval of persons with PPA-S and PPA-L, or, alternatively, the sample length did not allow these differences to emerge.

Conclusions

Analysis of a short language sample such as the Cookie Theft picture can supplement language and cognitive test batteries and add to the reliability of assessment in differentially diagnosing between persons with PPA versus healthy aging and MCI. The language measures that are particularly useful for this are WPM and proportion CIUs, followed by

counts of disfluencies, word retrieval errors, grammatical utterances and total errors. WPM, MLU and verbs per utterance differentiate persons with PPA-G from other PPA variants, while this short language sample did not differentiate between persons with PPA-S and PPA-L.

Limitations

A limitation of this study is that the Cookie Theft picture description was the only narrative available for all participant groups. This picture, which consists of a simple drawing of a familiar scene, elicited generally short and simple narratives with limited ideas and lexical diversity and frequent use of present progressive tense e.g., “The boy is standing. The stool is falling.”). Between-group differences in more sophisticated measures such as MATTR and sentence complexity may have emerged with a more complex narrative task. Further research is needed to compare the performance of these groups using complex picture description tasks, or story re-telling narrative tasks. Longer samples could potentially reveal larger between-group differences (especially between persons with PPA and AD, and among PPA variants) and improve diagnostic sensitivity. Longer narratives could also allow us to use other language measures such as a measure of lexical diversity called the D measure (Malvern & Richards, 1997) and a measure of syntactic ability, Developmental Sentence Scoring (DSS, Lee, 1971) which was used by Thorne and Faroqi-Shah (2016) for persons with aphasia.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to thank our participants and referring physicians for their dedication and interest in our study. This work was supported by grants from the Science of Learning Institute at Johns Hopkins University and by the National Institutes of Health (National Institute of Deafness and Communication Disorders) through award R01 DC014475 to KT.

REFERENCES

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, 329–336. [PubMed: 23794681]

Ash S, Moore P, Antani S, McCawley G, Work M, & Grossman M (2006). Trying to tell a tale: Discourse impairments in progressive aphasia and frontotemporal dementia. Neurology, 66(9): 1405–1413. [PubMed: 16682675]

Becker JT, Boiler F, Lopez OL, Saxton J, & McGonigle KL (1994). The natural history of Alzheimer’s disease: description of study cohort and accuracy of diagnosis. Archives of Neurology, 51(6), 585–594. [PubMed: 8198470]

Bird H, Lambdon-Ralph MA, Patterson K, & Hodges JR (2000). The rise and fall of frequency and imageability: Noun and verb production in semantic dementia. Brain and Language, 73, 17–49. [PubMed: 10872636]

Boersma Paul (2001). Praat, a system for doing phonetics by computer. Glot International. 5:9/10, 341–345.

Bortfeld H, Leon SD, Bloom JE, Schober MF, & Brennan SE (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. Language and speech, 44(2), 123–147. [PubMed: 11575901]

Bose A, & Buchanan L (2007). A cognitive and psycholinguistic investigation of neologisms. Aphasiology, 21(6–8), 726–738.

Brown C, Snodgrass T, Kemper SJ, Herman R, & Covington MA (2008). Automatic measurement of propositional idea density from part-of-speech tagging. Behavior Research Methods, 40(2), 540–545. [PubMed: 18522065]

Brown R (1973). A first language: The early stages. Cambridge, MA: Harvard University Press.

Bschor T, Kuhl KP, & Reischies FM (2001). Spontaneous speech of patients with dementia of the Alzheimer type and mild cognitive impairment. International Psychogeriatrics, 13(3), 289–298. [PubMed: 11768376]

Butterworth B (1992). Disorders of phonological encoding. Cognition, 42(1–3), 261–286. [PubMed: 1582159]

Choi H (2009). Performances in a picture description task in Japanese patients with Alzheimer's disease and with mild cognitive impairment. Korean J Commun Disord, 14, 326–337.

Clark HH (2002). Speaking in time. Speech communication, 36(1), 5–13.

Clark HH, & Fox Tree JE (2002). Using uh and um in spontaneous speaking. Cognition, 84(1), 73–111. [PubMed: 12062148]

Connor LT, Spiro A, Obler LK, & Albert ML (2004). Change in object naming ability during adulthood. The Journals of Gerontology Series B: Pscyhological Sciences and Social Sciences, 59(5), P203–P209.

Covington MA, & McFall JD (2010). Cutting the Gordian knot: The moving-average type–token ratio (MATTR). Journal of Quantitative Linguistics, 17(2), 94–100.

Croot K, Hodges JR, Xuereb J, & Patterson K (2000). Phonological and articulatory impairment in Alzheimer's disease: A case series. Brain and Language, 75, 277–309. [PubMed: 11049669]

Dell GS, Schwartz MF, Martin N, Saffran EM, & Gagnon DA (1997). Lexical access in aphasic and nonaphasic speakers. Psychological review, 104(4), 801. [PubMed: 9337631]

Dennis PA, & Hess TM (2016). Aging-related gains and losses associated with word production in connected speech. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn, 23(6), 638–650. doi: 10.1080/13825585.2016.1158233 [PubMed: 26963869]

Drummond C, Coutinho G, Fonseca RP, Assunção N, Teldeschi A, de Oliveira-Souza R, … & Mattos P (2015). Deficits in narrative discourse elicited by visual stimuli are already present in patients with mild cognitive impairment. Frontiers in aging neuroscience, 7, 96. [PubMed: 26074814]

Emery VOB (2000). Language impairment in dementia of the Alzheimer type: a hierarchical decline?. The International Journal of Psychiatry in Medicine, 30(2), 145–164. [PubMed: 11001278]

Ferreira VS (2010). Language production. Wiley Interdisciplinary Reviews: Cognitive Science, 1(6), 834–844. [PubMed: 26271781]

Folstein MF, Folstein SE, & McHugh PR (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr Res, 12(3): 188–199.

Forbes-McKay KE & Venneri A (2005). Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. Neurological Science, 26, 243–254.

Fraser KC, Meltzer JA, Graham NL, Leonard C, Hirst G, Black SE, & Rochon E (2014). Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex, 55, 43–60. [PubMed: 23332818]

Garrard P & Forsyth R (2010). Abnormal discourse in semantic dementia: A data-driven approach. Neurocase, 16(6), 520–528. [PubMed: 20614366]

Kavé G, & Dassa A (2018). Severity of Alzheimer's disease and language features in picture descriptions. Aphasiology, 32(1), 27–40. doi: 10.1080/02687038.2017.1303441

Goodglass H & Kaplan E (1983). Boston diagnostic aphasia examination. Media, PA: Williams & Wilkins.

Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF…Grossman M (2011). Recommendations for the classification of primary progressive aphasia and its variants. Neurology, 76(11), 1006–1014. [PubMed: 21325651]

Graham NL, Patterson K, Hodges JR (2004). When more yields less: speaking and writing deficits in nonfluent progressive aphasia. Neurocase, 10, 141–55. [PubMed: 15788253]

Hallgren KA (2012). Computing inter-rater reliability for observational data: an overview and tutorial. Tutorials in quantitative methods for psychology, 8(1), 23. [PubMed: 22833776]

Horton WS, Spieler DH, & Shriberg E (2010). A corpus analysis of patterns of age-related change in conversational speech. Psychology and Aging, 25(3), 708–713. doi: 10.1037/a0019424 [PubMed: 20677883]

Jarrold WL, Peintner B, Yeh E, Krasnow R, Javitz HS, & Swan GE (2010, 8). Language analytics for assessing brain health: Cognitive impairment, depression and pre-symptomatic Alzheimer's disease In International Conference on Brain Informatics (pp. 299–307). Springer Berlin Heidelberg.

Kavé G, Leonard C, Cupit J, & Rochon E (2007). Structurally well-formed narrative production in the face of severe conceptual deterioration: A longitudinal case study of a woman with semantic dementia. Journal of Neurolinguistics, 20(2), 161–177.

Kavé G, & Nussbaum S (2012). Characteristics of noun retrieval in picture descriptions across the adult lifespan. Aphasiology, 26(10), 1238–1249. doi: 10.1080/02687038.2012.681767

Kemper S, Greiner LH, Marquis JG, Prenovost K, & Mitzner TL (2001). Language decline across the life span: findings from the Nun Study. Psychology and aging, 16(2), 227–239. [PubMed: 11405311]

Kemper S, Thompson M and Marquis J (2001). Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content. Psychology and Aging 16, 600–614. [PubMed: 11766915]

Knibb JA, Woollams AM, Hodges JR, & Patterson K (2009). Making sense of progressive non-fluent aphasia: an analysis of conversational speech. Brain, 132, 2734–46. [PubMed: 19696033]

Lee LL (1966). Developmental sentence types: A method for comparing normal and deviant syntactic development. Journal of speech and hearing disorders, 31(4), 311–330. [PubMed: 5922999]

Lee LL, & Canter SM (1971). Developmental sentence scoring: A clinical procedure for estimating syntactic development in children's spontaneous speech. Journal of Speech and Hearing Disorders, 36(3), 315–340. [PubMed: 5566112]

Levelt WJ (1999). Models of word production. Trends in cognitive sciences, 3(6), 223–232. [PubMed: 10354575]

Lu X (2010). Automatic analysis of syntactic complexity in second language writing. International Journal of Corpus Linguistics, 15(4), 474–496.

Leyton CE, Hodges JR, Piguet O, & Ballard KJ (2017). Common and divergent neural correlates of anomia in amnestic and logopenic presentations of Alzheimer's disease. Cortex, 86, 45–54. doi: 10.1016/j.cortex.2016.10.019 [PubMed: 27875715]

Leyton CE, Landin-Romero R, Liang CT, Burrell JR, Kumfor F, Hodges JR, & Piguet O (2019). Correlates of anomia in non-semantic variants of primary progressive aphasia converge over time. Cortex, 120, 201–211. doi: 10.1016/j.cortex.2019.06.008 [PubMed: 31325799]

Mack JE, Chandler SD, Meltzer-Asscher A, Rogalski E, Weintraub S, Mesaulam M, & Thompson CK (2015). What do pauses in narrative production reveal about the nature of word retrieval deficits in PPA? Neuropsychologia, 77, 211–222. [PubMed: 26300385]

MacWhinney B (2000). The CHILDES project: Tools for analyzing talk, Vol 1: Transcription format and programs. 3 Mahwah, NJ: Erlbaum.

MacWhinney B, Fromm D, Holland A, Forbes M, & Wright H (2010). Automated analyses of the Cinderella Story. Aphasiology, 24, 856–868. [PubMed: 25067870]

Malvern DD, & Richards BJ (1997). A new measure of lexical diversity In Ryan A & Wray A (Eds.), Evolving models of language (58–71). Clevedon, UK: Multilingual Matters.

Marshall J (2011). Disorders of Sentence Processing in Aphasia In Papathanasiou I, Coppens P, & Potages (Eds.) Aphasia and Related Neurogenic Communication Disorders (pp.197–213). Burlington, MA: Jones & Bartlett Learning; 1st edition.

Mayer M (1969). Frog, where are you? New York: Penguin Books.

McHugh ML (2012). Interrater reliability: the kappa statistic. Biochemia medica, 22(3), 276–282. [PubMed: 23092060]

McKhann G, Knopman D, Chertkow H, Hyman B, Jack C Jr., Kawas C.…Phelps C (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging – Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 1;7(3), 263–269.

Mesulam MM (1982). Slowly progressive aphasia without generalized dementia. Annals of Neurology, 11(6), 592–8. [PubMed: 7114808]

Mesulam MM (2001). Primary progressive aphasia. Annals of Neurology, 49(4), 425–32. [PubMed: 11310619]

Mesulam M, Wieneke C, Rogalski E, Cobia D, Thompson C, & Weintraub S (2009). Quantitative template for subtyping primary progressive aphasia. Archives of neurology, 66(12), 1545–1551. [PubMed: 20008661]

Mesulam MM, Wieneke C, Thompson C, Rogalski E, Weintraub S (2012). Quantitative classification of primary progressive aphasia at early and mild impairment stages. Brain, 135, 1537–1553. [PubMed: 22525158]

Meteyard L & Patterson K (2009). The relation between content and structure in language production: An analysis of speech errors in semantic dementia. Brain and Language, 110(3), 121–134. [PubMed: 19477502]

Mortensen L, Meyer AS, & Humphreys GW (2006). Age-related effects on speech production: A review. Language and Cognitive Processes, 21(1–3), 238–290. doi: 10.1080/01690960444000278

Nicholas LE, & Brookshire RH (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. Journal of Speech, Language, and Hearing Research, 36(2), 338–350.

Patterson K & McDonald MC (2006). Sweet nothings: narrative speech in semantic dementia In: Andrews S (Ed.), Inkmarks to Ideas: Current Issues in Lexical Processing (pp. 299–317). Hove, UK: Psychology Press.

Peintner B, Jarrold W, Vergyri D, Richey C, Gorno-Tempini ML, & Ogar J (2008). Learning diagnostic models using speech and language measures. In: Engineering in Medicine and Biology Society, 2008. 30th Annual International Conference of the IEEE: 4648–4651.

Petersen RC (2004). Mild cognitive impairment as a diagnostic entity. Journal of Internal Medicine, 256, 183–194. [PubMed: 15324362]

Plante E, & Vance R, (1994). Selection of preschool language tests: A data-based approach. Language. Speech, and Hearing Services in Schools 25, 15–24.

Riello M, Faria AV, Ficek B, Webster K, Onyike CU, Desmond J, … Tsapkini K (2018). The Role of Language Severity and Education in Explaining Performance on Object and Action Naming in Primary Progressive Aphasia. Frontiers in Aging Neuroscience, 10(346). doi: 10.3389/fnagi.2018.00346

Riley KP, Snowdon DA, Desrosiers MF, & Markesbery WR (2005). Early life linguistic ability, late life cognitive function, and neuropathology: findings from the Nun Study. Neurobiology of Aging, 26, 341–347. [PubMed: 15639312]

Roark B, Mitchell M, Hosom JP, Hollingshead K, & Kaye J (2011). Spoken language derived measures for detecting mild cognitive impairment. IEEE Transactions on Audio, Speech, and Language Processing, 19(7), 2081–2090.

Sajjadi SA, Patterson K, Tomek M, & Nestor PJ (2012). Abnormalities of connected speech in semantic dementia vs Alzheimer's disease. Aphasiology, 26(6), 847–866.

Snowdon DA, Greiner LH & Markesbery WR (2000). Linguistic ability in early life and the neuropathology of Alzheimer's disease and cerebrovascular disease: findings from the nun study. Annals of the New York Academy of Sciences, 903, 34–38. [PubMed: 10818486]

Spieler DH, & Griffin ZM (2006). The influence of age on the time course of word preparation in multiword utterances. Language and Cognitive Processes, 21(1–3), 291–321.

Taler V, & Phillips NA (2008). Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review. Journal of clinical and experimental neuropsychology, 30(5), 501–556. [PubMed: 18569251]

Templin MC (1957). Certain language skills in children: Their development and interrelationships. Minneapolis: University of Minnesota Press.

Themistocleous C, Ficek B Webster K, den Ouden DB, Hillis AE, & Tsapkini K (under review). Automated speech segmentation and classification of Primary Progressive Aphasia variants from a picture description task via Deep Neural Network Architecture.

Thompson CK, Cho S, Hsu CJ, Wieneke C, Rademaker A, Weitner BB,…Weintraub S (2012). Dissociations between fluency and agrammatism in primary progressive aphasia. Aphasiology, 26(1), 20–43. [PubMed: 22199417]

Thorne J, & Faroqi-Shah Y (2016). Verb production in aphasia: Testing the division of labor between syntax and semantics. Seminars in Speech and Language, 37(1), 23–33. [PubMed: 26882362]

Tsapkini K, Faria A, Harris A, Desmond J, Rapp B, Frangakis C, … Hillis A (2016). Imaging tDCS intervention effects in primary progressive aphasia. Frontiers in Psychology. doi: 10.3389/conf.fpsyg.2016.68.00135

Toutanova K, Klein D, Manning CD, & Singer Y (2003, 5). Feature-rich part-of-speech tagging with a cyclic dependency network In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1 (pp. 173–180). Association for Computational Linguistics.

Tsantali E, Economidis D, & Tsolaki M (2013). Could language deficits really differentiate Mild Cognitive Impairment (MCI) from mild Alzheimer's disease?. Archives of gerontology and geriatrics, 57(3), 263–270. [PubMed: 23628238]

Tyas SL, Snowdon DA, Desrosiers MF, Riley KP, & Markesbery WR (2007). Healthy ageing in the Nun Study: definition and neuropathologic correlates. Age and ageing, 36(6), 650–655. [PubMed: 17906306]

Westbury C, & Bub D (1997). Primary progressive aphasia: a review of 112 cases. Brain & Language, 60(3), 381–406. [PubMed: 9398390]

Wilson SM, Henry ML, Besbris M, Ogar JM, Dronkers NF, Jarrold W,…Gorno-Tempini ML (2010). Connected speech production in three variants of primary progressive aphasia. Brain, 133, 2069–2088. [PubMed: 20542982]

Winblad B, Palmer K, Kivipelto M, Jelic V, Gratiglioni L,, Wahlund LO, …Peterson RC (2004). Mild cognitive impairment—beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. Journal of Internal Medicine, 256, 240–246. [PubMed: 15324367]

Highlights

- PPA scores are lower than healthy adults and MCI on fluency, word retrieval, and grammaticality
- PPA and AD groups did not differ on language measures
- Agrammatic PPA scores are lower than logopenic and semantic PPA on rate, mean length and verbs/utterance

Table 1 –

Age and Education Data for Experimental and Control Groups

	Controls (n=24)	MCI (n=20)	AD (n=20)	PPA (n=26)
Age (years) Mean (SD)	68.75 (7.21)	66.10 (9.18)	68.00 (7.37)	68.62 (7.79)
Education (years) Mean (SD)	15.04 (2.87)	14.65 (3.27)	14.55 (2.37)	16.12 (2.92)

Table 2 –

Language analysis measures used in the current study.

Domain	Measure	Calculation Method
Fluency	*Words per minute*	Total # of words/total length of audio in minutes The total number of intelligible, whole words per 5-utterance segment was manually counted and divided by the speaker's total speaking time as measured in Praat (Boersma, 2001).
	Total number of disfluencies	Disfluencies were manually coded and their numbers were extracted using CLAN's FREQ command
Word Retrieval	*Moving Average Type Token Ratio (MATTR)*	KidEVAL utility in CLAN provides the value
	Idea Density	# of propositions (non-noun entities)/ 10 words EVAL utility in CLAN provides the value
	Correct Information Units (CIU)	Total # correct information units/ Total # words in sample This was manually calculated
	Total number of semantic errors	Semantic paraphasias were manually coded and their numbers were extracted using CLAN's FREQ command
	Total number of phonological errors	Phonological paraphasias, phonemic paraphasias and neologisms were manually coded, and their numbers were extracted using CLAN's FREQ command
	Total number of word retrieval errors	The numbers of semantic, phonological and mixed paraphasias were extracted using CLAN's FREQ command and added up
Syntax	*Mean length of utterance (MLU)*	Total # of morphemes in a sample/ total # of utterances in a sample EVAL utility in CLAN provides the value
	Proportion of grammatical utterances	Manual coding of grammaticality of each utterance EVAL utility in CLAN provides the value
	Verbs per utterance	Average # verbs/ Total # utt EVAL utility in CLAN provides the value
	Sentence Complexity	Manual coding of complex sentence types (e.g., sentences containing embedded clauses, modifying clauses, passive sentence construction) The number of sentences is extracted using CLAN's FREQ command
	Total number of morphological errors	Morphological errors were manually coded and their numbers were extracted using CLAN's FREQ command
General	*Total number of errors*	Sum of disfluencies, semantic, phonological and morphological errors

Abbreviations: CLAN = Computerized analysis of language (MacWhinney, 2000), EVAL & FREQ = utilities in CLAN. EVAL computes a wide range of language measures, FREQ computes the number of instances of a specified language unit.

Table 3 –

Language scores for the participant groups and the results of the between-group statistical comparisons (right-most column)

Measures	Healthy Mean (SD) [95% CI]	PPA Mean (SD) [95% CI]	MCI Mean (SD) [95% CI]	AD Mean (SD) [95% CI]	ANOVA F p
Fluency					
Average WPM	141.73 (39.51) [126.09 – 157.37]	76.91 (39.50)[a, b] [61.88 – 91.94]	123.76 (31.21) [106.63 – 140.90]	116.61 (42.81) [99.47 – 133.74]	F(3,86) = 12.61 $p < .001$ *
Total Disfluencies	9.46 (7.10) [3.98 – 14.93]	27.57 (19.17)[a, b] [22.32 – 32.83]	10.50 (9.26) [4.50 – 16.50]	12.90 (13.88) [6.90 – 18.90]	F(3,86) = 9.71 $p < .001$ *
Word Retrieval					
MATTR	.86 (.03) [.84 – .88]	.83 (.08) [.81 – .85]	.86 (.04) [.84 – .88]	.87 (.04) [.84 – .80]	F(3,86) = 2.20 $p = .093$
Idea Density	.41 (.04) [.38 – .43]	.39 (.08) [.37 – .42]	.39 (.04) [.36 – .42]	.40 (.06) [.37 – .43]	F(3,86) = .27 $p = .870$
Proportion CIUs	.82 (.10) [.75 – .87]	.63 (.20)[a, b] [.57 – .68]	.80 (.12) [.74 – .87]	.72 (.18) [.65 – .78]	F(3,86) = 6.88 $p < .001$ *
No. Semantic Errors	.00 (.00) [−.24 – .24]	.46 (.81) [.22–.70]	.00 (.00) [−.27 – .27]	.25 (.91) [−.02 – .52	F(3,86) = 3.18 $p = .028$
No. Phonological Errors	.04 (.20) [−.24 – .33]	.69 (1.28) [.41 – .97]	.00 (.00) [−.31 – .31]	.05 (.22) [−.26 – .36]	F(3,86) = 5.35 $p = .002$ *
Total Word Retrieval Errors	.04 (.20) [−.38 – .46]	1.23 (1.75)[a, b] [.82 – 1.63]	.00 (.00) [−.46 – .46]	.30 (.92) [−.26 – .76]	F(3,86) = 7.37 $p < .001$ *
Syntax					
MLU, morphemes	9.93 (2.13) [8.70 – 11.16]	9.85 (3.81) [8.60 – 11.03]	11.05 (2.94) [9.70 – 12.39]	9.67 (2.85) [8.32 – 11.02]	F(3,86) = .87 $p = .458$
Proportion grammatical utterances	.89 (.16) [.80 – .96]	.66 (.28)[a, b] [.58 – .73]	.89 (.12) [.79 – .97]	.82 (.18) [.73 – .91]	F(3,86) = 6.92 $p < .001$ *
Verbs/utterance	1.22 (.27) [1.03 – 1.41]	1.40 (.68) [1.21 – 1.58]	1.30 (.35) [1.08 – 1.51]	1.23 (.44) [1.02 – 1.42]	F(3,86) = .69 $p = .557$
Complex Utterances	.47 (1.02) [.17 – .74]	.08 (.27) [−.19 – .35]	.60 (.75) [.28 – .91]	.35 (.58) [.03 – .66]	F(3,86) = 2.30 $p = .083$
Total Morphological Errors	.04 (.20) [−.21 – .29]	.50 (.99) [.25 – .74]	.05 (.22) [−.22 – .32]	.20 (.61) [−.07 – .47]	F(3,86) = 2.88 $p = .040$
General					
Total Errors	9.54 (7.16) [3.87 – 15.20]	29.31 (19.79)[a, b] [23.86 – 34.74]	10.55 (9.23) [4.34 – 16.75]	13.40 (14.77) [7.19 – 19.60]	F(3,86) = 10.82 $p < .001$ *

Abbreviations: AD= Alzheimer's Dementia, CIU = correct information units, MATTR = moving average type-token ratio, PPA= Primary Progressive Aphasia, MCI= Mild Cognitive Impairment, WPM = words per minute,

* $p < .003$,

[a] Post-hoc comparison with healthy group, $p < .05$,

[b] Post-hoc comparison with MCI, $p < .05$

Table 4.

Cut-off scores for significantly different language measures based on 95% confidence interval and the corresponding sensitivity and specificity values for differentiating between groups.

	Healthy vs PPA				MCI vs PPA		
Measure	**Cut-off**	**Sens**	**Spec**	**Z**	**Cut-off**	**Sens**	**Spec**
WPM	< 98.6	65.4	79.2	−1.6	< 99.74	65.4	75.0
Total Disfluencies	> 19.9	57.7	95.8	2.5	> 19.51	57.7	85.0
Proportion CIUs	< .7	65.4	87.5	−1.4	< .7	65.4	80.0
Total Retrieval Errors	> .6	53.8	95.8	5.9	> .61	53.8	100.0
Prop. Grammatical Utt.	< .8	57.7	83.3	−1.4	< .8	57.7	80.0
Total Errors	> 21.4	57.7	95.8	2.7	> 21.1	57.7	87.5

Abbreviations: CIU = correct information units, Sens = Sensitivity, Spec = Specificity, WPM = words per minute, Z= z-score indicating how different the PPA group was from the neurologically healthy group.

Table 5 –

Language scores for PPA variants and the results of the between-group statistical comparisons (right-most column)

Measures	PPA -G Mean (SD) [95% CI]	PPA-L Mean (SD) 95% CI	PPA-S Mean (SD) 95% CI	K-W χ^2 (df =2) p
Fluency				
Avg WPM	43.71 (17.70)[a] [31.04 – 56.37]	85.08 (31.11) [61.16 – 109.01]	113.83 (34.03) [82.41 – 145.26]	15.01 $p = .001$ *
Total Disfluencies	28.80 (22.42) [12.76 – 44.84]	33.33 (17.74) [19.69 – 46.98]	8.43 (14.40) [5.10 – 31.75]	1.95 $p = .376$
Word Retrieval				
MATTR	.79 (.11) [.70 – .86]	.85 (.55) [.80 – .89]	.87 (.01) [.84 – .89]	4.61 $p = .100$
Idea Density	.34 (0.06) [.30 – .38]	.44 (.79) [.37 – .49]	.43 (.44) [.39 – .47]	11.04 $p = .004$
Proportion CIUs	.66 (.16) [.54 – .77]	.63 (.18) [.48 – .76]	.59 (.17) [.43 – .74]	.71 $p = .699$
No. Semantic Errors	.40 (.69) [−.10 – .90]	.33 (.70) [−.21 – .88]	.71 (1.11) [−.31 – 1.74]	0.75 $p = .685$
No. Phonological Errors	.90 (1.28) [−.02 – 1.82]	.67 (1.65) [−.61 – 1.94]	.43 (0.78) [−.30 – 1.16]	1.35 $p = .507$
Total Word Retrieval Errors	1.40 (1.83) [.09 – 2.71]	1.00 (1.65) [−.27 – 2.27]	1.29 (1.97) [−.54 – 3.11]	0.78 $p = .670$
Syntax				
MLU, morphemes	6.51 (2.95)[a] [4.39 – 8.62]	12.44 (2.77) [10.31 – 14.56]	11.30 (2.45) [9.02 – 13.56]	13.98 $p = .001$ *
Proportion grammatical utterances	.51 (.32) [.27 – .74]	.71 (.25) [.51 – .89]	.81 (.15) [.67 – .95]	4.56 $p = .510$
Verbs/utterance	.75 (.48)[a] [.40 – 1.09]	1.80 (.43) [1.46 – 2.13]	1.82 (.46) [1.40 – 2.20]	16.48 $p < .001$ *
Complex Utterances	.00 (.0) -	.11 (.33) [−.15 – .37]	.14 (.37) [−.21 – .49]	1.35 $p = .508$
Total Morphological Errors	.40 (.69) [−.10 – .90]	.78 (1.39) [−.29 – 1.85]	.29 (.75) [−.41 – .98]	0.72 $p = .695$
General				
Total Errors	30.60 (22.37) [14.59 – 46.61]	35.11 (18.65) [20.77 – 49.45]	20.00 (16.22) [4.99 – 35.01]	1.97 $p = .372$

Mean (SD) for measures in between groups comparisons. Abbreviations: K-W χ^2 = Kruskal-Wallis chi-square; PPA-G = agrammatic PPA; PPA-L = logopenic PPA; PPA-S = semantic PPA,

* $p < .003$,

[a] Post-hoc comparison with PPA-S and PPA-L groups, $p < .05$.