The Role of Word Properties in Performance on Fluency Tasks in People with Primary Progressive Aphasia

Adrià Rofes[a,b,*], **Vânia de Aguiar**[b], **Bronte Ficek**[b], **Haley Wendt**[b], **Kimberly Webster**[b], and **Kyrana Tsapkini**[b,1]

[a]Global Brain Health Institute, Trinity College Dublin, Dublin, Ireland [b]Department of Neurology, Johns Hopkins Medicine, Baltimore, MD, USA

Abstract

People with primary progressive aphasia (PPA) present language difficulties that require lengthy assessments and follow-ups. Despite individual differences, people with PPA are often classified into three variants that present some distinctive language difficulties. We analyzed the data of 6 fluency tasks (i.e., "F", "A", "S", "Fruits", "Animals", "Vegetables"). We used random forests to pinpoint relevant word properties and error types in the classification of the three PPA variants, conditional inference trees to indicate how relevant variables may interact with one another and ANOVAs to cross-validate the results. Results indicate that total word count helps distinguish healthy individuals (N = 10) from people with PPA (N = 29). Furthermore, mean familiarity differentiates people with svPPA (N = 8) from people with lvPPA (N = 10) and nfvPPA (N = 11). No other word property or error type was relevant in the classification. These results relate to previous literature, as familiarity effects have been reported in people with svPPA in naming and spontaneous speech. Also, they strengthen the relevance of using familiarity to identify a specific group of people with PPA. This paper enhances our understanding of what determines word retrieval in people with PPA, complementing and extending data from naming studies.

Keywords

Category; familiarity; fluency; letter; primary progressive aphasia

INTRODUCTION

People with primary progressive aphasia (PPA) have a neurodegenerative disease that first and foremost affects language abilities [1, 2]. Despite individual differences, these individuals are clinically divided into three variants presenting distinctive language difficulties [3, 4]. Individuals with semantic variant PPA (svPPA) tend to have damage in lexical-semantics, as shown by low object naming scores and difficulties with object

[1]Kyrana Tsapkini, PhD, Department of Neurology, Cerebrovascular Division, 600 N. Wolfe Street, Phipps 488, Baltimore, MD 21287, USA. tsapkini@jhmi.edu.. *Correspondence to: Adrià Rofes, PhD, Global Brain Health Institute, Trinity College Dublin, College Green, 2, Dubin, Ireland. Tel.: +353 0 1 896 85 00; a.rofes@atlanticfellows.org.

knowledge; people with logopenic variant PPA (lvPPA) have difficulties in lexical-phonology and short-term verbal working memory, as shown by phonological errors in spontaneous speech and low scores in sentence repetition. Finally, individuals with non-fluent variant PPA (nfvPPA) have difficulties in certain aspects of grammar (e.g., verb inflection) and motor aspects of speech.

Throughout the course of the disease, people with PPA undergo multiple language assessments tapping into different levels (e.g., lexical-semantics, phonology, grammar). This is necessary to understand the progression of the disease, to plan behavioral treatments, and to obtain objective information relative to treatment outcomes. These assessments are often lengthy, require the input of a language expert, and many of the tasks that are administered use predetermined stimuli (e.g., pictures), constraining the responses of patients [5–7].

In this paper, we wanted to understand whether it is possible to classify PPA variants by using fluency tasks, enhancing our understanding of what determines word retrieval in people with PPA. In fluency tasks, individuals are given one minute to say as many words as possible starting with a specific letter of the alphabet (e.g., "F", "A", "S"), or to say words of a specific kind or category (e.g., animals, fruits, nouns, verbs). We chose these tasks because they can be rapidly administered and scored, are commonly used among clinicians, do not require predetermined stimuli, potentially do not bias the results, and because they may be less prone to test re-test effects, hence, suitable to be used multiple times throughout the course of the disease [8–12].

To do that, we ran a classification analysis using machine learning algorithms on the word properties and error types that these individuals make in fluency tasks. We looked at word properties of fluency tasks because these are indicative of specific difficulties at an underlying language level [3, 13]. For example, people with a lexical-semantic impairment tend to show effects for semantic word properties such as imageability or familiarity, while individuals with impairments in post-lexical or output buffer damage may show difficulties with word length as measured with phonemic counts [13–16]. This information is relevant to identify the kind of impairment that a patient has, and manipulations of these word properties are typically included in standardized tests to identify such impairments [17].

Furthermore, there is a literature on the use of word properties in fluency tasks in people with neurodegeneration. Forbes-McKay et al. [18] found that people with Alzheimer's disease (AD) produced items that were more frequent, more typical, shorter in length, and acquired earlier than healthy individuals. In addition, age-of-acquisition was a significant predictor of disease severity and the best predictor of group membership compared to the other variables. Marczinski and Kertesz [9] found that people with AD and people with svPPA did not differ from one another but produced fewer and more frequent words than healthy individuals. Finally, Vita et al. [19] found that people with AD and people with mild cognitive impairment (MCI) produced words with higher typicality than healthy individuals (e.g., pigeons are more representatives of the category bird than ostriches). Also, in a two-year follow-up, individuals with MCI who produced more typical words were more prone to develop AD compared to individuals with MCI with lower typicality values.

Two issues of analyzing word properties are that there are many word properties that are commonly reported, and that many are intercorrelated [20–24]. For fluency tasks, this means that it is hard to decide which word properties may be more informative to classify individuals with a specific neurological disorder, or even if any word property may be more informative than commonly used measures (e.g., the total number of words). To shed light on this, we used random forests analysis—a machine learning algorithm that deals well with relatively small datasets and intercorrelated values. Additionally, we used another machine learning algorithm, namely conditional inference trees, to understand whether mean scores above or below a certain number for a specific variable are more likely to occur in one PPA variant as opposed to others.

To the best of our knowledge, this is the first study to investigate a large number of word properties of fluency tasks in people with PPA. We do not envision these tasks to substitute comprehensive language assessments as a whole. However, they may facilitate the referral of patients with language impairments to more specialized services, and to provide hints regarding the kind of language impairments patients may have. Similar papers have looked at the total number of words [25], the latency in recalling words [26], error types [27], clusters and switches of words starting with the same phoneme or words of the same category [28], and even neuroanatomical differences between tasks or participants [29]. Below is a summary of the literature on word properties of fluency tasks in people with neurodegeneration and a section on initial predictions of which word properties may show effects in people with PPA.

Effects of word properties in people with neurodegeneration

Age of acquisition ratings indicate when individuals learned a word in the spoken or written form [30–32]. They have been typically related to processes of word retrieval, particularly in the output lexicon (for a discussion, see [32]). Age of acquisition predicts object naming accuracy in people with svPPA [33].

Concreteness ratings indicate "the degree to which the concept denoted by a word refers to a perceptible entity" [34]. For example, "couch" and "cobra" are concrete because we directly experience them through our senses. Contrary to that, "hope" and "ideal" are abstract words. Individuals with and without brain damage typically perform better on comprehension tasks with concrete words [35, 36]. The reverse pattern, that is, better performance for abstract words as opposed to concrete words, has been reported in people with svPPA [37].

Frequency estimates are obtained by counting how many times a word appears in a corpus. Wilson et al. [38] studied spontaneous speech during the description of a picture and found that individuals with svPPA produced nouns of higher frequency compared to healthy controls, lvPPA, nfvPPA, and behavioral variant frontotemporal dementia (bvFTD). The findings for svPPA of Wilson et al. [38] resemble an early study by Bird et al. [39].

Familiarity is a measure of how often people are in contact with or use certain words. For example, the "lemur" and "vertex" are lower in familiarity than "zebra" or "quarter" [40]. In spontaneous speech, Fraser et al. [14] found that people with svPPA use words that are higher in familiarity (and higher frequency) compared to healthy individuals. Similar

findings for familiarity have been reported in previous studies, including studies with naming tasks [15, 16].

Imageability ratings indicate how well a word gives rise to a sensory experience or mental image. For example, "house" and "apple" typically rate high in imageability and "hope" or "fact" rate low in imageability [41]. Healthy individuals as well as people with post-stroke aphasia have more difficulty processing low imageability words compared to high imageability words [23, 42, 43]. However, Bird et al. [39] found that individuals with svPPA produced words of low imageability during picture description. This same pattern has been reported in healthy individuals [44], people with post-stroke aphasia [45], and in individuals with semantic dementia [46] in naming.

Length in phonemes is measured by counting the number of phonemes of a word (e.g., apple/(ə)/has 4 phonemes). Lambon Ralph et al. [15] presented data from an individual with semantic dementia who named pictures with longer names better. We argue that this could be due to a deficit in the phonological output lexicon, as longer words have fewer phonological neighbors than shorter words and, therefore, are easier to activate than shorter words. Also, Fraser et al. [14] indicated that individuals with nfvPPA produce shorter words in spontaneous speech compared to healthy individuals. Semantic association is an objective measure of word relatedness [47]. In spontaneous speech, individuals with svPPA produced groups of words with lower semantic association values compared to healthy controls [48].

Orthographic similarity and phonological similarity are measures of lexical neighborhood. The former is obtained by counting the number of words that arise by substituting one letter of the target word in a given corpus [49]. The latter is obtained with the same procedure, but instead of substituting one letter, we substitute a phoneme. We used these measures to understand whether participants used clusters of words that are phonologically or orthographically similar, as this is a strategy reported to occur in letter fluency tasks [50]. Therefore, participants that use this strategy will tend to produce words with higher orthographic or phonological neighbors.

Predictions for people with PPA

Word properties that are associated with language performance in the different variants of PPA may add value to classification analyses. For example, word properties such as imageability, familiarity, concreteness, and semantic association show effects in people with lexical-semantic impairments. Therefore, differences in the mean value of these variables may accurately distinguish between individuals with svPPA and other PPA variants. Other word properties such as frequency, length in phonemes, orthographic and phonological similarity show effects in people with lexical-phonological impairments. Therefore, differences in this variable may be relevant to classify people with lvPPA. Finally, in this study it may be hard to point to variables that classify individuals with nfvPPA other than by exclusion. A variable such as verb proportion could have been considered as an indicator of grammatical processing. However, this measure would have only been valuable in letter fluency, as in most category fluency tasks all words are nouns. Also, in the presence of noun-verb homonymy it would have been impossible to decide whether patients were producing nouns or verbs (e.g., "ferment" versus "to ferment", "saw" versus "to saw"). In

Fig. 1, we present a summary of the predictions for which word properties may show effects in people with PPA in this study.

In sum, our three predictions are that 1) mean value familiarity, imageability, and concreteness, may accurately distinguish between individuals with svPPA and other PPA variants; 2) mean frequency, length in phonemes, orthographic and phonological similarity may distinguish between people with lvPPA and other PPA variants; and 3) we may not accurately classify nfvPPA with the current word properties/tasks.

METHODS

Participants

39 right-handed native English speakers participated in this study. Twenty-nine of these participants were people with PPA; the remaining 10 were non-brain damaged individuals matched for age (U = 122.500, p = 0.787 two-tailed) and education (U = 63.000, p = 0.752 two-tailed) with the PPA group. The group of individuals with PPA was composed of 12 females and 17 males with mean age 68 (sd = 7), mean education 17 years (sd = 2), and mean years post diagnosis 4.9 (sd = 3). Their language score in the FTLD-CDR was 1.9 (sd = 0.81) and their mean overall severity was 6.95 (sd = 3.7) on the FTLD-CDR scale [51]. Participants with a Mini-Mental State Examination score below 13 were excluded from the study. Based on the Gorno-Tempini et al. [1] classification, an assessment of language production and comprehension distinguished 10 people as lvPPA, 11 nfvPPA, and 8 svPPA. Individuals with apraxia of speech as their main or only deficit were not included in this study. Apraxia of speech was measured with an apraxia battery [52]. Demographic and test information per group is included in Table 1. Atrophy profiles per participant were determined by the referring neurologist. All procedures involving experiments on human subjects were done in accordance with the ethical standards of The Johns Hopkins Medicine Institutional Review Boards (JHM IRBs).

Fluency tasks

Six different fluency tasks were administered. These included three letter fluency tasks (in one minute, say as many words as possible starting with "F", "A", "S") and three category fluency tasks (in one minute, say as many words as possible that are "Fruits", "Animals", "Vegetables"). The tasks were part of a broader assessment protocol. They were administered consecutively in the same session and in the same order to all participants (i.e., "F", "A", "S", "Fruits", "Animals", "Vegetables"). The instructions for the phonological fluency tasks were as follows: "I'd like to you tell me as many words as you can that start with a certain letter. I'll give you one minute. OK? Tell me as many words as you can that start with the letter F (or A, or S)." These instructions are similar to Nasreddine et al. [64]. Note that proper nouns and words beginning with the same sound and having a different derivational suffix (e.g., *friend, friendly, friendliness*) were counted as correct, and instructions saying that those were incorrect were not included. This was done to facilitate the understanding of the instructions, given that some of our participants had comprehension difficulties. The instructions for the semantic tasks were as follows: "Can you tell me as

many fruits (or animals, or vegetables) as you can think of? They can start with any letter."
These instructions are the same as Mioshi et al. [65].

Fluency scoring

Words starting with the target letter of the alphabet or corresponding to the target semantic
category were considered correct. We also counted as correct proper nouns (e.g., *Anthony,
Alabama, Antarctica*), derived words (e.g., *friend, friendly, friendliness*), different types of
the same food (e.g., *Jack cheese* and *Manchego cheese*), animals that differ by sex (e.g., *bull*
and *cow*) and counted in both the superordinate category and its members (e.g., *bird* and
robin, sparrow). We considered as incorrect: repetitions, inflected words (e.g., *banana,
bananas*), fragments (i.e., retrieving the beginning of the word, one of more initial phonemes
—the patient is typically trying to find the rest of the word, e.g., "co" for "cockatoo"),
phonological paraphasias (i.e., switching, omitting, or adding phonemes to less than 50% of
the word's root, e.g., "lettush" for "lettuce"), neologisms (i.e., using a word that is
unintelligible and without any resemblance to the target word, more than 50% of the
phonemes are changed, e.g., "seebesh", "giger-hand"), and words beginning with a different
letter or from another word category.

Word properties

Correct words were rated for nine word properties extracted from different databases: age of
acquisition [66], concreteness [34], familiarity (MRC database [67]), frequency (CELEX
database containing written and spoken corpora [68]), imageability (6,377 word database,
combining Coltheart et al. [67], Juhasz et al. [69], and Stadthagen-Gonzalez & Davis [70]),
length in phonemes (N-WATCH [60]), semantic association (LSA [46]), orthographic
similarity [71], and phonological similarity [71]. Age of acquisition and imageability scores
were obtained for all words in the singular form (e.g., apple) as the databases did not include
ratings of the same words in the plural form (e.g., apples). Specific to the imageability
scores, any time we encountered repeated values between databases we maintained the
scores by Coltheart et al. [67]. Frequency scores for plural words were reported when they
were included in the database. Semantic Association scores were obtained with the LSAfun
package in R [47].

Statistics

We conducted three separate analyses. First, we analyzed the data for letter and category
fluency together. Second, we analyzed the data for letter fluency separately. Third, we
analyzed category fluency separately. All analyses were conducted with the R software [72]
and the statistical method for each of the analyses was the same:

First, for each analysis, the database included all the correct words said by each participant.
For each word, we obtained 9 word properties (frequency, imageability, age-of-acquisition,
familiarity, concreteness, length in phonemes, orthographic neighborhood, phonological
neighborhood, semantic association) and 6 error types (repetition, fragment, phonological
paraphasia, neologism, wrong category, wrong letter). Provided that some of the variables
were obtained with questionnaires, as opposed to being collected in large corpora, we could
not always retrieve values for each word property. In cases where missing values did not

exceed 50% [73], we estimated missing values using random forests imputation (rfImpute function [21]).

Second, random forests was used to rank variables (total number of words, 9 word properties of fluency tasks, 6 error types) in terms of their relative importance in the classification of PPA variant. This algorithm was used over other machine learning algorithms because it effectively deals with large numbers of predictors that may be inter-correlated and small sample sizes [74]. Random forests is a machine learning algorithm that provides a ranking of variable importance [75]. In this ranking, variables that are above a certain threshold show up as informative to explain/classify a dependent variable. For example, we may find that familiarity, imageability, and length in phonemes can be used to classify individuals with PPA, while factors such as frequency or orthographic similarity may not be as informative.

For the computation of random forests, we performed the following 7 analysis steps; for additional details, see [76, 77].

1. Estimation of missing values using random forests for data imputation [21].

2. Generation of a random forest with unbiased conditional inference trees [75], using the cforest function [78].

3. Extraction of the relative importance of each of the predictors using conditional permutation variable importance [75]. To do so, we used the varimp function [78]. Importance is a measure of prediction of the dependent variable (in our case, PPA variant). In that sense, when the variable is important, the model gains in accuracy [75]. While traditional variable importance measures over-estimate variable importance in the presence of correlations between the predictors, the varimp function uses *conditional* importance. Conditional importance mimics the behavior of partial correlations, and can therefore better account for the independent contributions of each variable, even if there are correlations between predictors [75].

4. Evaluation of the imputation processes by repeating data imputation and the previous two steps 20 times. This generated 20 separate datasets. We chose one dataset to be used in the following analyses by calculating the distance to the mean in variable importance across 20 datasets, so that the dataset used was that with the most representative values.

5. Generation of a random forest, as in step (1), using the selected dataset.

6. Extraction of variable importance, as in step (3).

7. Calculation of the estimation predictor accuracy including only potentially informative predictors using leave-one-out cross-validation. This means that the classifier is trained on a dataset where one data point is left out. The value of the observation left out is predicted and saved. This procedure is repeated until each data point has been left out once. Finally, we compared the actual values with the predicted values and, in this way, we evaluated the accuracy of predictions. For this comparison, we computed a confusion matrix to obtain sensitivity and specificity values for the classification of PPA variant. Accuracy scores of 100%

would reflect a perfect classification, along with sensitivity and specificity scores of 1.

Third, to illustrate how variables interact—a feature that random forests does not allow—we implemented another machine learning algorithm, the conditional inference trees [78]. These perform a series of statistical tests to identify points along a variable's scale (a split point) at which the prediction of values in the dependent measure change significantly. For example, individuals belonging to a certain group may more typically provide words with a word property higher than a certain value, while the rest of individuals provide words with a word property lower or equal than a certain value. This procedure is done for all variables selected based on their variable importance ranking, and hence a hierarchical, tree-like representation is produced, with nodes representing split points for significant variables. Conditional inference trees use inferential statistics to determine whether a split point results in significantly different distributions of the dependent measure, using the ANOVA F statistic for continuous variables and the χ^2 statistic for categorical variables [78].

Fourth, we cross-validated the results with a factorial ANOVA for each dependent measure deemed important by Random Forests. The ANOVA included each dependent measure (number of words, word properties, error types) and the factor participant (4 levels: logopenic, semantic, nonfluent, non-brain-damaged controls). *Post-hoc* TukeyHSD tests with adjusted p-values for multiple comparisons were conducted to study pairwise differences within the factor Variant.

RESULTS

In Table 2, we included the mean and standard deviation per task and dependent variable across groups.

Classification for letter and category fluency combined

Random forests indicated that, after leave-one-out cross-validation, the accuracy of predictions was 51% (CI: 0.3478–0.6758); note that chance level given 4 groups of ~10 participants is 25%. The sensitivity and specificity to classify each group was as follows: svPPA (sensitivity = 0.44, specificity = 0.86), lvPPA (sensitivity = 0.034, specificity = 0.77), nfvPPA (sensitivity = 0.34, specificity = 0.74), and NBD (sensitivity = 0.83, specificity = 1). See confusion matrix in Table 3. The most informative variables in the classification of PPA variant were total number of words, familiarity, length in phonemes, frequency, age of acquisition, repetition, concreteness, semantic association, and imageability; the rest of the variables were not informative.

A conditional inference tree was computed by using the variable ranking produced by random forests. The conditional inference tree identified an interaction between total number of words and familiarity (Fig. 2). The interaction indicated that NBD individuals are classified when they produce more than 75 words, while people with PPA are classified as such when they produce 75 or fewer words ($\chi^2 = 30.668$, $p = 0.001$). Additionally, a greater proportion of individuals are classified as having svPPA when their mean word familiarity is above 557.6 ($\chi^2 = 10.201$, $p = 0.01$).

Cross-validating the results, the ANOVAs indicated that for total number of words, there was a main effect of Variant (F(3,35) = 48.8, p = 0.0001) whereby NBD individuals produced significantly more words than people with svPPA (p = 0.001), nfvPPA (p = 0.001), and lvPPA (p = 0.001). The other comparisons were not significant (Fig. 3a). Also, for familiarity, there was a main effect of variant (F(3,35) = 8.812, p = 0.0001). People with svPPA produced more familiar words than nfvPPA (0.004), lvPPA (0.005), and NBD (0.001). The other differences were not significant (Fig. 3b).

Classification for letter fluency

Random forests indicated that, after leave-one-out cross-validation, the accuracy of predictions was 49% (CI: 0.3242–0.6522). The sensitivity and specificity to classify each group was as follows: svPPA (sensitivity = 0.42, specificity = 0.84), lvPPA (sensitivity = 0.00, specificity = 0.69), nfvPPA (sensitivity = 0.42, specificity = 0.8), and NBD (sensitivity = 0.83, specificity = 1). See confusion matrix in Table 4. The most informative variables in the classification of PPA variant were total number of words, length in phonemes, age of acquisition; the rest of the variables were not informative.

A conditional inference tree was computed by using the variable ranking produced by random forests. The conditional inference tree showed that NBD individuals are classified when they produce more than 38 words, while people with PPA are classified as such when they produce 38 or less words (χ^2 = 28.5577, p = 0.001). We found no significant results for length in phonemes or age of acquisition. Additionally, cross-validating the results, the ANOVAs indicated that for total number of words, there was a main effect of Variant (F(3,35) = 35.28, p = 0.0001) whereby NBD individuals produced significantly more words than people with svPPA (p = 0.001), nfvPPA (p = 0.001), and lvPPA (p = 0.001). Also, people with svPPA produced significantly more words than nfvPPA (p = 0.02). The other comparisons were not significant.

Classification for category fluency

Random forests indicated that, after leave-one-out cross-validation, the accuracy of predictions was 41% (CI: 0.2557–0.579). The sensitivity and specificity to classify each group was as follows: svPPA (sensitivity = 0.14, specificity = 0.78), lvPPA (sensitivity = 0.33, specificity = 0.75), nfvPPA (sensitivity = 0.31, specificity = 0.73), and NBD (sensitivity = 0.77, specificity = 1). See confusion matrix in Table 5. The most informative variables in the classification of PPA variant were total number of words, length in phonemes, age of acquisition, semantic association, repetition, phonological paraphasia; the rest of the variables were not informative.

A conditional inference tree was computed by using the variable ranking produced by random forests. The conditional inference tree showed that NBD individuals are classified when they produce more than 25 words, while people with PPA are classified as such when they produce 25 or fewer words (χ^2 = 30.0234, p = 0.001). We found no significant results for length in phonemes, age of acquisition, semantic association, repetition, phonological paraphasia. Cross-validating the results, the ANOVAs indicated that for total number of words, there was a main effect of Variant (F(3,35) = 43.91, p = 0.0001) whereby NBD

individuals produced significantly more words than people with svPPA ($p = 0.001$), nfvPPA ($p = 0.001$), and lvPPA ($p = 0.001$). The other comparisons were not significant.

DISCUSSION

In this study, we used random forests to pinpoint relevant word properties and error types in the classification of three PPA variants, conditional inference trees to indicate how relevant variables may interact with one another, and ANOVAs to cross-validate the results. The goal of this paper was to use fluency tasks to provide further understanding of what determines word retrieval in people with PPA, complementing studies with that have used other tasks.

When considering the mean values for each of the word properties in letter and category fluency, total word count and familiarity were the most relevant variables to classify PPA variant. Similar results for the total word count were also obtained when considering the data for letter and category fluency tasks separately. The results for total word counts match the current literature, whereby people with language impairments and neurodegeneration produce fewer words in fluency tasks than healthy individuals [9]. These results were also reported in people with other neurological disorders [79–82].

The results for familiarity are novel for this population and task. These results match prediction 1, where we considered word properties that could be potentially affected in each of the PPA variants, and particularly, variables that may distinguish people with svPPA from other variants. We found that individuals with svPPA were most commonly classified based on their familiarity scores—a word property that is affected in people with lexical-semantic deficits. Further evidence for the relevance of familiarity to classify individuals with svPPA was found in the results of the ANOVAs. It is worth noting that familiarity effects have also been reported in people with svPPA in other tasks, namely, picture naming and spontaneous speech [14–16]. Such reports strengthen the validity of our findings in fluency tasks. However, in comparison to previous work, we considered a large number of word properties and statistical methods that deal well with intercorrelated values. Therefore, our results imply that familiarity is a particularly good predictor to classify people with svPPA, in comparison to other word properties, including other lexical-semantic properties such as imageability and concreteness where we found no effects (even in *post-hoc* ANOVA analyses). We will speculate that these word properties are different. For example, words such as "for", "any", "after", "and", "so" where used by many participants in the fluency tasks. These words are low in imageability and concreteness, as they do not refer to sensory experiences or perceptible entities. However, they are high in familiarity, as people often use them. Differences between familiarity, concreteness, and other word properties are assessed and discussed in much detail by Gernsbacher [83]. Even though the example we provided applies only to letter fluency, it is interesting to see that we only find familiarity effects when we consider letter and category fluency combined. Hence, our data do not support arguments on different search strategies between letter and category fluency, as separate analyses for these types of fluency tasks did not individualize relevant word properties.

Furthermore, word properties reflecting lexical-phonological damage such as word length, orthographic and phonological similarity, as well as frequency were not significant to

identify individuals with lvPPA. Consequently, prediction 2 regarding people with lvPPA was not entirely correct, as individuals with lvPPA tend to have more problems with sentence repetition and produce phonological paraphasias in spontaneous speech [1] and these characteristics do not necessarily need to map with a frequency or length effect or other when producing single words. In fact, the evidence pointing to specific word properties being damaged in people with lvPPA is scarce [38] and, to the best of our knowledge, there is no previous description of an extensive number of word properties of fluency tasks in people with lvPPA.

Also, confirming prediction 3 and as opposed to Fraser et al. [14] where it was indicated that people with nfvPPA produced shorter words compared to controls in spontaneous speech, we found no particular relevant role of word length to classify this variant. We did not predict finding such an effect of word length in people with nfvPPA, as we argued that the word properties we entered in this study were not particularly suited to classify the nfvPPA variant. In fact, the sensitivity to classify individuals with nfvPPA was particularly low. That is, none of our models provided high sensitivity to identify which individuals in our sample had nfvPPA.

None of the error types showed as relevant in the classification of PPA variants. This is possibly the case because the number of errors produced averaged around 10% (percentage of errors by total number of words), hence, it was rather low. Having a closer look at the data, people with nfvPPA produced the highest percentage of errors (24%), among these, the most common errors were repetitions (38%) and use of the wrong category (14%). Contrary to what would be expected given Gorno-Tempini et al. [1], people with lvPPA did not produce many more phonological paraphasias than the other groups (percentage of errors per group: lvPPA = 7%, nfvPPA = 4%, svPPA = 0%, NBD = 6%). This may be due to the fact that the error types that we are more commonly reported in naming and spontaneous speech do not always or necessarily translate to fluency tasks. Having said that, people with lvPPA produced relatively more neologisms (37% of all errors) than the other groups (nfvPPA = 10%, svPPA = 28%, NBD = 4%), which could be seen in line with difficulties at the phonological level. Finally, for people with svPPA, the most common error types were neologisms (28%) and use of the wrong category (22%), and the most common error amongst individuals without brain damage were repetitions (77%).

In summary, other than finding differences between healthy individuals and people with PPA in word counts, the main finding from this study is that a word property that has been related to lexical-semantic processing can help us to classify individuals with svPPA, a PPA variant that is typically associated with lexical-semantic damage. The fact that people with svPPA produce more familiar words (instead of less familiar words) than the other variants may be due to lexical or semantic difficulties (or in the mappings between them). For example, in the process of producing words in a fluency task, damage to the semantic system or its connection to the phonological output lexicon may generate less activation to the phonological output lexicon, making it easier to produce more familiar words. Also, semantic representations of low familiarity words may be more susceptible to damage, hence, producing more familiar words [15, 33, 82].

From a clinical standpoint, it is worth noting that these results were obtained only by adding an analysis of word properties to an assessment that takes 6–10 min to administer (~1 min for each fluency task and an additional time to explain the task). Also, the study of word properties of fluency tasks is promising as it is largely unencumbered by experimental stimuli. This is in contrast to other commonly used tasks such as picture naming, as these tasks use specific visual stimuli that sometimes may preclude the observation of certain effects [24]. For example, differences in imageability (i.e., how a word gives rise to a sensory experience or mental image) are hard to detect with picture naming, as words that are easy to draw are typically high in imageability, while words that are low in imageability are also difficult to draw, and many times not possible to use in picture naming experiments. Differently from picture naming, in fluency tasks, participants only need to say words of a specific category or starting with a specific letter. For example, if the task is to say words starting with "F", participants can say words with low imageability such as "for", "fuss", and "from" and words with high imageability such as "finger", "father", or "fireplace".

Similar arguments to the benefits of using fluency tasks over picture naming hold for spontaneous speech assessments, where participants are less constrained to produce specific words than in picture naming [30, 38, 48]. However, spontaneous speech assessments are very tedious to transcribe and to analyze, require analyses by a language expert, and seem more prone to error than transcribing the results of a fluency task and calculating the value of word properties based on already collected corpora. Nonetheless, spontaneous speech analyses have been shown to provide comparable results to regular language assessments [7], while such a study does not yet exist for fluency tasks.

Finally, some of the limitations of this study are that while random forests produced a ranking of variable importance, the predictive ability of the model was around 50% and, therefore, low. However, given 4 groups with of circa 10 participants, the results are above chance level (25%). Also, the confidence interval for the reported accuracy (i.e., CI: 0.3478–0.6758)] is above the 25% chance level. A related issue was the lack of separate sets of observations for feature selection versus calculating prediction accuracy. Given the small sample size, splitting the data further would have created fewer stable results. However, the chosen approach may have led to some degree of over fitting, and hence reduced generalizability of the accuracy data. Thus, future research should evaluate prediction accuracy on an independent set of observations. Another issue is that total word count and familiarity may not be unique to identifying individuals with svPPA or even PPA. To answer this question, similar studies adding other neurological populations with similar deficits should be encouraged. A follow-up to this study could include people with MCI and/or people with AD, as these individuals also have neurodegenerative disorders that are difficult to identify solely by their pathological profile. In this study we used a cut-off for patient severity and measured severity scores on the FTLD scale. However, attention should be paid to translating these results to the individual level, as particularly severe participants may produce too little words to obtain reliable results on any of the word properties. Finally, the low classification accuracy of the random forest may reflect issues with sample size, potential issues with the use of PPA type instead of nature of impairment as a predictor [4], or the value of fluency tasks or the word properties we used to classify PPA variant. Further work on these issues, along with replication of these findings in languages other than

English, should focus on increasing the overall accuracy of the models and the classification sensitivity for people with PPA.

Conclusions

The study of 9 word properties of fluency tasks indicated that total word number and mean word familiarity were the most relevant factors to classify individuals with PPA and especially to distinguish svPPA from the other variants. These findings are similar to studies on naming and spontaneous speech, providing further evidence to understand what determines word retrieval in people with PPA. Replication of these results in English and other languages seems necessary to validate the classification method. The familiarity finding has significant implications for treatment, since people with svPPA may have better outcomes, treatment should aim at preserving familiar items.

ACKNOWLEDGMENTS

REFERENCES

[1]. Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa S, Manes F (2011) Classification of primary progressive aphasia and its variants. Neurology 76, 1006–1014. [PubMed: 21325651]

[2]. Mesulam MM (1987) Primary progressive aphasia – differentiation from Alzheimer's disease. Ann Neurol 22, 533–534. [PubMed: 3324947]

[3]. Whitworth A, Webster J, Howard D (2014) A cognitive neuropsychological approach to assessment and intervention in aphasia: A clinician's guide, 2nd edition. Psychology Press, Hove, UK.

[4]. Wicklund MR, Duffy JR, Strand EA, Machulda MM, Whitwell JL, Josephs KA (2014) Quantitative application of the primary progressive aphasia consensus criteria. Neurology 82, 1119–1126. [PubMed: 24598709]

[5]. Kurland J, Reber A, Stokes P (2014) Beyond picture naming: Norms and patient data for a verb-generation task. Am J Speech Lang Pathol 23, S259–S270. [PubMed: 24686752]

[6]. McRae K, Hare M, Elman JL, Ferretti T (2005) A basis for generating expectancies for verbs from nouns. Mem Cognit 33, 1174–1184.

[7]. Rofes A, Talacchi A, Santini B, Pinna G, Nickels L, Bastiaanse R, Miceli G (2018) Language in individuals with left hemisphere tumors: Is spontaneous speech analysis comparable to formal testing? J Clin Exp Neuropsychol 40, 722–732. [PubMed: 29383968]

[8]. Cooper DB, Lacritz LH, Weiner MF, Rosenberg RN, Cullum CM (2004) Category fluency in mild cognitive impairment: Reduced effect of practice in test-retest conditions. Alzheimer Dis Assoc Disord 18, 120–122. [PubMed: 15494616]

[9]. Marczinski CA, Kertesz A (2006) Category and letter fluency in semantic dementia, primary progressive aphasia, and Alzheimer's disease. Brain Lang 97, 258–265. [PubMed: 16325251]

[10]. Rabin LA, Barr WB, Burton LA (2005) Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. Arch Clin Neuropsychol 20, 33–65. [PubMed: 15620813]

[11]. Rofes A, Mandonnet E, Godden J, Baron MH, Colle H, Darlix A, Wager M (2017) Survey on current practices within the European Low Grade Glioma Network: Language and cognitive assessment part. Acta Neurochir (Wien) 159, 1167–1178. [PubMed: 28474122]

[12]. Woods SP, Scott JC, Sires DA, Grant I, Heaton RK, Tröster AI, HIV Neurobehavioral Research Center (HNRC) Group (2005) Action (verb) fluency: Test–retest reliability, normative standards, and construct validity. J Int Neuropsychol Soc 11, 408–415. [PubMed: 16209421]

[13]. Shallice T (1988) From Neuropsychology to Mental Structure Cambridge University Press, Cambridge, UK.

[14]. 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]

[15]. Lambon Ralph MA, Howard D, Nightingale G, Ellis AW (1998) Are living and non-living category-specific deficits causally linked to impaired perceptual or associative knowledge? Evidence from a category-specific double dissociation. Neurocase 4, 311–338.

[16]. Woollams AM, Cooper-Pye E, Hodges JR, Patterson K (2008) Anomia: A doubly typical signature of semantic dementia. Neuropsychologia 46, 2503–2514. [PubMed: 18499196]

[17]. Swinburn K, Porter G, Howard D (2004) Comprehensive aphasia test (Vol 40), Psychology Press, Hove, UK.

[18]. Forbes-McKay KE, Ellis AW, Shanks MF, Venneri A (2005) The age of acquisition of words produced in a semantic fluency task can reliably differentiate normal from pathological age related cognitive decline. Neuropsychologia 43, 1625–1632. [PubMed: 16009244]

[19]. Vita MG, Marra C, Spinelli P, Caprara A, Scaricamazza E, Castelli D, Canulli S, Gainotti G, Quaranta D (2014) Typicality of words produced on a semantic fluency task in amnesic mild cognitive impairment: Linguistic analysis and risk of conversion to dementia. J Alzheimers Dis 42, 1171–1178. [PubMed: 25024315]

[20]. Cutler A (1981) Making up materials is a confounded nuisance, or: Will we be able to run any psycholinguistic experiments at all in 1990? Cognition 10, 65–70. [PubMed: 7198562]

[21]. Cutler F, Wiener R (2015) RandomForest: Breiman and Cutler's Random Forests for Classification and Regression (Version 46–12)

[22]. Gilhooly KJ, Watson FL (1981) Word age-of-acquisition effects: A review. Curr Psychol Rev 1, 269–286.

[23]. Nickels L, Howard D (1995) Aphasic naming: What matters? Neuropsychologia 33, 1281–1303. [PubMed: 8552229]

[24]. Nickels L, Howard D (1994) A frequent occurrence? Factors affecting the production of semantic errors in aphasic naming. Cogn Neuropsychol 11, 289–320.

[25]. Canning SD, Leach L, Stuss D, Ngo L, Black SE (2004) Diagnostic utility of abbreviated fluency measures in Alzheimer disease and vascular dementia. Neurology 62, 556–562. [PubMed: 14981170]

[26]. Rohrer D, Wixted JT, Salmon DP, Butters N (1995) Retrieval from semantic memory and its implications for Alzheimer's disease. J Exp Psychol Learn Mem Cogn 21, 1127. [PubMed: 8744958]

[27]. Mickanin J, Grossman M, Onishi K, Auriacombe S, Clark C (1994) Verbal and nonverbal fluency in patients with probable Alzheimer's disease. Neuropsychology 8, 385–395.

[28]. Pakhomov SV, Hemmy LS (2014) A computational linguistic measure of clustering behavior on semantic verbal fluency task predicts risk of future dementia in the Nun Study. Cortex 55, 97–106. [PubMed: 23845236]

[29]. Biesbroek JM, van Zandvoort MJ, Kappelle LJ, Velthuis BK, Biessels GJ, Postma A (2016) Shared and distinct anatomical correlates of semantic and phonemic fluency revealed by lesion-symptom mapping in patients with ischemic stroke. Brain Struct Funct 221, 2123–2134. [PubMed: 25939335]

[30]. Bird H, Franklin S, Howard D (2001) Age of acquisition and imageability ratings for a large set of words, including verbs and function words. Behav Res Methods 33, 73–79.

[31]. Carroll JB, White MN (1973) Word frequency and age of acquisition as determiners of picture-naming latency. Q J Exp Psychol 25, 85–95.

[32]. Hodgson C, Ellis AW (1998) Last in, first to go: Age of acquisition and naming in the elderly. Brain Lang 163, 146–163.

[33]. Ralph MAL, Graham KS, Ellis AW, Hodges JR (1998) Naming in semantic dementia—what matters? Neuropsychologia 36, 775–784. [PubMed: 9751441]

[34]. Brysbaert M, Warriner AB, Kuperman V (2014) Concreteness ratings for 40 thousand generally known English word lemmas. Behav Res Methods 46, 904–911. [PubMed: 24142837]

[35]. Jefferies E, Patterson K, Jones RW, Lambon Ralph MA (2009) Comprehension of concrete and abstract words in semantic dementia. Neuropsychology 23, 492–499. [PubMed: 19586212]

[36]. Sandberg C, Kiran S (2014) Analysis of abstract and concrete word processing in persons with aphasia and age-matched neurologically healthy adults using fMRI. Neurocase 20, 361–388. [PubMed: 23548150]

[37]. Grossman M, Ash S (2004) Primary progressive aphasia: A review. Neurocase 10, 3–18. [PubMed: 15849155]

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

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

[40]. Noble CE (1953) The meaning-familiarity relationship. Psychol Rev 60, 89–98. [PubMed: 13056100]

[41]. Paivio A, Yuille JC, Madigan SA (1968) Concreteness, imagery, and meaningfulness values for 925 nouns. J Exp Psychol 76, 1–25.

[42]. Cortese MJ, Schock J (2013) Imageability and age of acquisition effects in disyllabic word recognition. Q J Exp Psychol 66, 946–972.

[43]. Luzzatti C, Raggi R, Zonca G, Pistarini C, Contardi A, Pinna GD (2002) Verb–noun double dissociation in aphasic lexical impairments: The role of word frequency and imageability. Brain Lang 81, 432–444. [PubMed: 12081411]

[44]. Tyler LK, Moss HE, Galpin A, Voice JK (2002) Activating meaning in time: The role of imageability and form-class. Lang Cogn Process 17, 471–502.

[45]. Warrington EK (1981) Concrete word dyslexia. Br J Psychol 72, 175–196. [PubMed: 7248674]

[46]. Breedin SD, Saffran EM, Coslett HB (1994) Reversal of the concreteness effect in a patient with semantic dementia. Cogn Neuropsychol 11, 617–660.

[47]. Günther F, Dudschig C, Kaup B (2015) LSAfun-An R package for computations based on Latent Semantic Analysis. Behav Res Methods 47, 930–944. [PubMed: 25425391]

[48]. Hoffman P, Meteyard L, Patterson K (2014) Broadly speaking: Vocabulary in semantic dementia shifts towards general, semantically diverse words. Cortex 55, 30–42. [PubMed: 23261549]

[49]. Coltheart M, Davelaar E, Jonasson JT, Besner D, Dornic S (1977) Attention and performance Lawrence Erlbaum, Mahwah, NJ, USA.

[50]. Troyer AK, Moscovitch M, Winocur G (1997) Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. Neuropsychology 11, 138. [PubMed: 9055277]

[51]. Knopman DS, Kramer JH, Boeve BF, Caselli RJ, Graff-Radford N R, Mendez MF, Miller BL, Mercaldo N (2008) Development of methodology for conducting clinical trials in frontotemporal lobar degeneration. Brain 131, 2957–2968. [PubMed: 18829698]

[52]. Dabul B (1979) Apraxia Battery for Adults PRO-ED, Austin, TX, USA.

[53]. Kaplan E, Goodglass H, Weintraub S, Goodglass H (1983) Boston naming test Lea & Febiger, Philadelphia.

[54]. Breining BL, Tippett DC, Davis C, Posner J, Sebastian R, Oishie K (2015) Assessing dissociations of object and action naming in acute stroke. Clinical Aphasiology Conference, Monterey, CA, USA.

[55]. Howard D, Patterson KE (1992) The Pyramids and Palm Trees Test: A test of semantic access from words and pictures Thames Valley Test Company, Suffolk, UK.

[56]. Bak TH, Hodges JR (2003) Kissing and dancing—a test to distinguish the lexical and conceptual contributions to noun/verb and action/object dissociation. Preliminary results in patients with frontotemporal dementia. Thames Valley Test Company 16, 169–181.

[57]. Hillis A (2015) Sentence Repetition Test. NACC UDS-FTLD Neuropsychological battery instructions (form C1-F) (3.0) University of Washington, Washington, USA.

[58]. Goodman RA, Caramazza A (1985) The Johns Hopkins University Dysgraphia Battery Baltimore, MD, USA.

[59]. Love T, Oster E (2002) On the categorization of aphasic typologies: The SOAP (A Test of Syntactic Complexity). J Psycholinguist Res 31, 503–529. [PubMed: 12528429]

[60]. Schmidt M (1996) Rey auditory verbal learning test: A handbook Western Psychological Services, Los Angeles, CA, USA.

[61]. Wechsler D (1981) Wechsler Adult Intelligence Scale-Revised Psychological Corporation, San Antonio, TX, USA.

[62]. Corsi PM (1972) Human memory and the medial region of the brain. Diss Abstr Int 34, 819B.

[63]. Parkington JE, Leiter RG (1949) Parkington's pathway test. Psychol Serv Center Bull 1, 9–20.

[64]. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H (2005) The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. J Am Geriatr Soc 53, 695–699. [PubMed: 15817019]

[65]. Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR (2006) The Addenbrooke's Cognitive Examination Revised (ACE-R): A brief cognitive test battery for dementia screening. Int J Geriatr Psychiatry 21, 1078–1085. [PubMed: 16977673]

[66]. Kuperman V, Stadthagen-Gonzalez H, Brysbaert M (2012) Age-of-acquisition ratings for 30,000 English words. Behav Res Methods 44, 978–990. [PubMed: 22581493]

[67]. Coltheart M (1981) The MRC psycholinguistic database. Q J Exp Psychol 33, 497–505.

[68]. Baayen RH, Piepenbrock R, Gulikers L (1995) The CELEX lexical database (release 2). Distributed by the Linguistic Data Consortium, University of Pennsylvania, PA, USA.

[69]. Juhasz BJ, Lai YH, Woodcock ML (2015) A database of 629 English compound words: Ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. Behav Res Methods 47, 1004–1019. [PubMed: 25361864]

[70]. Stadthagen-Gonzalez H, Davis CJ (2006) The Bristol norms for age of acquisition, imageability, and familiarity. Behav Res Methods 38, 598–605. [PubMed: 17393830]

[71]. Davis CJ (2005) N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. Behav Res Methods 37, 65–70. [PubMed: 16097345]

[72]. R Core Team (2017) R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria.

[73]. Shah AD, Bartlett JW, Carpenter J, Nicholas O, Hemingway H (2014) Comparison of random forest and parametric imputation models for imputing missing data using MICE: A CALIBER study. Am J Epidemiol 179, 764–774. [PubMed: 24589914]

[74]. Breiman L (2001) Random forests. Mach Learn 45, 5–32.

[75]. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A (2008) Conditional variable importance for random forests. BMC Bioinformatics 9, 307. [PubMed: 18620558]

[76]. Tagliamonte SA, Baayen RH (2012) Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. Lang Var Change 24, 135–178.

[77]. Zhang HR, Min F (2016) Three-way recommender systems based on random forests. Knowl Based Syst 91, 275–286.

[78]. Hothorn T, Hornik K, Strobl C, Zeileis A (2017) Party: A Laboratory for Recursive Partytioning (Version 12–4)

[79]. Clark DG, Kapur P, Geldmacher DS, Brockington JC, Harrell L, DeRamus TP, Blanton PD, Lokken K, Nicholas AP, Marson DC (2014) Latent information in fluency lists predicts functional decline in persons at risk for Alzheimer disease. Cortex 55, 202–218. [PubMed: 24556551]

[80]. Pettit L, McCarthy M, Davenport R, Abrahams S (2013) Heterogeneity of letter fluency impairment and executive dysfunction in Parkinson's disease. J Int Neuropsychol Soc 19, 986–994. [PubMed: 23953916]

[81]. Zhao Q, Guo Q, Hong Z (2013) Clustering and switching during a semantic verbal fluency test contribute to differential diagnosis of cognitive impairment. Neurosci Bull 29, 75–82. [PubMed: 23322003]

[82]. Howard D, Gatehouse C (2006) Distinguishing semantic and lexical word retrieval deficits in people with aphasia. Aphasiology 20, 921–950.

[83]. Gernsbacher MA (1984) Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. J Exp Psychol Gen 113, 256–281. [PubMed: 6242753]
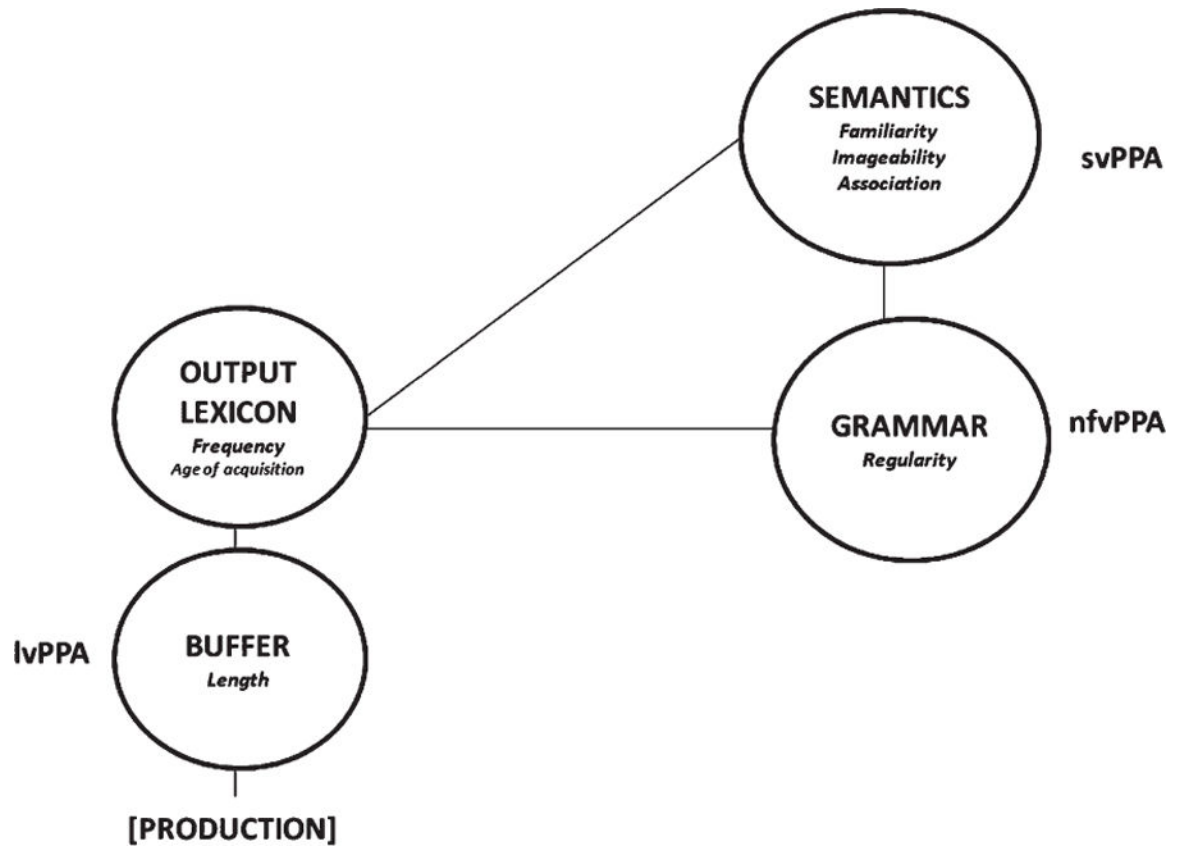
Fig. 1.
Summary of word properties that may show effects in people with PPA. Other word properties could apply.
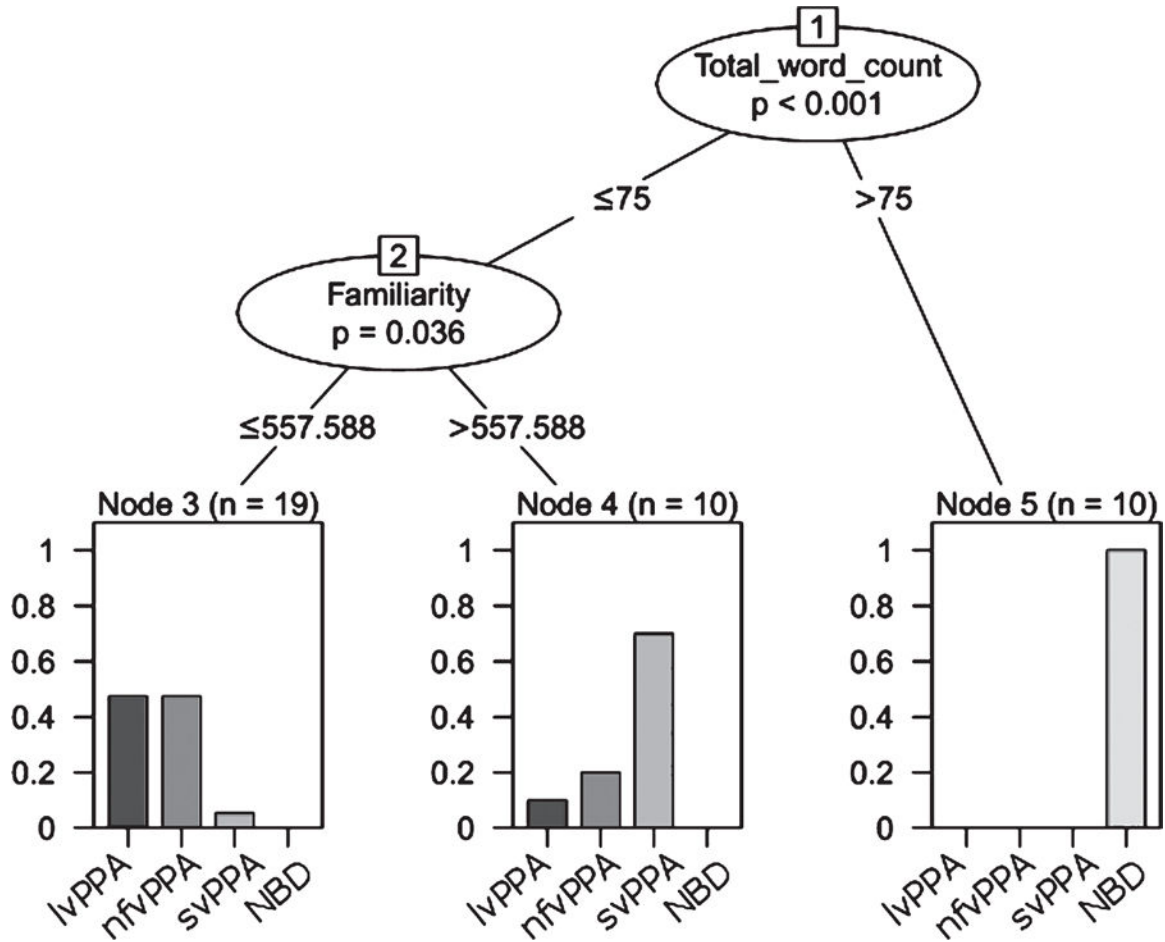
Fig. 2.
Conditional Inference Tree to classify PPA variant including interactions. The circles represent significant variables and include a p-value. Below each node, the values represent a point in each variable where individuals can be significantly separated into PPA groups. Each bar plot illustrates the distribution of PPA variant within each branch of the tree. More specifically, the figure shows a first significant split for the total word count. When participants produce more than 75 words, you go right to node 5 and that splits off all 10 non-brain-damaged individuals from the participants with PPA. If word count is less or equal than 75 words, then you go left and check familiarity, where we also found a significant split. If mean familiarity is greater than 557, then you go to node 4 and that splits off 10 more participants. The majority of these participants are people with svPPA (0.7 or 70%), then nfvPPA (20%) and lvPPA (10%). Finally, if mean familiarity is less than or equal than 557, then you go left to node 3 and that splits off the remaining 19 participants. Here we find similar percentages of people with lvPPA and nfvPPA (45%) and a much smaller percentage of people with svPPA (5%).
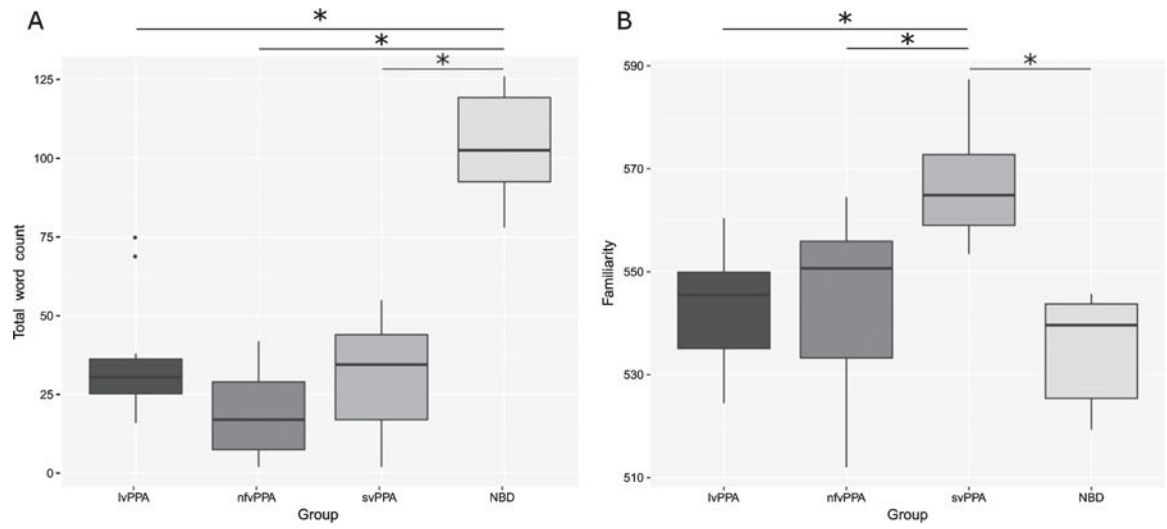
Fig. 3.
Total word count and mean familiarity of words produced for letter and category fluency.

Table 1

Demographic and test data on the participants

	lvPPA	nfvPPA	svPPA	NBD	Significance
Demographic					
Number of participants	10	8	11	10	
Age	66(9)	69(6)	69(6)	68(7)	p = 0.64
Sex (M/F)	5/5	4/4	7/4	4/6	p = 0.75
Education in years	16(2)	16(3)	17(2)	17(1)	p = 0.37
Status					
Language FTLD-CDR[1]	1.6(.9)	2(.8)	2.3(.6)		p = 0.26
Total FTLD-CDR	6.4(3.8)	6.1(3.9)	7.9(3.5)		p = 0.53
Years post diagnosis	4.7(3.7)	5.1(2.8)	4.9(2.6)		p = 0.98
Naming					
Objects (BNT, N = 30)[2]	20.8(9.1)[b]	15.4(9.2)[a]	2.6(3.2)[a,c]		p = 0.0003*
Actions (HANA, N = 35)[3]	21.7(9.4)[b]	15.5(9.6)	7.4(7.2)[c]		p = 0.01*
Word association					
Objects (PPT, N = 15)[4]	14.1(2.3)	13.7(2.7)	12.6(1.9)		p = 0.43
Actions (KD, N = 15)[5]	13.7(2)	12.6(1.9)	11.5(2.5)		p = 0.21
Repetition					
Sentence repetition (N of words)[6]	27(7.7)	21.6(11.8)[b]	33.9(2.4)[a]		p = 0.02*
Spelling to dictation[7]					
Words (% correct)	82.7(21.8)	81.6(17.5)	91.4(7.3)		p = 0.47
Pseudowords (% correct)	76.4(17.4)[a]	55.4(18.8)[b,c]	90.9(7.4)[a]		p = 0.001*
Grammar					
Sentence comprehension (SOAP, N = 40)[8]	28.4(6.1)	25.9(7.9)	30.6(7.4)		p = 0.42
Memory & other					
RAVLT (sum of 5 trials)[9]	16(11)	30(25)	17(14)		p = 0.35
RAVLT (Delayed recall)	4(3.79)	6.2(6.1)	2.3(2)		p = 0.71

	lvPPA	nfvPPA	svPPA	NBD	Significance
RAVLT (Written recognition)	2.6(5.6)	6.2(7.9)	0(6)		$p = 0.38$
Digit span forward[10]	$3.80(1.3)^b$	$3.18(1.2)^b$	$6.38(1.6)^{ac}$		$p = 0.0001*$
Digit span backward	$2.55(0.6)^b$	$1.95(1.2)^b$	$4.63(1.8)^{ac}$		$p = 0.0003*$
Spatial span forward[11]	$3.81(1.1)^a$	$2.55(1.2)^{b,c}$	$4.64(1)^a$		$p = 0.003*$
Spatial span backward	3.50(1.13)	$2.59(1.46)^b$	$4.5(1.29)^a$		$p = 0.03*$
TMT-A[12]	104(118)	56(15)	65(42)		$p = 0.47$
TMT-B	194(128)	250(88)	132(94)		$p = 0.24$

[1] FTLD-CDR, Score for the language item only and the total sum of all items in the Frontotemporal Lobal Degeneration version of the Clinical dementia Rating scale [51];

[2] BNT, Accuracy of picture naming of objects in the Boston Naming Test [53];

[3] HANA, Accuracy of picture naming of actions in the Hopkins assessment of Naming actions [54];

[4] PPT, Conceptual knowledge of objects assessed with picture association through the Pyramids and Palm Trees Test [55];

[5] KD, Conceptual knowledge of actions assessed with picture association through the Kissing and Dancing Test [56];

Number of words repeated correctly during oral repetition of sentences [57];

[7] Accuracy of spelling to dictation words and pseudowords from the Johns Hopkins University dysgraphia battery [58]

[8] SOAP, Auditory comprehension of sentences using a sentence-to-picture matching task with thematic and unrelated distractors, the Subject relatives, Object relatives, Actives, Passives test [59];

[9] Sum of words recalled in Rey auditory verbal learning test trials 1 to 5, including each trial [60];

[10] Length of digit span from the Wechsler Adult Intelligence Scale – Revised [61];

[11] Length of spatial span from the Corsi block test [62];

[12] Response times in parts A and B of the Trail Making Test [63]. Superscript letters indicate significantly impaired (or different, in the case of demographic data) relative to

[a] nfvPPA

[b] svPPA

[c] lvPPA, or

[d] NBD.

Table 2

Mean and standard deviation per task and dependent variable across groups

	Words	Frequency	Familiarity	Imag.	Concr.	SA	AoA	Length	Orth.S	Phon.S
Letter & Category										
lvPPA	36 (20)	2.8 (0.5)	544 (11)	544 (42)	4.11 (0.4)	0.42 (0.1)	5.53 (0.5)	4.62 (0.8)	4.65 (2)	10.2 (3.3)
nfvPPA	18 (14)	3.2 (1.1)	544 (18)	521 (88)	3.99 (0.7)	0.45 (0.2)	5.55 (1)	4.22 (0.7)	5.05 (2.6)	11.89 (4.7)
svPPA	30 (20)	3.9 (0.5)	567 (11)	515 (53)	3.66 (0.6)	0.34 (0.1)	5.1 (0.9)	3.97 (0.6)	6.04 (1.7)	11.59 (3.2)
NBD	104 (17)	2.4 (0.3)	535 (11)	533 (15)	3.98 (0.2)	0.33 (0.1)	6.59 (0.6)	6.33 (3)	6.43 (1)	9.68 (1.9)
Letter										
lvPPA	17 (9)	3.85 (0.9)	552 (13)	483 (52)	3.53 (0.6)	0.24 (0.1)	6.11 (0.8)	4 (0.8)	4.81 (2.1)	11.31 (3.9)
nfvPPA	8 (6)	3.95 (1.3)	547 (28)	470 (81)	3.44 (0.6)	0.21 (0.1)	6.03 (1.2)	3.8 (0.8)	5.46 (2.4)	13.43 (4.5)
svPPA	22 (15)	4.39 (0.5)	567 (12)	451 (35)	3.16 (0.2)	0.27 (0.1)	5.69 (0.7)	3.8 (0.5)	5.43 (1.67)	12.12 (3.6)
NBD	51 (10)	3.36 (0.6)	549 (9)	447 (32)	3.08 (0.2)	0.25 (0.1)	7.38 (0.8)	5.3 (0.5)	6.15 (1.5)	9.47 (2.6)
Category										
lvPPA	19 (13)	1.94 (0.4)	537 (18)	604 (11)	4.88 (0.1)	0.6 (0.1)	5.01 (0.6)	5.02 (1)	4.78 (2.8)	9.44 (3.9)
nfvPPA	10 (9)	2.18 (0.8)	544 (27)	613 (13)	4.92 (0.1)	0.7 (0.1)	4.65 (1)	4.66 (0.6)	5.31 (3.1)	9.96 (3.7)
svPPA	8 (6)	2.77 (0.9)	563 (27)	617 (15)	4.96 (0.1)	0.6 (0.2)	4.17 (0.9)	4.69 (0.9)	6.4 (3.5)	9.41 (4.4)
NBD	53(11)	1.61 (0.3)	523 (13)	597 (6)	4.86 (0.1)	0.4 (0.1)	5.81 (0.6)	7.16 (5)	6.82 (1.2)	9.94 (1.8)

lvPPA, logopenic variant PPA; nfvPPA, nonfluent variant; svPPA, semantic variant; NBD, non brain-damaged individuals. Acc, accuracy; Part, participant; Words, total number of words; Imag, imageability; Fam, familiarity; Concr, concreteness; SA, semantic association AoA, age of acquisition; Length, length in Phonemes; Orth.S, orthographic similarity; Phon.S, phonological similarity.

Table 3

Confusion matrix for letter and category fluency combined

Predicted class	Actual class			
	lvPPA	*nfvPPA*	*svPPA*	*NBD*
lvPPA	3	4	1	2
nfvPPA	4	3	4	0
svPPA	2	2	4	0
NBD	0	0	0	10

Table 4

Confusion matrix for letter fluency

Predicted class	Actual class			
	lvPPA	*nfvPPA*	*svPPA*	*NBD*
lvPPA	0	6	4	0
nfvPPA	5	6	0	0
svPPA	1	2	3	2
NBD	0	0	0	10

Table 5

Confusion matrix for category fluency

Predicted class	Actual class			
	lvPPA	**nfvPPA**	**svPPA**	**NBD**
lvPPA	2	5	1	2
nfvPPA	3	4	3	1
svPPA	3	3	2	0
NBD	2	0	0	8