The use of spelling for variant classification in primary progressive aphasia: Theoretical and practical implications

Kyriaki Neophytou[a,*], **Robert W. Wiley**[a,b], **Brenda Rapp**[a], **Kyrana Tsapkini**[a,c,1]

[a]Department of Cognitive Science, Johns Hopkins University, Baltimore, MD, USA

[b]Department of Psychology, University of North Carolina at Greensboro, Greensboro, NC, USA

[c]Department of Neurology, Johns Hopkins Medicine, Baltimore, MD, USA

Abstract

Currently, variant subtyping in primary progressive aphasia (PPA) requires an expert neurologist and extensive language and cognitive testing. Spelling impairments appear early in the development of the disorder, and the three PPA variants (non-fluent - nfvPPA; semantic - svPPA; logopenic - lvPPA) reportedly show fairly distinct spelling profiles. Given the theoretical and empirical evidence indicating that spelling may serve as a proxy for spoken language, the current study aimed to determine whether spelling performance alone, when evaluated with advanced statistical analyses, allows for accurate PPA variant classification. A spelling to dictation task (with real words and pseudowords) was administered to 33 PPA individuals: 17 lvPPA, 10 nfvPPA, 6 svPPA. Using machine learning classification algorithms, we obtained pairwise variant classification accuracies that ranged between 67 and 100%. In additional analyses that assumed no prior knowledge of each case's variant, classification accuracies ranged between 59 and 70%. To our knowledge, this is the first time that all the PPA variants, including the most challenging logopenic variant, have been classified with such high accuracy when using information from a single language task. These results underscore the rich structure of the spelling process and support the use of a spelling task in PPA variant classification.

Keywords

Primary progressive aphasia; Spelling; Variant classification; Language; Semantic variant; Non-fluent variant; Logopenic variant

---

*Corresponding author. Department of Cognitive Science, 239 Krieger Hall, 3400 N. Charles Street, Baltimore, MD, 21218, USA. kneophy1@jhu.edu (K. Neophytou).

1. Introduction

Approximately 1–9 per 100,000 people suffer from Primary Progressive Aphasia (PPA),[2], an age-related degenerative neurological syndrome, mainly characterized by a gradual deterioration of language functions (Mesulam, 1987, 1982). PPA is very heterogeneous with regards to the clinical, imaging, cognitive, and pathological profile of patients. In most classifications it is divided into three subtypes, or variants, each associated with distinct regions of brain atrophy, diverse pathologies, as well as with diverse neuropsychological profiles. These are: non-fluent variant PPA (nfvPPA), semantic variant PPA (svPPA) and logopenic variant PPA (lvPPA). Even though these subtypes are widely used by clinicians and researchers across the world, there are a number of reasons that make subtyping intrinsically challenging. Because PPA is a degenerative disorder, subtypes that are assigned to individuals are not necessarily stable as the symptoms may change with disease progression and may also be affected by disease severity. Furthermore, because the underlying disease process does not necessarily respect the boundaries of functional neural networks, individuals may suffer from multiple subtypes. Nonetheless, despite these issues, subtyping has been found to be very useful for clinicians because they often rely on these classifications for better prognosis, as variants may be associated with different survival times (e.g., Matias-Guiu et al., 2015), and for developing more targeted treatments (e.g., Graham, 2014).

Currently, PPA variant subtyping most often relies on the criteria listed in the consensus paper of Gorno-Tempini et al. (2011) [3]. Based on these consensus criteria, diagnosis requires an expert neurologist and many hours of language and cognitive testing along with imaging and pathological testing. Previous attempts to use simpler tools that reduce the number of tasks and/or use automated methods for PPA subtype classification have had important limitations. Either they did not achieve high accuracy levels for all three variants, with lvPPA being the most 'problematic' variant, or they required extensive testing to achieve accurate classification (Fraser et al., 2014; Mesulam et al., 2009, 2012; Wilson et al., 2009). Clearly, therefore, there is still a need to develop convenient and effective tools for PPA variant classification.

Previous research has provided evidence that the spelling profiles of the three PPA variants are fairly distinct from one another (for a review, see Graham, 2014). Further, previous studies have argued on theoretical and empirical grounds that spelling shares a number of its components with the spoken language system (e.g., Henry et al., 2011), supporting the notion that spelling performance could serve to index spoken language deficits. On these bases, this study investigated whether a spelling task could be used for accurate classification of PPA variants. Using advanced statistical analyses, we provide the first evidence that spelling data alone allows for high accuracy in the PPA variant classification, including lvPPA, which has been consistently reported to be the most challenging variant to classify. We also find good classification accuracies when we approximate the clinical

---

[2]Information obtained from orpha.net.
[3]Even based on the Gorno-Tempini et al. (2011)'s consensus criteria there may be a considerable percentage of unclassifiable cases, ranging from 4 to 41% for any given study or cohort (see Gil-Navarro et al., 2013; Leyton et al., 2011; Mesulam et al., 2012; Harris et al., 2013; Sajjadi et al., 2012; Wicklund et al., 2014).

scenario where the goal is to identify the variant assuming no prior knowledge of it. The results show that a single spelling task provides high classification accuracy and thus support the clinical use of a spelling task for the differential diagnosis of PPA variants.

1.1. PPA variants & their classification

Individuals with PPA are classified into the three variants based on a set of cognitive, imaging and pathological criteria that were agreed upon by an international consensus group of clinicians and researchers in 2011 (Gorno-Tempini et al., 2011; see also Mesulam and Weintraub (2014) for suggested revisions of these criteria). According to their guidelines, nfvPPA is characterized by agrammatism in language production, as well as effortful and halting speech. Single word comprehension is intact. The primary areas of atrophy are the left posterior fronto-insular regions, the insula and the premotor and supplementary motor areas. svPPA, previously known as semantic dementia, is characterized by impaired semantic knowledge leading to single-word comprehension difficulties and impaired spoken naming. Speech production and grammatical knowledge are largely intact in svPPA. The primary area of atrophy is the anterior temporal lobe bilaterally, although damage is expected to be greater in the left hemisphere. lvPPA is characterized by impaired phonological processing, manifested in impaired repetition as well as word finding difficulties. Single word comprehension and speech articulation are relatively intact. The primary area of atrophy is the left posterior perisylvian and/or parietal regions, including the posterior temporal lobe, the supramarginal and angular gyri. Importantly, previous studies argue that these distinct patterns that distinguish the three PPA variants become less distinctive as brain degeneration becomes more severe (see Rogalski et al., 2011).

One key criticism of the Gorno-Tempini et al. (2011) guidelines is that they do not indicate which specific tests need to be administered, nor do they specify cut-off scores for the relevant tests (Mesulam et al., 2012). Also, in order to be able to reliably classify an individual with PPA into a variant category, extensive testing is required, including neuropsychological, imaging and pathological assessments. Another criticism is that the guidelines lack specificity with regards to classifying lvPPA (Harris et al., 2013). Many of the criteria used to identify lvPPA describe aspects of language performance that are spared in these individuals while the criteria identifying nfvPPA and svPPA mostly refer to impairments in specific language areas. This can create ambiguity in distinguishing lvPPA from the other two variants, leading to mixed PPA cases (for discussion see Harris et al., 2013; Mesulam et al., 2012).

Mesulam et al. (2009) tried to address the need for a short and objective procedure for classifying PPA individuals into the three variants, by proposing a quantitatively-based tool for PPA variant classification. Mesulam and colleagues provided a 2-dimensional template based on two tasks: a grammatical production task and a single word comprehension task. To validate this approach, they first classified 16 PPA individuals into the three variants using five language tests (auditory single word comprehension, grammaticality of syntax production, naming, fluency and repetition). They then created an 'orthogonal mapping' of the patients' performance on the two tasks of interest, by plotting one against the other. Based on the plot, they found that the three variants created three distinct clusters: the lvPPA

group scored above 60% on both tasks, the nfvPPA group scored above 60% on the single word comprehension task and below 60% on the syntax production task, and the svPPA group showed the reverse pattern from that of the nfvPPA group, scoring below 60% on the single word comprehension task and above 60% on the syntax production task. In a follow-up study (Mesulam et al., 2012), Mesulam and colleagues used the same two-task classification approach on a different set of 21 PPA individuals with an accuracy threshold of 80% (instead of 60%). They also allowed for a 'grey zone' area when accuracy on the two tasks was between 80 and 90%. The reason for using a different threshold than in the earlier study was because the sample of participants was less impaired compared to those in their previous study. However, seven individuals, one third of the total participants, fell into the 'grey zone': 4 of the 6 lvPPA, 2 of the 9 nfvPPA and 1 of the 4 svPPA individuals. When solely relying on the two tasks, the accuracy scores for the three groups were 78% for the nfvPPA, 75% for the svPPA, and 33% (i.e., at chance) for the lvPPA. The particularly low classification accuracy for the lvPPA patients is illustrative of the difficulties that have been encountered in classifying this variant. In order to better classify these individuals into one of the three variants, the researchers considered their performance on other tasks, namely repetition and speech production, indicating that the two-task classification was not sufficient.

Given the number of advanced automated tools for classification, some researchers have tried to use machine learning algorithms for PPA variant classification. Fraser et al. (2014) and Fraser et al. (2013) used syntactic and semantic features extracted from narrative speech to distinguish between nfvPPA and svPPA. The authors tested several different machine learning algorithms, and their classification accuracies for the pairwise comparison of nfvPPA with svPPA (the only pair of variants they considered) ranged from 71 to 79%. Wilson et al. (2009) also utilized machine learning algorithms using pairwise classification to categorize PPA variants, but instead of language data, they used MRI grey matter images to compare the atrophy patterns for each pair of variants. The accuracies for variant classification were 81.3% for the lvPPA *vs* nfvPPA comparison, 89.1% for the nfvPPA *vs* svPPA comparison, and 93.8% for the lvPPA *vs* svPPA comparison. For the nfvPPA vs svPPA comparison, if various linguistic variables were also included, accuracy reached 96.2%.

The results of the above studies suggest that advanced statistical tools may provide an important approach for PPA variant classification. However, further research is needed to identify a set of performance features that would allow for high classification accuracies for all three variants that would not require neuroimaging measures, because imaging is expensive, for some patients it may contraindicated and it may be not available to all clinicians, especially in remote or underdeveloped settings. The current study addresses this need by using a single and easily administered task: spelling. In the discussion that follows, we describe the cognitive architecture of spelling and elaborate on the theoretical and empirical bases of our decision to consider spelling as a classification task.

1.2. The cognitive architecture of spelling

Studies of acquired dysgraphia (primarily subsequent to stroke) have formed the basis for the development of the cognitive architecture of spelling depicted in Fig. 1 (see Caramazza and Miceli, 1990; Rapp et al., 2015; Tainturier and Rapp, 2001). The spelling system schematized here is a 'dual route' system in which the input for the task of spelling to dictation is acoustic, while the output is orthographic. In an intact spelling system, if the input is a familiar real word, such as 'window', it can be successfully processed by the *Lexical Processing Route*. Namely, after acoustic and phonological processing, the lexical phonological representation is activated in phonological long-term memory, providing access to the lexical-semantic representation which, in turn forms the basis for the retrieval of the lexical orthographic representation from orthographic long-term memory. The graphemic buffer, also referred to as *orthographic working memory* (Buchwald and Rapp, 2004), is responsible for maintaining the activation levels of the letter identities during the time needed to produce each letter in the target word. Unfamiliar words or pseudowords, such as 'foit' do not have phonological, semantic or orthographic representations to be retrieved from long-term memory and, for these, the *Sublexical Processing Route* generates a plausible spelling from the acoustic input. Specifically, the unfamiliar phonological string is held in *phonological working memory* while the phonemes are mapped onto single or multiple-letter orthographic units, the graphemes, via *phonology-to-orthography conversion* (POC) processes that rely on information learned about the systematic relationships between sounds and letters. The sequence of graphemes that is generated in this way is processed by orthographic working memory, as for real words. The letter representations generated by both lexical and sublexical processes correspond to abstract letter representations (lacking form or sound) that are then assigned specific formats depending on the modality of output, e.g., letter shapes for written spelling and letter names for oral spelling. Note that, in the case of real words, there is only one correct output, while in the case of pseudowords (in languages such as English) multiple outputs may be plausible. For instance, both FAM and PHAM would count as correct spellings of the pseudoword/fam/, but only GRAPH (and not GRAFF) is considered to be a correct spelling for the real word/graef/.

Selective impairments affecting each component of the spelling system give rise to characteristic patterns of performance which, in the context of PPA, may prove to be useful in distinguishing between PPA variants. Thus, impairment affecting the lexical route will result in an effect of *lexicality*, with worse performance on real words than pseudowords. On the other hand, selective impairment in the sublexical route will also result in an effect of lexicality but in the opposite direction, with worse performance on pseudowords compared to real words (e.g., Bub and Kertesz, 1982; Goodman-Schulman and Caramazza, 1987; Shallice, 1981).

If the lexical route is affected, the spelling of real words can be also affected by lexical variables, such as *frequency*, and semantic variables, such as *imageability*. In the case of a disruption to any of the sub-components of the lexical route, higher frequency words are spelled more accurately than lower frequency words, as their representations are more robust, and therefore more resilient to damage (Oldfield and Wingfield, 1965; Rapp and Fischer-Baum, 2015; Wilson et al., 2010). In the case of a semantic impairment,

imageability may affect performance in either a positive or negative manner, presumably depending on specific aspects of the brain damage. Thus, there are cases of semantic impairment where words corresponding to more imageable concepts are relatively better preserved than less imageable, ones and other cases in which the opposite is observed, with the latter referred to as the 'reverse concreteness effect' (Breedin et al., 1994; Crutch and Warrington, 2005; Warrington, 1981). We discuss this in more detail in the General Discussion. In the case of a damaged lexical route, if the sublexical system is intact it can generate plausible spellings for the target, producing an effect of *PG conversion probability*. More specifically, when spelling relies on the POC process, PG mappings that are more common (and therefore more probable) in the language (e.g., *f* for /f/) may be produced more often than mappings that are less probable (e.g., *ph* for/f/), resulting in higher accuracy for (words with) higher PG conversion probabilities and in phonologically plausible errors (PPEs) for words with lower PG conversion probabilities (e.g., 'yacht' spelled as YOT).

In cases of selective impairment of the graphemic buffer, performance is usually affected by the *length* of the word, in terms of the number of letters in the word. Since the graphemic buffer has limited capacity to hold letter representation, letters in longer words are more susceptible to error than letters in shorter words (Buchwald and Rapp, 2004; Schiller et al., 2001; Tainturier and Rapp, 2003). The performance on the individual letters of a word is also differentially affected based on their position in the word. For some individuals with disruption affecting the graphemic buffer spelling accuracy decreases with letter position (e.g., Schiller et al., 2001). This is referred to as the *linear position* effect, with position being defined starting from the left of a string. Other individuals have lower spelling accuracy in the middle positions of words (e.g., Buchwald and Rapp, 2004; Friedmann and Gvion, 2001; Jones et al., 2009; Wing and Baddeley, 1980). This *quadratic position* effect creates a bow-shaped accuracy function across letter positions. The two patterns may also be observed in combination.

Two further variables that may be relevant to characterizing spelling performance are *orthographic* and *phonological neighborhood density* (i.e., the number of words that are orthographically or phonologically very similar to a given string). For instance, *wreath* (/riθ/) and *breath* (/brɛθ/) are orthographic neighbors because they only differ with respect to one letter (w- *vs* b-), while *wreath* (/riθ/) and *teeth* (/tiθ/) are phonological neighbors because they only differ with respect to one phoneme (/r/ *vs* /t/). There have been several lines of evidence indicating that lexical neighbors of a target word are active during retrieval from orthographic long-term memory (Folk et al., 2002; Goldrick et al., 2010; McCloskey et al., 2006; Roux and Bonin, 2009) providing the basis for neighborhood density to influence spelling accuracy and error types. Goldrick et al. (2010) and Sage and Ellis (2004) showed better performance in written production for target words in high-density compared to low-density neighborhoods. Furthermore, Tainturier (2013) demonstrated the influence of phonological lexical neighbors on pseudoword spelling.

The patterns of spelling performance reviewed in this section have been primarily identified in individuals suffering from stroke-induced dysgraphia. However, in the context of the current investigation, we will consider the patterns of spelling performance associated with

the three variants of PPA, and how these map onto the different components of the spelling architecture.

1.3. The spelling profile of the three PPA variants

Various attempts have been made to describe the spelling profile/s of individuals with PPA. Graham (2014) reviewed several studies investigating spelling impairments in the three PPA variants (e.g., Faria et al., 2013; Sepelyak et al., 2011; Shim et al., 2012) and provided a description of the distinct spelling impairments of each of the three PPA variants, both with respect to the variables affecting their spelling performance, as well as the errors they make.

Despite considerable variability in deficits and performance even within the same variant, evidence indicates a general pattern associated with each variant. Of the three variants, svPPA seems to be most clearly associated with a specific spelling pattern. A characteristic feature of svPPA is the clinical syndrome of "surface dysgraphia", which is characterized by difficulty in spelling real words that have low PG conversion probabilities (i.e., words with irregular orthography, such as 'yacht' - henceforth, *irregular words*). Their spelling is also characterized by a high prevalence of PPEs, signaling an intact sublexical system. Surface dysgraphia is so prominent in svPPA, that is considered to be one of the main diagnostic criteria for this variant (Gorno-Tempini et al., 2011). In contrast, the spelling performance of individuals with nfvPPA is characterized by worse performance on pseudowords than on words. However, additionally, like svPPA, for words, performance is worse with irregular compared to regular words. Also, in contrast to the svPPA, in nfvPPA spelling errors (for both words and pseudowords) are mostly phonologically implausible strings, known as non-phonologically plausible errors (henceforth, nonPPEs, e.g., the pseudoword 'donsept' spelled out as DONSIT), while PPEs, although also reported, are less prevalent. The combination of impaired performance on pseudowords and the high frequency of nonPPEs for pseudoword targets is consistent with the clinical category of phonological dysgraphia. The third variant, lvPPA, has been the most challenging to accurately describe. As Graham (2014) noted, the spelling profile of these individuals is very similar to nfvPPA, in that they exhibit worse performance with pseudowords than words, producing a mixture of PPEs and nonPPEs. However, the next most common language impairment reported in lvPPA is surface dysgraphia, which is the characteristic feature of svPPA. In other words, the spelling profile of lvPPA seems to be an amalgamation of the profiles of svPPA and nfvPPA.

Overall, the spelling profiles of the svPPA and the nfvPPA have been shown to be somewhat distinct and relatively consistent across studies, with the former having a primary impairment affecting the lexical route and the latter the sublexical route. On the other hand, the spelling profile of the lvPPA group is not as distinct and has not been easy to identify. Given some of these differences, Shim et al. (2012) suggested that spelling might be an important source of evidence regarding PPA variant classification. However, to the best of our knowledge, their proposal of using spelling performance as a tool for variant classification has not been tested.

1.4. Spelling as a proxy for spoken language

There are two (non-mutually exclusive) accounts for why spelling may serve as a useful proxy for the spoken language system: a) spelling and spoken language share language and other cognitive processes, and b) written and spoken language processes are distinct and independent but they are instantiated in nearby cortex and, therefore, both will tend to be affected by atrophy in a given region.

In most cognitive theories, both spelling and spoken language comprehension and production are assumed to share semantic, and phonological components - corresponding to the phonological longterm memory, working memory and semantic system depicted in Fig. 1 (Henry et al., 2011; Rapp and Lipka, 2011). Furthermore, evidence from studies of PPA support the assumption of additional shared processes. For instance, Shim et al. (2012), who studied the relationship between spelling and other language measures, reported a positive correlation between the spelling of pseudowords and syntax production and repetition, and also a positive correlation between the spelling of irregular words and single word comprehension and naming. They suggested that the association in performance between pseudoword spelling, repetition and syntax production occurs because all three employ rule-based processes that may be controlled by the same underlying cognitive or neural mechanism.

Given the theoretical and empirical evidence that spelling shares at least some of the same cognitive mechanisms as other spoken language systems, it provides the basis for using spelling as a proxy for spoken language processing in PPA variant prediction.

1.5. The current study

The current study aimed to determine whether spelling performance, evaluated with advanced statistical analyses and automated classification tools, can be a useful tool in PPA variant classification. If successful, this research would constitute a "proof of concept" and a significant first step towards the future development of a simpler and more clinically appropriate spelling-based classification tool.

To this end, we implemented a four-step analysis approach. First, for each individual with PPA we measured the relationship between a set of lexical, sublexical and semantic variables and their spelling performance. Second, in order for the variant classification to be as efficient as possible and avoid overfitting, we carried out variable selection by identifying which variables were most informative for distinguishing the three variants. Third, using the variables identified in the previous step, we attempted to predict the variant of every participant in pairwise comparisons (lvPPA *vs* nfvPPA, lvPPA *vs* svPPA, and nfvPPA *vs* svPPA). Finally, we extended these results in a 'real-life' scenario where we assumed no prior knowledge of each individual's variant, thus simulating a more clinically realistic scenario where we could assess how many cases could be correctly classified, misclassified or not classified on the basis of the spelling data.

2. Materials & methods

2.1. Participants

The current study included data from 42 individuals with PPA who participated at a clinical trial on the effects of transcranial direct current stimulation in PPA (ClinicalTrials.gov Identifier: NCT02606422). To determine the "ground truth" variant classification to be used in evaluating the accuracy of the spelling-based classification, the PPA variants of these individuals were identified based on the Gorno-Tempini et al. (2011) consensus guidelines. Because these participants were recruited over a period of five years the available tasks for each of them slightly varied. We retrieved all the information that was available for each individual and used data from a time point closest to the point in time that the spelling data were collected. All data used in this study were collected prior to any treatment. The Gorno-Tempini et al. classification was based on the following features: grammaticality of sentence production, effortful speech, word finding difficulty, single word comprehension, naming, repetition, syntax comprehension, object knowledge, surface dyslexia/dysgraphia, and phonemic speech errors. The specific tasks and performance criteria used to quantify each of the Gorno-Tempini et al. features, as well as each participant's performance relative to each of the Gorno-Tempini et al. classification criteria, are presented in Appendix A.

Based on these criteria, the 42 PPA individuals were subtyped as follows: 17 lvPPA, 10 nfvPPA, 6 svPPA, 4 were characterized as mixed cases and 5 as unclassified cases. The mixed cases (henceforth *M*) were individuals whose behavior fit two variant profiles, while unclassified cases (henceforth *UC*) were cases that did not fully fit a single variant profile, meeting some but not all the criteria for one or more variants. In the Gorno-Tempini et al. (2011) paper, M and UC cases are together referred to as 'PPA unclassifiable' cases, whose syndromes might become clearer at later stages of the disease, if they were at the very early stages, or alternatively, their syndromes had become indistinguishable because they were in later, more severe stages[4]. The percentage of classified cases in the current dataset (i.e., excluding M and UC cases) was at 79%. Similar to other classification studies that excluded M and/or UC cases (e.g., Mesulam et al., 2012; Wilson et al., 2009), all the analyses discussed below only included the 33 classified cases.

2.2. Data collection and scoring

Participants were recruited on a continuous basis over five years and their spelling performance was assessed using a spelling to dictation task that included real words and pseudowords, ranging from 73 to 134 and 19–34 items, respectively.

Spelling accuracy was evaluated at the letter level, rather than the word level. Briefly, scoring involved assigning - for each letter in a target word - half credit for producing the correct letter identity and half credit for the correct position. For a detailed description of the scoring procedure see Caramazza and Miceli (1990) and Tainturier and Rapp (2003).

---

[4]The overall severity scores of the M and UC cases fell on either of the two ends of the overall severity spectrum as measured by FTD Clinical Dementia Rating (FTD-CDR 'sum of Boxes', see Knopman et al., 2008): two individuals scored between 18.5 and 19, six individuals scored between 1 and 4, and one individual scored 8. Based on these scores, one could suggest that scores equal or below 4 and scores equal or above 18 could be potentially used as cutoff points prior to classification. However, further research is required to empirically test this.

Error types were also recorded. Besides null responses (i.e., no response), errors were categorized into seven types: five error types for real words and two for pseudowords. For the real words: PPEs (e.g., 'fame' spelled as PHAME), lexical substitutions (e.g., 'knob' spelled as KNOCK), morphological substitutions (e.g., 'fight' spelled as FIGHTING), semantic substitutions (e.g., 'tiger' spelled as LION) and pseudowords (e.g., 'member' spelled as MOMER). For the pseudowords: the errors were some other pseudoword (e.g., 'donsept' spelled as DOMSIT), or lexicalizations (e.g., 'donsept' spelled as CONCEPT). Error type proportions were calculated out of the total number of stimuli in the category, for words or pseudowords separately. For instance, if an individual was administered 80 real words and made 10 PPEs, 4 lexical substitutions and 6 pseudoword for word errors, the values used for the analysis were: PPEs = 0.13 (i.e., 10/80); lexical substitutions = 0.05 (i.e., 4/80); pseudoword for word = 0.08 (i.e., 6/80).

2.3. Data analysis

The data analysis consisted of four main steps: (1) for each individual, quantify the contribution of each of nine lexical, semantic and sublexical variables (see Table 1) to their spelling performance using Linear Mixed-Effects Modeling, (2) to reduce the number of variables to be used in subsequent analyses, we evaluate which of 14 variables (the nine variables discussed in Step 1, plus the accuracy difference in performance between words and pseudowords and the percentages of four error types) differed across the three variants, using a one-way Analysis of Variance per variable, (3) classify every individual in two pairwise comparisons relevant to their variant[5] (e.g., for an lvPPA individual: lvPPA *vs* nfvPPA and lvPPA *vs* svPPA, but not nfvPPA *vs* svPPA) using a leave-one-out cross-validation process by training binomial models with the values of the variables selected in step 2 and then assessing the statistical significance of the classification accuracy, and, (4) extend these results to a 'real-life' scenario, assuming no prior knowledge of the variant in each case by predicting a label for every individual for *all three* pairwise comparisons. This allows us to identify correctly classified, misclassified, as well as unclassified cases.

2.3.1. Step 1: individual variable weightings - linear mixed-effects models—
The first step was to quantify the extent to which each of 9 lexical, semantic and sublexical variables discussed in section 1.2. affected the spelling performance of each individual. To achieve this, we ran linear mixed-effects models (LMEM) (Baayen et al., 2008) utilizing the lmer function of the 'lme4' package in R (Bates and Maechler, 2009), to get a coefficient for each variable for each participant[6]. The variables (Table 1) included in this analysis were selected based on their relevance to the different components of the spelling system presented above. For the current analysis, we ran two models for each participant: Model 1 analyzed spelling responses to both real words and pseudowords, while Model 2 analyzed only spelling responses to real words. Since Model 1 was designed to evaluate the effect of predictors that are relevant to both real words and pseudowords, coefficients for this model

---

[5]Initially, we performed a single, 3-way analysis with all the three variants together, rather than creating three pairwise comparisons. However, the variant classification accuracies with the 3-way analysis were relatively low compared to the pairwise comparisons and, for this reason, we adopted a pairwise classification approach. Similarly, in their MRI-based PPA variant classification study, Wilson et al. (2009) also report variant classification accuracies from pairwise comparisons.
[6]Interested readers are referred to Wiley and Rapp (2018) for a more in-depth explanation of the application and effectiveness of the LMEM analysis.

were estimated with fixed effects for: lexicality, length, position and position-quadratic. Random effects were: by-item (i.e., for each word or pseudoword) slopes for position and position-quadratic as well by-item intercepts. Model 2 was designed to evaluate the effect of predictors that are specific to real words only, with fixed effects coefficients estimated for: imageability, frequency, PG conversion probability, orthographic and phonological neighborhood density. Random effects were: by-item intercepts. For both models, individual letter accuracy was used as the dependent variable. As previously indicated, individual letter accuracy was evaluated for all the letters of all the stimuli (with letter scores ranging from 0 to 1) and logistic regression (binomial family) was used to model these accuracies (as bounded between 0 and 1). The goal of this analysis was to obtain a beta coefficient value for each variable, for every individual.

2.3.2.   Step 2: variable reduction – one-way ANOVAs—The second step was to reduce the number of variables to be used for variant classification, by identifying the variables on which the variants most clearly differed. To this end, a series of one-way ANOVAs, with the single between-groups factor of variant as a predictor, were used to determine which variables had mean values that differed, using a threshold[8] of $p < .05$. For this analysis we used the 'stats' package in R (Bates and Maechler, 2009).

We evaluated fourteen variables: the nine lexical, semantic and sublexical variables presented in Table 1, whose relevance was discussed earlier with respect to the cognitive architecture of spelling. We also included the *word-pseudoword grapheme accuracy* variable. This variable was calculated by subtracting the percentage of correct graphemes in pseudowords from the percentage of correct graphemes in real words. The goal of this variable was to capture the difference in impairment between the lexical and the sublexical routes. Although it might seem very similar to the *lexicality* variable, the two variables are conceptually different. The *lexicality* variable quantifies the effect of the lexicality of the input, whether it is a real word or a pseudoword, while the *word-pseudoword grapheme accuracy* evaluates the relative intactness of the two routes, the lexical and the sublexical subsystems, regardless of total accuracy. The final four variables corresponded to error type percentages. These were included because of the potential utility of error type in distinguishing between deficit loci and, therefore, PPA variants. Although there were seven different types of errors (see section 2.2.), for this analysis we only included four. For the real words: PPEs, lexical substitutions and pseudowords. For the pseudowords, the only error type included was other pseudoword error responses. Semantic substitution errors and lexicalizations were excluded because they were extremely rare. Morphological substitution errors were also very limited and were therefore included with lexical substitutions.

2.3.3.   Step 3: variant classification – binomial cross classification—The third step of the analysis was to investigate the extent to which the variables identified in Step 2 allow for accurate PPA variant classification. To this end, we evaluated three distinct models, one for each of the three pairs of variants. The predictors in these models were the

---

[8]Note that because the goal is simply variable reduction the threshold can be arbitrary and is not subject to concerns about multiple comparisons. These concerns would be relevant only if claims were to be made about the significance of the variables identified at this analysis step, something which we are not doing.

variables identified in Step 2. For each pairwise comparison, we followed the same process: for variant classification we used binomial model analysis, and since variant classification requires predicting the variant using a training data set and a test dataset, we used the 'leave-one-out' cross validation procedure to split the data into test and training datasets. This means that, for each pairwise comparison, we trained a binomial log-linear model on the data of all the individuals in the two variants except for one (i.e., the left-out participant). Then, the model was tested by predicting the variant of the left-out participant. This 'leave-one-out' procedure was repeated as many times as needed to predict the variant of each of the participants for each pairwise comparison (i.e., 27 times for the lvPPA *vs* nfvPPA comparison, 23 times for the lvPPA *vs* svPPA comparison, and 16 times for the nfvPPA *vs* svPPA comparison). The predictions (classifications) were compared to the 'ground truth' variant of each individual, as defined following the Gorno-Tempini et al. (2011) criteria, to calculate the overall classification accuracy for each model, for each variant. For the binomial model fitting and the classification, we used the 'multinom' and 'predict' functions from the *nnet* package (Venables and Ripley, 2002) in R (R Core Team, 2013).

We then evaluated the statistical significance of the classification accuracies using a Monte-Carlo permutation test with 1000 permutations. In each permutation, the variant type labels for each individual were randomly assigned and, following the just described 'leave-one-out' cross validation procedure, classification accuracy values were obtained.

2.3.4.   Step 4: classification in a 'real-life' scenario—The fourth and final step was to evaluate classification accuracy assuming no prior knowledge about the variant of each individual. This better approximates a 'real-life' scenario, in which a patient walks in into a clinic and the therapist needs to determine their variant[9]. To this end, we expanded on the results of Step 3, by also predicting a label for each individual for the pairwise comparison of 'no interest'. For instance, for an individual categorized by the Gorno-Tempini et al. criteria as svPPA, the pairwise comparison of 'no interest' would be the lvPPA *vs* nfvPPA comparison, because we know that whatever the predicted label would be, it would be incorrect. The models we used to get a label for the comparison of 'no interest' were trained on *all* the data from the two non-target variants. For instance, to predict the label for an svPPA individual, the model would have been trained on the data from all the lvPPA and nfvPPA individuals, without leaving anyone out. For the binomial model fitting and the classification, the same software packages were used as in Step 3.

This procedure allowed us to have three variant labels per participant, one for each pairwise comparison. However, given that one of the comparisons was of 'no interest', only two out of the three predicted labels could be correct. Therefore, an *accurately classified case* required 2-out-of-3 predicted labels to be correct. Except for accurately classified cases, this approach also allows us to evaluate *misclassified* cases, as well as *unclassified* cases. Misclassified cases are cases for which 2-out-of-3 labels are the same, but they do not correspond to the target label. For instance, if an svPPA individual is classified as lvPPA in the lvPPA *vs* svPPA comparison and it is also classified as lvPPA in the comparison

---

[9]It is important to make clear that this analysis included the same set of individuals as the previous analysis, namely only the 33 classified cases from our original dataset.

of 'no interest' (lvPPA *vs* nfvPPA), then there are 2-out-of-3 labels that match, which means that this person would be classified as lvPPA. However, this classification would be incorrect, and this would be considered a misclassified case. Finally, the unclassified cases are the ones for which each of the three pairwise comparisons predicted a different label. In that case, the individual could not be classified, and this would be considered to be an unclassified case. The percentage of unclassified cases would index the ability of this approach to distinguish PPA variants based on their spelling performance, assuming that the correct variant classification for each individual is indeed the one assigned by the Gorno-Tempini et al. (2011) criteria. A high percentage of unclassified cases would indicate that the spelling profiles of these variants were not distinct enough to form different categories, while a low percentage of unclassified cases would indicate distinctive spelling profiles. With this approach we were able to measure how many of the cases were correctly classified, misclassified or unclassified. This more fine-grained evaluation of the predicted classifications allows us to better assess how confident we can be that a given classification is correct in a 'real-life' situation, which is an important goal of this type of research.

3.   Results

3.1.   Step 1: individual variable weightings

For each individual we obtained coefficients for each of the nine variables in Table 1, using LMEM models. Results are presented in Table 2. The pattern of results across the three variants will be discussed in the General Discussion.

3.2.   Step 2: variable reduction

As discussed in section 2.3.2, there was an initial set of 14 variables: the nine lexical, semantic and sublexical variables presented in Table 1, *word-pseudoword grapheme accuracy*, and the percentages of each of four error types.

Similar to Table 2, which reports of the coefficient values for the nine lexical, semantic and sublexical variables, Table 3 presents the summary of the values for the *word-pseudoword grapheme accuracy* variable, as well as the percentages of error types.

Table 4 reports the variables that were selected as the most important predictors for distinguishing the three variants, as described in Section 2.3.2.

3.3.   Step 3: variant classification

Three binomial analyses were conducted, one for each pairwise comparison of the three variants. The predictors consisted of the set of variables extracted from the ANOVAs (Table 4). In Table 5 we report the accuracies of each model both overall, as well as for each variant separately.

The Monte-Carlo permutation testing showed that classification accuracies (for all three pairwise comparisons) that were equal or better than those actually obtained were rarely observed in analyses based on 1000 randomly assigned categories: 1 out of 1000 times for the *nfvPPA vs svPPA comparison (p = .001)*, 39 out of 1000 times for the *lvPPA vs svPPA comparison (p = .039)*, and 10 out of 1000 times for the *lvPPA vs nfvPPA comparison (p*

= .01). Fig. 2 presents the distribution of classification values obtained from data sets with randomly scrambled variant labels.

3.4. Step 4: variant classification in a 'real-life' scenario

The fourth and final step was to investigate the accuracy of variant classification assuming no prior knowledge about the variant of each individual. To this end, we expanded on the results of Step 3 by also predicting a label for each individual for the pairwise comparison of 'no interest'. In other words, in Step 4, we obtained a predicted label for each of the *three* pairwise comparisons per individual, instead of just two per individual (see Appendix B). As discussed before, this allowed us to approximate a 'real-life' scenario of PPA variant classification which can lead to three distinct possible outcomes: (a) accurately classified cases, (b) misclassified cases, and (c) unclassified cases[10]. Table 6 presents the confusion matrix with the classification accuracy obtained from the pairwise classifications, which derived from the three predicted labels per individual evaluated against the "ground truth" classifications based on the Gorno-Tempini et al. (2011) classification. Overall, across the three variants, 64% of the cases were accurately classified (white cells), 30% of the cases were misclassified (light-grey cells), and 6% of the cases remained unclassified (dark-grey cells).

The Monte-Carlo permutation testing showed that the overall classification accuracy of 64% or better across the three variants was only observed 1 time in 1000 randomly permuted data sets (p = .001) (see Fig. 3).

4. Discussion

This study investigated the accuracy of PPA variant classification with data from a single spelling task. Given that spelling recruits many different language and cognitive components and that the three PPA variants have been shown to have distinct spelling profiles affecting different components of the spelling process, we tested the hypothesis that spelling performance is able to distinguish between the three variants. Indeed, the results of this study show that applying advanced statistical methods to spelling performance and using automated classification tools allows classification of PPA variants with relatively high pairwise classification accuracies, ranging from 67 to 100% (Table 5). This is the first time, at least to our knowledge, that using pairwise classification, every PPA variant is classified with such high accuracy when only using information from a single language task, including lvPPA, which is considered to be the most challenging variant to classify.

In the last part of our analysis, which allowed us to approximate a more realistic clinical scenario in which variant pairwise classification occurs without any prior knowledge of the variant[11], the classification accuracies for the three variants were 70%, 66% and 59%

---

[10]Unclassified cases in this analysis, as described in section 2.3.4., refer to individuals who could be classified based on the Gorno-Tempini et al. (2011) criteria (i.e., they belong in the main analysis group of 33 classified individuals), but when only their spelling data are taken into account, they could not be classified. These cases are not related to the UC cases discussed in section 2.1.

[11]Note that this same assumption underlies 3-way classification approaches, which also does not rely on prior knowledge of a participant's variant. However, as we noted earlier, we and others have found that the attempts of using 3-way classification of PPA variants has not allowed for the same levels of accuracy.

for nfvPPA, svPPA and lvPPA, respectively (Table 6). This analysis allowed us to divide classification outcomes into three categories: accurately classified cases, misclassified cases and unclassified cases. By accurately classifying two-thirds (64%) of the cases, the results of this analysis reveal the considerable differentiation of the spelling profiles of the three PPA variants and provide evidence that the rich structure of spelling can capture the different components of the spoken language system. This research is an important step towards developing a simpler and more clinically appropriate classification tool that is both theoretically and empirically motivated. The process described above could be automated providing a simple and quick variant classification tool for clinicians.

4.1.   The spelling profiles of the three PPA variants

In the discussion that follows, we assess how the distribution of values of the various variables (Tables 2-3) provide a characterization of the spelling profiles of the three PPA variants.

As discussed in the Introduction, consistent with the previous literature, the nfvPPA individuals showed clear impairment in the sublexical route. This is manifested (as seen in Table 2) in the nfvPPA group's higher grapheme accuracy for real words compared to pseudowords (a large and positive beta value for the *lexicality* variable), they had the highest proportion of pseudoword for pseudoword errors of the three groups and the lowest rate of phonologically plausible errors, and, finally, the *word-pseudoword grapheme accuracy* difference was an important variable in distinguishing nfvPPA from the other two groups (Table 3). With regard to imageability, the nfvPPA group had the strongest positive relationship between imageability and spelling accuracy, with better spelling for high imageability words, serving as additional evidence for an impairment to the lexical route, although, the specific nature and locus of this impairment requires further investigation. Finally, with regards to *length* and *position*, the beta values reported in Table 2, indicate that nfvPPA had worse performance on longer words compared to shorter words and also on the final letters of a word compared to the initial letters. This could be driven by either an impairment in phonological or orthographic working memory, or both (see Fig. 1). Further research, with greater numbers of participants, should investigate these possible interpretations of the length and position effects.

The svPPA individuals showed less impairment in the lexical route with relatively intact sublexical processing. Support for this comes from the fact that they showed the reverse pattern of performance than the nfvPPA group with higher grapheme accuracy for pseudowords compared to real words (a negative *word-pseudoword grapheme accuracy* difference value; Table 3). Similarly, the beta values for *lexicality* (Table 2), also indicated better performance for pseudowords compared to real words. Consistent with this, the svPPA group made the most PPEs and the fewest nonPPEs. As we have seen, *imageability* was another lexical variable that was found to be significant in differentiating the three groups. The reason imageability was so useful in classification was because it had a negative effect for svPPA while it had a positive effect for lvPPA and nfvPPA. In other words, svPPA individuals had worse performance on more concrete compared to more abstract words, while nfvPPA and lvPPA individuals showed the opposite pattern. These results suggest

that while the latter two groups benefited from the perceptual feature salience of concrete words, the svPPA exhibited 'the reversal of concreteness effect' (Breedin et al., 1994). This effect has been reported both in healthy individuals in certain task contexts (Romani et al., 2008; Tyler et al., 2002) as well as in post-stroke aphasia (Crutch and Warrington, 2005, 2003; Warrington, 1981) and semantic dementia (Bonner et al., 2009; Breedin et al., 1994; Warrington, 1975; Yi et al., 2007). Although the reversal effect has not always been reported in this population (see Hoffman and Lambon Ralph, 2011), the current findings do constitute further support for this feature in svPPA. More importantly, these results stress the importance of imageability as a predictor in PPA variant classification. Fraser et al. (2013) had also identified imageability as a significant predictor for distinguishing svPPA from nfvPPA, but the direction of the effect for the two groups was not reported. With regard to length, unlike the other two groups, the svPPA group exhibited a "reverse" length effect, such that performance was better on longer compared to shorter words. Whether this effect is statistically significant and whether it could be due to confounding co-variates, is an issue that will require further study.

Finally, the lvPPA group, like the other two variants was also shown to have impairments that affected both the lexical and the sublexical routes. In contrast to the other two variants, however, the two routes seem quite comparably affected, as indicated by the close to zero value for the word/pseudoword grapheme accuracy variable (Table 3) and beta value for the lexicality variable (Table 2). Also, compared to the other two variants, lvPPA had intermediate rates of PPEs and pseudoword for pseudoword errors. Overall, the lvPPA group was more similar to the nfvPPA group, in terms of impairment to the sublexical route as well as the direction of the length, position and imageability effects. Nonetheless, lvPPA and nfvPPA could be distinguished based on the lvPPA's more modest differences between word and pseudoword accuracy and the milder imageability effects. The dual (lexical/sublexical) but milder impairments of the lvPPA individuals allows them to be distinguished from the other two variants.

4.2. The 'challenging' logopenic variant

As discussed above, lvPPA has been the most challenging variant to distinguish from the other two variants. The few previous attempts that were made to develop an automated tool for variant classification either involved neuroimaging data (Wilson et al., 2009), which the current study tries to avoid, or excluded this variant from their analysis (Fraser et al., 2013, 2014). The 2012 study by Mesulam and colleagues, which was the first attempt to implement the method developed in their 2009 paper, is the study most comparable to the current study as they also only used behavioral data for the classification and also included the lvPPA. In the Mesulam et al. (2012) work, when only the data from the two relevant tasks were used, the classification accuracy for the lvPPA individuals was 33%. However, Step 3 of the analysis we report on here yielded pairwise classification accuracies between 76 and 82% for lvPPA (Table 5) and the classification accuracy we obtained for lvPPA at Step 4, the analysis which more closely resembles a 3-way classification approach, yielded an accuracy of 59% for this group - higher than what has been previously reported. It is important to note that a direct comparison between the 3-way classification approach of Mesulam and colleagues and our Step 4 analysis is difficult to make. This is largely because

we used a leave-one-out cross-validation method. This allows, on each iteration, applying what the model has learned on the basis of all but one cases, to the classification of a 'new' case. In contrast, a number of other approaches (such as those used by Mesulam and colleagues) take all individuals into account at once without any validation on new cases. For this reason, the leave-one-out cross-validation method allows us to build a more robust classification system which should be more accurate in classifying new individuals.

The results of the current study also validate the proposal to use the spelling performance of individuals with PPA for variant prediction. While the general language profile of lvPPA has been described as a mixture of the nfvPPA and the svPPA profiles (Gorno-Tempini et al., 2011), the exact language and cognitive mechanisms that are impaired for this population have not yet been defined. In accordance with previous studies on the spelling of PPA individuals, and as discussed in the previous section, the current findings indicate that lvPPA have an impairment that affects both lexical and the sublexical spelling routes. There are two alternative interpretations for this relationship: there may be a single impaired process that is used by both routes, or their impairment affects independent processes that are supported by nearby substrates that have high likelihood of both being affected by the underlying disease. Although we cannot currently adjudicate between these two possibilities, the multilevel and rich structure of spelling does allow us to capture the complexity of the deficits in lvPPA providing enough information to achieve reasonable classification success.

4.3. The merit of the methods

As discussed, we believe that spelling tasks can be particularly useful for PPA variant classification because they can serve to index spoken language processes. Besides using spelling data though, the analytic approach was critical to the high classification accuracy we achieved. In this section, we discuss the methodological components that seem to have been particularly useful in this regard.

First, the three variants were not classified within a single comparison, but rather, similar to Wilson et al. (2009), we constructed three pairwise comparisons. Given the similarities between lvPPA and both other variants, it may not be surprising that previous attempts to distinguish the three variants in a single comparison did not yield high accuracies, especially for lvPPA (see for example Mesulam et al., 2012). The pairwise comparison approach allows the specific differences between pairs of variants to be exploited to obtain high classification accuracies. Given this, for studies that originally used 3-way classification, it would be important to see how classification accuracy rates would change if pairwise comparisons were used instead.

Second, similar to other studies of PPA, the sample sizes were relatively small - 33 PPA individuals, with the within-variant sample sizes ranging between 6 and 17 individuals. Therefore, every piece of data was extremely valuable for an accurate classification. The use of a 'leave-one-out' training, rather than some other splitting procedures (e.g., 80% of data for training and 20% of data for testing) allowed use of the maximum amount of available data for training.

In summary, the high classification accuracies we obtained were the result of both the richness of the spelling data set, but also of certain methodological steps that can be considered in future research.

4.4. Study limitations

The current study is the first key step towards the development of a simpler and more clinically suitable spelling-based tool for PPA variant classification. However, there are certain issues that must be addressed before this type of tool would be clinically useful. First, even though we used data from a single task, a spelling to dictation task, the participants spelled an average of approximately a hundred words. In order for the spelling task to be practical and useful for clinicians, the spelling list needs to be shorter. Second, the analytic methods were fairly elaborate, as they required scoring every individual letter of every word, advanced statistical analyses for quantifying the contribution of the various variables and multiple pairwise comparisons. Parts of this pipeline, like the letter scoring, can be automated while the use of simpler analytic approaches should be investigated. Third, we did not compare the relative benefits of spelling compared to other single tasks in the context of the particular analytic approach we used and thus cannot conclude that spelling is the best single task.

Another limitation of the study stems from the unbalanced sample sizes across the three variant groups and the relatively small sample size of the svPPA group. Because of this, it is hard to know the extent to which these results would generalize to a new cohort. We did assess generalization across the cohort via internal validation, using the leave-one-out cross-validation procedure. However, generalizability of the classification approach is best evaluated through external validation, by using a truly independent sample.

Finally, it is important to acknowledge, as indicated in the Introduction, that there are certain intrinsic shortcomings associated with PPA classification. These stem not only from the fact there is not (as yet) a direct relationship between underlying pathology and clinical subtype, but also due to the progressive nature that produces changing symptoms and mixed and uncertain classifications. These challenges are not limited to this study, but face all work involving clinical subtyping in PPA.

5. Conclusions

Normally, an extensive, comprehensive battery of language and cognitive tests would be needed to accurately classify PPA variants, yet we have shown that a single spelling task provides a wealth of useful information. Specifically, spelling performance on words *vs.*pseudowords, phonologically plausible errors, as well as the effects of imageability, length, position and lexicality were the six variables that allowed for high classification accuracy for each of the three PPA variants. While the analysis process described in this paper is complex, it is nonetheless an objective approach that does not require prior notions regarding the critical features needed for variant classification.

The high classification accuracy obtained using spelling data from the three PPA variants strengthens the importance of spelling in the domain of language research and provides the first empirical evidence that spelling can serve as an efficient PPA classification tool.

Acknowledgments

Appendix A

1. Classification criteria for each variant as proposed in Gorno-Tempini et al. (2011). For classifying an individual as **lvPPA** both core features need to present, plus a minimum of 3 optional features, either from the optional-absent or the optional-present features. For classifying an individual as **nfvPPA** either of the core features need to present, plus a minimum of 2 optional features, either from the optional-absent or the optional-present features. Finally, for classifying an individual as **svPPA** both core features need to present, plus a minimum of 3 optional features, either from the optional-absent or the optional-present features. *Note: following the original* Gorno-Tempini et al. (2011) *criteria, for svPPA only, 'ungrammaticality of sentence production' and 'effortful speech/impaired motor speech' are grouped together as a single optional feature, describing speech production as a whole.*

Classification criteria based on Gorno-Tempini et al. (2011) *guidelines*	PPA Variant		
	lvPPA	*nfvPPA*	*svPPA*
Ungrammaticality of sentence production: 1. Poor performance on passive sentences in sentence construction; 2.Agrammatic spontaneous speech, with short phrases and omissions of grammatical morphemes	optional - *absent*	core	optional - *absent*
Effortful spontaneous speech/impaired motor speech	optional - *absent*	core	
Prominent word finding difficulties in spontaneous speech	core		
Poor single word comprehension	optional - *absent*	optional - *absent*	core
Impaired naming			core
Impaired repetition of sentences	core		optional - *absent*
Poor syntax comprehension, with worse performance on passive and object-relative sentences compared to active and subject-relative sentences		optional - *present*	
Impaired object knowledge	optional - *absent*	optional - *absent*	optional - *present*
Evidence of surface dyslexia and/or dysgraphia			optional - *present*
Phonemic errors in spontaneous speech	optional - *present*		

2. The correspondence between the variant classification criteria proposed by Gorno-Tempini et al. and the specific language tasks used for the variant classification in the current study. We also, include individual participant performance scores on the relevant tasks.(1 - Sentence Anagarams (Johns Hopkins University Sentence Anagrams, an in-house test developed at Johns Hopkins University); 2 - Picture Description (Goodglass et al., 2000); 3 – BERNDT (Berndt et al., 1997); 4 – PALPA (Kay et al., 1996); 5 – Semantic Word-Picture Matching (taken from National Alzheimer's Coordinating Center (NACC) battery for Frontotemporal Dementia, available at www.alz.washington.edu); 6 – BNT (Kaplan et al., 2001); 7 – Sentence Repetition (Hillis, 2015); 8 – SOAP (Love and Oster, 2002); 9 –

Pyramids & Palm Trees (Howard and Patterson, 1992); 10 – Passage Reading (Dabul, 2000);
11 – Single Word Spelling (Goodman and Caramazza, 1985)

ID/ Features (*Task*)	Grammaticality of sentence production (*1. Sentence Anagrams[1] (A: active; P: passive); 2. Agrammatic speech in Picture Description[2]*)	Effortful speech (*Picture Description[2]*)	Word Finding Difficulty (*Picture Description[2]*)	Single Word Comprehension (*BERNDT[3] or PALPA[4] or Semantic Word-Picture Matching[5]*)	Naming (*Boston Naming Test (BNT)[6] (total:30)*)	Repetition (*Sentence Repetition[7] (total:37)*)	Syntax Comprehension (*SOAP test[8] (total: 40; A: active; P: passive; SR: Subject Relative; OR: Object Relative)*)	Object Knowledge (*Pyramids & Palm Trees[9] (total: 15)*)	
ABK	Sentence Anagrams: A: 0/5; P: 0/5; Agrammatic Speech	YES	YES		10/30	7/37	total: 16/40; A:7, P:3, SR:5, OR:1	15/15	Y
BIN	Sentence Anagrams: A: 5/5; P: 5/5	YES	NO	100%	29/30	36/37	total: 35/40; A:10, P:10, SR:10, OR:5	15/15	N
BLR	Sentence Anagrams: A:5/5; P: 3/5	NO	YES		26/30	31/37	total: 29/40; A:8, P:7, SR:9, OR:5	15/15	N
BNR	Sentence Anagrams: A: 5/5; P: 0/5			59%	3/30	12/37	total: 27/40; A:6, P:8, SR:7, OR:6	14/15	
CBN	Sentence Anagrams: A: 4/5; P: 0/5; Agrammatic Speech	YES	YES	100%	4/30	11/37	total: 26/40; A:8, P:5, SR:9, OR:4	15/15	Y
CBT	Sentence Anagrams: A: 5/5; P: 4/5	NO	YES	100%	16/30	27/37	total: 33/40; A:8, P:7, SR:9, OR:9	15/15	N
CKI	Sentence Anagrams: A: 4/5; P: 0/5; Agrammatic Speech	YES	YES		18/30	31/37	total: 25/40; A:8, P:5, SR:6, OR:6	15/15	N
DCN	Sentence Anagrams: A: 0/5; P: 0/5	NO	YES	25%	0/30	34/37	total: 24/40; A:8, P:2, SR:7, OR:5	10/15	Y
DEK	Sentence Anagrams: A: 5/5; P: 5/5	NO	YES	(notes on impaired comprehesnion by physician)	1/30	35/37	total: 39/40; A:9, P:10, SR:10, OR:10	12/15	N
DME	Sentence Anagrams:A: 0/5; P: 0/5	YES	YES		24/30	33/37	total: 21/40; A:9, P:5, SR:5, OR:2	15/15	N
DNE	Sentence Anagrams: A: 5/5; P: 2/5	NO	YES	92%	22/30	23/37	total: 31/40; A:8, P:9, SR:6, OR:8	14/15	N
DPD	Sentence Anagrams:	YES	YES		25/30	28/37	total: 31/40; A:10, P:8, SR:9, OR:4	15/15	Y
DPZ	Sentence Anagrams: 0/10; Agrammatic Speech	YES	NO		23/30	23/37	total: 20/40; A:7, P:6, SR:3, OR:4	12/15	Y

ID/ Features (*Task*)	Grammaticality of sentence production (*1. Sentence Anagrams[1] (A: active; P: passive); 2. Agrammatic speech in Picture Description[2]*)	Effortful speech (*Picture Description[2]*)	Word Finding Difficulty (*Picture Description[2]*)	Single Word Comprehension (*BERNDT[3] or PALPA[4] or Semantic Word-Picture Matching[5]*)	Naming (*Boston Naming Test (BNT)[6] (total:30)*)	Repetition (*Sentence Repetition[7] (total:37)*)	Syntax Comprehension (*SOAP test[8] (total: 40; A: active; P: passive; SR: Subject Relative; OR: Object Relative)*)	Object Knowledge (*Pyramids & Palm Trees[9] (total: 15)*)	S...
DRS	Sentence Anagrams: A: 5/5; P: 4/5	NO	YES	95%	28/30	26/37	total: 32/40; A:10, P:9, SR:10, OR:3	15/15	Y
DTL	Sentence Anagrams: A: 2/5; P: 0/5	YES	YES	(word comprehension was intact, based on physician's notes)	14/30	30/37	total: 32/40; A:9, P:9, SR:7, OR:7	13/15	N
DUE	Sentence Anagrams:A: 2/5; P: 0/5	NO	YES	63%	1/30	10/37	total: 19/40; A:6, P:2, SR:7, OR:4	15/15	Y
ERM	Sentence Anagrams: A: 5/5; P: 2/5; Agrammatic Speech	YES	NO	100%	27/30	35/37	total: 33/40; A:9, P:10, SR:9, OR:5	15/15	N
GFS	Sentence Anagrams: A: 4/5; P: 0/5	NO		37%	1/30	0/37	total: 17/40; A:3, P:4, SR:6, OR:4	15/15	
GSH	Sentence Anagrams: A: 5/5; P: 4/5	NO	YES		24/30	36/37	total: 35/40; A:10,P:10, SR:10, OR:5	15/15	N
IJN	Sentence Anagrams: A: 5/5; P: 0/5; Agrammatic Speech	NO	YES		16/30	24/37	total: 22/40; A:9, P:5, SR:6, OR:2	15/15	Y
JEE	Sentence Anagrams: A: 5/5; P: 0/5	NO	YES		0/30	30/37	total: 23/40; A:7, P:6, SR:5, OR:5	13/15	Y
JHR	Sentence Anagrams: A: 5/5; P: 5/5	NO	NO		25/30	35/37	total: 40/40; A:10, P:10, SR:10, OR:10	15/15	N
JJN	Sentence Anagrams: A: 3/5; P: 1/5			35%	2/30	33/37	total: 20/40; A:8, P:4, SR:3, OR:5	13/15	Y
JKA	Sentence Anagrams: A: 5/5; P: 5/5	NO	YES	85%	4/30	37/37	total: 37/40; A:10, P:9, SR:9, OR:9	15/15	N
JRD	Sentence Anagrams: A: 5/5; P: 3/5	NO	YES	95%	12/30	21/37		15/15	
JRE	Sentence Anagrams: A: 5/5; P: 4/5	NO	YES	97%	24/30	37/37	total: 34/40; A:10, P:10, SR:10, OR:4	15/15	N
JSS	Sentence Anagrams: A: 2/5; P: 1/5; Agrammatic Speech	NO	YES	95%	28/30	29/37	total: 27/40; A:7, P:7, SR:7, OR:6	14/15	N

ID/ Features (*Task*)	Grammaticality of sentence production *(1. Sentence Anagrams[1] (A: active; P: passive); 2. Agrammatic speech in Picture Description[2])*	Effortful speech *(Picture Description[2])*	Word Finding Difficulty *(Picture Description[2])*	Single Word Comprehension *(BERNDT[3] or PALPA[4] or Semantic Word-Picture Matching[5])*	Naming *(Boston Naming Test (BNT)[6] (total:30))*	Repetition *(Sentence Repetition[7] (total:37))*	Syntax Comprehension *(SOAP test[8] (total: 40; A: active; P: passive; SR: Subject Relative; OR: Object Relative)*	Object Knowledge *(Pyramids & Palm Trees[9] (total: 15))*	S...D...L...(A...r...S...w...s...
JWE	Sentence Anagrams: A: 5/5; P: 4/5	NO	YES	100%	5/30	35/37	total: 39/40; A:9, P:10, SR:10, OR:10	11/15	N
KBG	Sentence Anagrams: A: 5/5; P: 0/5	NO	NO	100%	28/30	35/37	total: 18/40; A:7, P:1, SR:6, OR:4	15/15	N
KCE	Agrammatic Speech	YES		80%	2/30	19/37	total: 15/40; A:4, P:3, SR:6, OR:2	13/15	Y
LCR	Sentence Anagrams:A: 1/5; P: 0/5	NO	YES	60%	9/30	4/37	total: 27/40; A:7, P:7, SR:5, OR:8	12/15	Y
MOR	Sentence Anagrams: A: 2/5; P: 4/5	YES	YES	95%	25/30	28/37	total: 29/40; A:9, P:7, SR:5, OR:8	15/15	
MPI	Sentence Anagrams: A: 5/5; P: 4/5	YES	YES		14/30	29/37	total: 34/40; A:10, P:9, SR:8, OR:7	15/15	Y
MVR	Sentence Anagrams: A: 4/5; P: 0/5	NO	YES		27/30	32/37	total: 21/40; A:5, P:8, SR:4, OR:4	8/15	N
NCG	Sentence Anagrams: A: 1/5; P: 0/5	NO	YES	75%	0/30	33/37	total: 27/40; A:8, P:6, SR:10, OR:3	11/15	Y
RFH	Sentence Anagrams: A: 5/5; P: 0/5	NO	YES	*(notes by physician that comprehension was impaired for instructions and less familiar words)*	2/30	31/37	total: 25/40; A:9, P:2, SR:9, OR:5	14/15	N
RVT/RWT	Sentence Anagrams: A: 4/5; P: 0/5			62%	25/30		total: 21/40; A:7, P:3, SR:9, OR:2	15/15	
SKR	Agrammatic Speech	YES	YES	73%	14/30	15/37		15/15	Y
SLR	Sentence Anagrams: A: 5/5; P: 5/5			100%	9/30	36/37	total: 38/40; A:10, P:9, SR:10, OR:9	15/15	
TBD	Sentence Anagrams: A: 5/5; P: 4/5	NO	YES	98%	29/30	31/37	total: 34/40; A:9, P:8, SR:10, OR:4	15/15	N
TBE	Sentence Anagrams: A: 5/5; P: 5/5	NO	YES		8/30	32/37	total: 35/40; A:10, P:10, SR:10, OR:5	15/15	N
TBT	Sentence Anagrams: A: 4/5; P: 4/5	NO	YES	100%	30/30				N

Appendix B.

Predicted variant for each pairwise comparison (L = lvPPA; NF = nfvPPA; S = svPPA, U = unclassified). Final classification is defined by the number of overlapping pairwise variant predictions: 2 overlapping predictions = accurately classified (✓) or misclassified (x); 0 overlapping predictions = unclassified (?)

| ID	Actual Variant	lvPPA vs svPPA	nfvPPA vs svPPA	lvPPA vs nfvPPA	Final Classification	Classification Accuracy
ABK	nfvPPA	L	NF	NF	NF	✓
BIN	nfvPPA	L	L	NF	L	x
BLR	lvPPA	S	L	S	S	x
BNR	lvPPA	L	L	NF	L	✓
CBN	nfvPPA	L	NF	NF	NF	✓
CBT	lvPPA	S	L	S	S	x
CKI	nfvPPA	L	NF	NF	NF	✓
DCN	svPPA	S	L	S	S	✓
DEK	svPPA	S	L	S	S	✓
DME	lvPPA	L	NF	NF	NF	x
DNE	lvPPA	L	L	NF	L	✓
DPD	nfvPPA	L	L	NF	L	x
DPZ	nfvPPA	L	NF	NF	NF	✓
DRS	lvPPA	S	L	S	S	x
DUE	lvPPA	L	L	NF	L	✓
ERM	nfvPPA	L	NF	NF	NF	✓
GFS	lvPPA	L	L	NF	L	✓
IJN	lvPPA	L	L	S	L	✓
JJN	svPPA	L	NF	NF	NF	x
JKA	svPPA	S	L	S	S	✓
JRD	lvPPA	L	L	NF	L	✓
JRE	lvPPA	L	L	S	L	✓
JSS	lvPPA	L	L	NF	L	✓
KBG	nfvPPA	L	L	NF	L	x
MOR	lvPPA	L	NF	NF	NF	x
MPI	lvPPA	L	L	NF	L	✓
MVR	lvPPA	L	NF	S	U	?
NCG	svPPA	L	L	S	L	x
RFH	svPPA	S	L	S	S	✓
RVT	nfvPPA	L	NF	NF	NF	✓
SKR	nfvPPA	L	NF	NF	NF	✓
TBD	lvPPA	L	L	NF	L	✓
TBE	lvPPA	S	L	NF	U	?

References

Baayen RH, Davidson DJ, Bates DM, 2008. Mixed-effects modeling with crossed random effects for subjects and items. J. Mem. Lang. 59, 390–412.

Baayen RH, Piepenbrock R, Gulikers L, 1995. The CELEX Lexical Database (Release 2). Distrib. Linguist. Data Consort. Univ. Pa.

Balota DA, Yap MJ, Hutchison KA, Cortese MJ, Kessler B, Loftis B, Neely JH, Nelson DL, Simpson GB, Treiman R, 2007. The English lexicon project. Behav. Res. Methods 39, 445–459. [PubMed: 17958156]

Bates D, Maechler M, 2009. Package 'lme4'(Version 0.999375-32): Linear Mixed-Effects Models Using S4 Classes. Orgwebpackageslme4lme4 Pdf Available 4 2011 httpcran R-Proj.

Berndt RS, Mitchum CC, Haendiges AN, Sandson J, 1997. Verb retrieval in aphasia. 1. Characterizing single word impairments. Brain Lang. 56, 68–106. [PubMed: 8994699]

Bonner MF, Vesely L, Price C, Anderson C, Richmond L, Farag C, Avants B, Grossman M, 2009. Reversal of the concreteness effect in semantic dementia. Cogn. Neuropsychol. 26, 568–579. 10.1080/02643290903512305. [PubMed: 20183015]

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

Bub D, Kertesz A, 1982. Evidence for lexicographic processing in a patient with preserved written over oral single word naming. Brain 105, 697–717. [PubMed: 7139251]

Buchwald A, Rapp B, 2004. Rethinking the graphemic buffer? Brain lang. Special Abstract Issue: Academy of Aphasia 2004 Program 91, 100–101. 10.1016/j.bandl.2004.06.052.

Caramazza A, Miceli G, 1990. The structure of graphemic representations. Cognition 37, 243–297. [PubMed: 2282774]

Coltheart M, 1981. The MRC psycholinguistic database. Q. J. Exp. Psychol. 33, 497–505.

Crutch SJ, Warrington EK, 2005. Gradients of semantic relatedness and their contrasting explanations in refractory access and storage semantic impairments. Cogn. Neuropsychol. 22, 851–876. [PubMed: 21038279]

Crutch SJ, Warrington EK, 2003. Preservation of propositional speech in a pure anomic: the importance of an abstract vocabulary. Neurocase 9, 465–481. [PubMed: 16210229]

Dabul B, 2000. Apraxia Battery for Adults: Examiner's Manual. Pro-ed.

Faria AV, Crinion J, Tsapkini K, Newhart M, Davis C, Cooley S, Mori S, Hillis AE, 2013. Patterns of dysgraphia in primary progressive aphasia compared to post-stroke aphasia. [WWW Document]. Behav. Neurol. 10.3233/BEN-2012-110237.

Folk JR, Rapp B, Goldrick M, 2002. The interaction of lexical and sublexical information in spelling: what's the point? Cogn. Neuropsychol. 19, 653–671. [PubMed: 20957558]

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]

Fraser KC, Rudzicz F, Rochon E, 2013. Using text and acoustic features to diagnose progressive aphasia and its subtypes. In: INTERSPEECH, pp. 2177–2181.

Friedmann N, Gvion A, 2001. Letter position dyslexia. Cogn. Neuropsychol. 18, 673–696. [PubMed: 20945233]

Gil-Navarro S, Lladó A, Rami L, Castellví M, Bosch B, Bargalló N, et al. , 2013. Neuroimaging and biochemical markers in the three variants of primary progressive aphasia. Dement. Geriatr. Cognit. Disord. 35 (1–2), 106–117. [PubMed: 23392204]

Goldrick M, Folk JR, Rapp B, 2010. Mrs. Malaprop's neighborhood: using word errors to reveal neighborhood structure. J. Mem. Lang. 62, 113–134. 10.1016/j.jml.2009.11.008. [PubMed: 20161591]

Goodglass H, Kaplan E, Barresi B, 2000. Boston Diagnostic Aphasia Examination Record Booklet. Lippincott Williams & Wilkins.

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

Goodman-Schulman R, Caramazza A, 1987. Patterns of dysgraphia and the nonlexical spelling process. Cortex J. Devoted Study Nerv. Syst. Behav. 23, 143–148.

Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, Ogar JM, Rohrer JD, Black S, Boeve BF, Manes F, Dronkers NF, Vandenberghe R, Rascovsky K, Patterson K, Miller BL, Knopman DS, Hodges JR, Mesulam MM, Grossman M, 2011. Classification of primary progressive aphasia and its variants. Neurology 76, 1006–1014. [PubMed: 21325651]

Graham NL, 2014. Dysgraphia in primary progressive aphasia: characterisation of impairments and therapy options. Aphasiology 28, 1092–1111.

Hanna PR, 1966. Phoneme-grapheme Correspondences as Cues to Spelling Improvement.

Harris JM, Gall C, Thompson JC, Richardson AM, Neary D, du Plessis D, Pal P, Mann DM, Snowden JS, Jones M, 2013. Classification and pathology of primary progressive aphasia. Neurology 81, 1832–1839. [PubMed: 24142474]

Henry ML, Beeson PM, Alexander GE, Rapcsak SZ, 2011. Written language impairments in primary progressive aphasia: a reflection of damage to central semantic and phonological processes. J. Cogn. Neurosci. 24, 261–275. 10.1162/jocn_a_00153. [PubMed: 22004048]

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

Hoffman P, Lambon Ralph MA, 2011. Reverse concreteness effects are not a typical feature of semantic dementia: evidence for the hub-and-spoke model of conceptual representation. Cerebr. Cortex 21, 2103–2112.

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

Jones AC, Folk JR, Rapp B, 2009. All letters are not equal: subgraphemic texture in orthographic working memory. J. Exp. Psychol. Learn. Mem. Cogn. 35, 1389. [PubMed: 19857011]

Juhasz BJ, Lai Y-H, 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]

Kaplan E, Goodglass H, Weintraub S, 2001. Boston Naming Test. Pro-ed.

Kay J, Lesser R, Coltheart M, 1996. Psycholinguistic assessments of language processing in aphasia (PALPA): an introduction. Aphasiology 10, 159–180.

Knopman DS, Kramer JH, Boeve BF, Caselli RJ, Graff-Radford NR, Mendez MF, et al. , 2008. Development of methodology for conducting clinical trials in frontotemporal lobar degeneration. Brain 131 (11), 2957–2968. [PubMed: 18829698]

Love T, Oster E, 2002. On the categorization of aphasic typologies: the SOAP (A test of syntactic complexity). J. Psycholinguist. Res. 31, 503–529. 10.1023/A:1021208903394. [PubMed: 12528429]

Leyton CE, Villemagne VL, Savage S, Pike KE, Ballard KJ, Piguet O, et al. , 2011. Subtypes of progressive aphasia: application of the international consensus criteria and validation using β-amyloid imaging. Brain 134 (10), 3030–3043. [PubMed: 21908392]

Matias-Guiu JA, Cabrera-Martín MN, Moreno-Ramos T, García-Ramos R, Porta-Etessam J, Carreras JL, Matías-Guiu J, 2015. Clinical course of primary progressive aphasia: clinical and FDG-PET patterns. J. Neurol. 262 (3), 570–577. [PubMed: 25491078]

McCloskey M, Macaruso P, Rapp B, 2006. Grapheme-to-lexeme feedback in the spelling system: evidence from a dysgraphic patient. Cogn. Neuropsychol. 23, 278–307. [PubMed: 21049332]

Mesulam M, Wieneke C, Rogalski E, Cobia D, Thompson C, Weintraub S, 2009. Quantitative template for subtyping primary progressive aphasia. Arch. Neurol. 66, 1545–1551. [PubMed: 20008661]

Mesulam M-M, 1987. Primary progressive aphasia—differentiation from Alzheimer's disease. Ann. Neurol. 22, 533–534. 10.1002/ana.410220414. [PubMed: 3324947]

Mesulam MM, 1982. Slowly progressive aphasia without generalized dementia. Ann. Neurol. 11, 592–598. [PubMed: 7114808]

Mesulam MM, Weintraub S, 2014. Is it time to revisit the classification guidelines for primary progressive aphasia? Neurology 82, 1108–1109. [PubMed: 24598706]

Mesulam MM, Wieneke C, Thompson C, Rogalski E, Weintraub S, 2012. Quantitative classification of primary progressive aphasia at early and mild impairment stages. Brain J. Neurol. 135, 1537–1553.

Oldfield RC, Wingfield A, 1965. Response latencies in naming objects. Q. J. Exp. Psychol. 17, 273–281. [PubMed: 5852918]

Rapp B, Fischer-Baum S, 2015. Uncovering the cognitive architecture of spelling. In: The Handbook of Adult Language Disorders. Psychology Press, New York, pp. 59–86.

Rapp B, Lipka K, 2011. The literate brain: the relationship between spelling and reading. J. Cogn. Neurosci. 23, 1180–1197. [PubMed: 20433242]

Rapp B, Purcell J, Hillis AE, Capasso R, Miceli G, 2015. Neural bases of orthographic long-term memory and working memory in dysgraphia. Brain 139, 588–604. [PubMed: 26685156]

Rogalski E, Cobia D, Harrison TM, Wieneke C, Weintraub S, Mesulam MM, 2011. Progression of language decline and cortical atrophy in subtypes of primary progressive aphasia. Neurology 76, 1804–1810. [PubMed: 21606451]

Romani C, Mcalpine S, Martin RC, 2008. Concreteness effects in different tasks: implications for models of short-term memory. Q. J. Exp. Psychol. 61, 292–323.

Roux S, Bonin P, 2009. Neighborhood effects in spelling in adults. Psychon. Bull. Rev. 16, 369–373. [PubMed: 19293109]

Sage K, Ellis AW, 2004. Lexical influences in graphemic buffer disorder. Cogn. Neuropsychol. 21, 381–400. [PubMed: 21038212]

Sajjadi SA, Patterson K, Arnold RJ, Watson PC, Nestor PJ, 2012. Primary progressive aphasia: a tale of two syndromes and the rest. Neurology 78 (21), 1670–1677. [PubMed: 22573633]

Schiller NO, Greenhall JA, Shelton JR, Caramazza A, 2001. Serial order effects in spelling errors: evidence from two dysgraphic patients. Neurocase 7, 1–14. [PubMed: 11239072]

Sepelyak K, Crinion J, Molitoris J, Epstein-Peterson Z, Bann M, Davis C, Newhart M, Heidler-Gary J, Tsapkini K, Hillis AE, 2011. Patterns of breakdown in spelling in primary progressive aphasia. Cortex 47, 342–352. [PubMed: 20060967]

Shallice T, 1981. Phonological agraphia and the lexical route in writing. Brain J. Neurol. 104, 413–429.

Shim H, Hurley RS, Rogalski E, Mesulam M-M, 2012. Anatomic, clinical, and neuropsychological correlates of spelling errors in primary progressive aphasia. Neuropsychologia 50, 1929–1935. [PubMed: 22579708]

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

Tainturier M-J, 2013. Lexical neighborhood effects in pseudoword spelling. Front. Psychol. 4, 862. [PubMed: 24348436]

Tainturier M-J, Rapp B, 2003. Is a single graphemic buffer used in reading and spelling? Aphasiology 17, 537–562.

Tainturier M-J, Rapp B, 2001. The spelling process. Handb. Cogn. Neuropsychol. What Deficits Reveal Hum. Mind 263–289.

Team, R.C., 2013. R: A Language and Environment for Statistical Computing.

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.

Venables WN, Ripley BD, 2002. Modern Applied Statistics with S, fourth ed. Springer-Verlag, New York Statistics and Computing.

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

Warrington EK, 1975. The selective impairment of semantic memory. Q. J. Exp. Psychol. 27, 635–657. [PubMed: 1197619]

Wicklund MR, Duffy JR, Strand EA, Machulda MM, Whitwell JL, Josephs KA, 2014. Quantitative application of the primary progressive aphasia consensus criteria. Neurology 82 (13), 1119–1126. [PubMed: 24598709]

Wiley RW, Rapp B, 2018. Statistical analysis in Small-N Designs: using linear mixed-effects modeling for evaluating intervention effectiveness. Aphasiology 1–30. [PubMed: 33012945]

Wilson SM, Dronkers NF, Ogar JM, Jang J, Growdon ME, Agosta F, Henry ML, Miller BL, Gorno-Tempini ML, 2010. Neural correlates of syntactic processing in the nonfluent variant of primary progressive aphasia. J. Neurosci. Off. J. Soc. Neurosci. 30, 16845–16854.

Wilson SM, Ogar JM, Laluz V, Growdon M, Jang J, Glenn S, Miller BL, Weiner MW, Gorno-Tempini ML, 2009. Automated MRI-based classification of primary progressive aphasia variants. Neuroimage 47, 1558–1567. [PubMed: 19501654]

Wing AM, Baddeley AD, 1980. Spelling errors in handwriting: a corpus and a distributional analysis. In: Frith U (Ed.), Cognitive Processes in Sspelling. Academic Press, London, pp. 251–285.

Yi H-A, Moore P, Grossman M, 2007. Reversal of the concreteness effect for verbs in patients with semantic dementia. Neuropsychology 21, 9–19. 10.1037/0894-4105.21.1.9. [PubMed: 17201526]
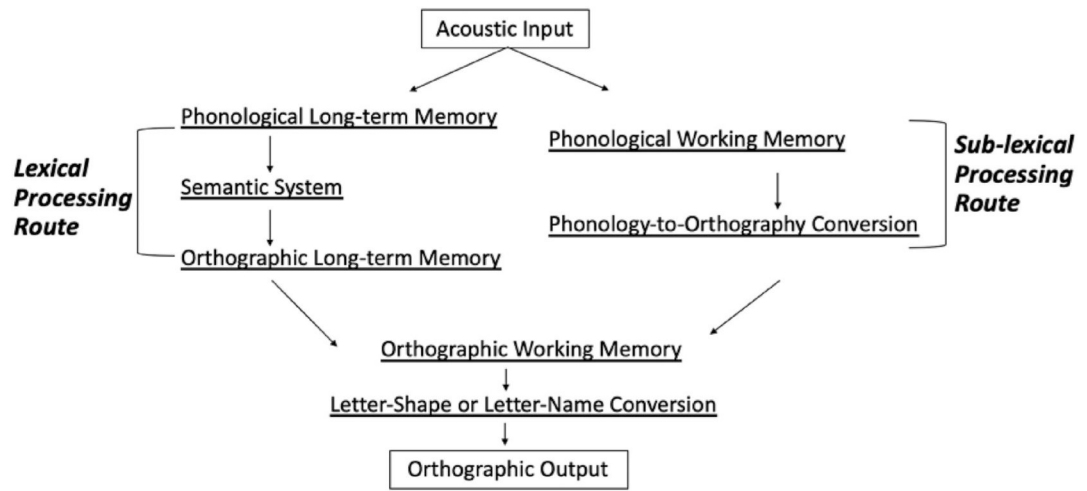
Fig. 1.
The cognitive architecture of spelling.
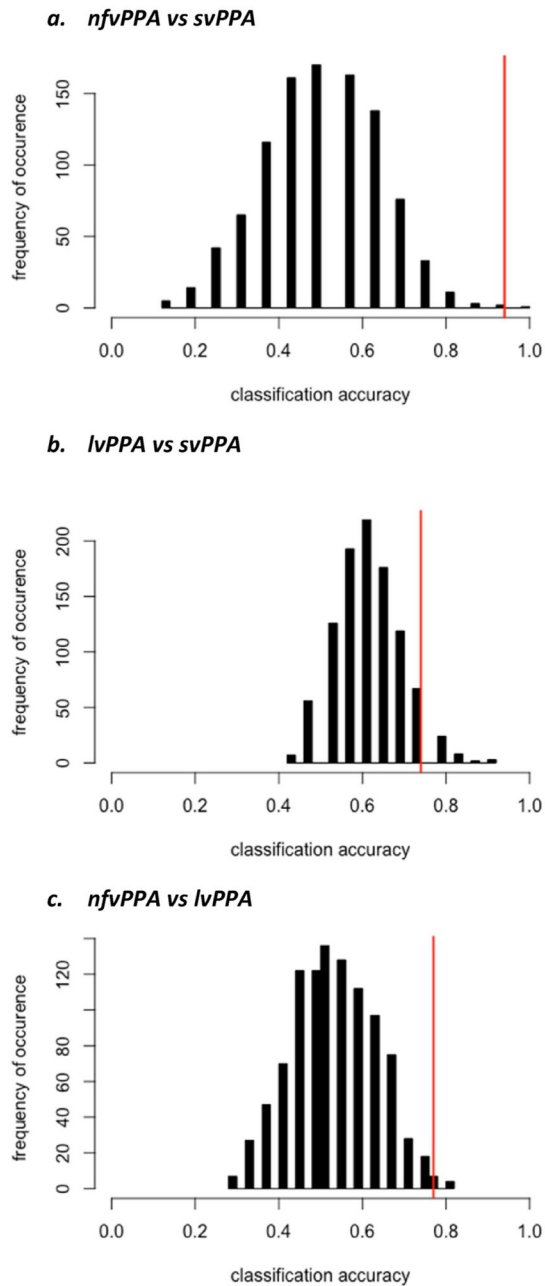
Fig. 2.
Distribution of accuracy values across the 1000 permutations per pairwise comparison. The thin, red line indicates the accuracy value obtained by the actual data. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
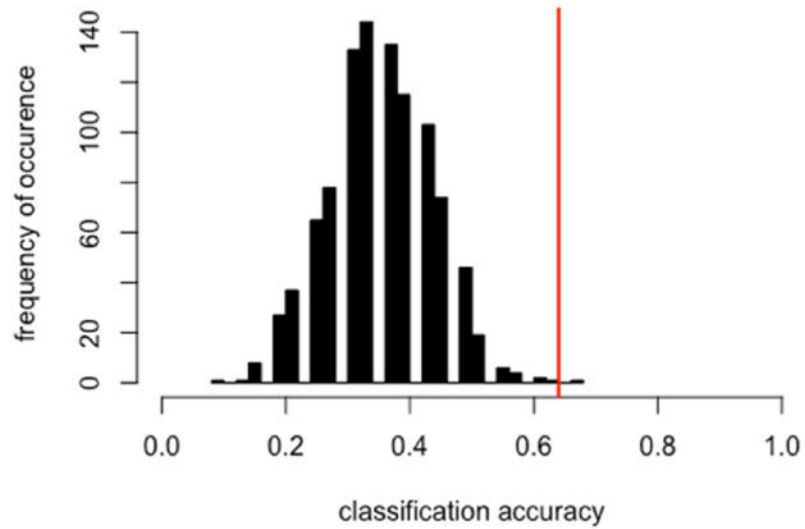
Fig. 3.
Distribution of classification accuracy values across the 1000 random permutations of variant label assignments, across the three variants. The thin, red line indicates the accuracy value obtained from the actual data. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1

Description of the variables evaluated in Step 1 of the analysis that served to obtain beta coefficient values for each variable for each participant on the basis of single-participant LMEM fitting.

Variable	Description
Lexicality	The lexicality status of the letter string, i.e., whether it was a real word or a pseudoword (coded as 1 for words and 0 for pseudowords).
Length	The length of each letter string in terms of the number of letters, ranging from 3 to 8 letters.
Position	The position of a letter in a word (starting from the left), scaled to the length of the word. Since the variable is scaled to the length of the word, it will always have a value of 0, the final letter a value of 1, and the letters falling in between will have values that will sum up to 1. For instance, the four letters of a 4-letter word were coded as follows: letter 1 = 0; letter 2 = 0.33; letter 3 = 0.66, letter 4 = 1. This variable quantifies if there was an increase or a decrease in performance across letter positions.
Position-quadratic	Position-quadratic is the quadratic form of the linear variable Position. To calculate the values for this variable, the linear position values were transformed into quadratic values. For instance, the four letters of a 4-letter word were coded as follows: letter 1 = (0^2) = 0; letter 2 = (0.33^2) = .11; letter 3 = (0.66^2) = 0.44, letter 4 = (1^2) = 1. By performing a regression with a quadratic term (i.e., a non-linear term), we investigate the possibility of the relation between position and accuracy being non-linear. Since an equation with a quadratic term gives a parabola, we can see whether performance on the two ends of a word is better than the middle positions (i.e. an upward parabola), as has been reported for some individuals (e.g., Buchwald and Rapp, 2004; Tainturier and Rapp, 2003).
Imageability	The imageability rating (ranging from 174 to 360) of each word from the MRC database (Coltheart, 1981). If values were absent from this database, the Stadthagen-Gonzalez and Davis (2006), and the Juhasz et al. (2015) [7] databases were used.
Frequency	The written frequency of each word (ranging from 0.72 to 2614.04) from the CELEX database (Baayen et al., 1995) and log transformed.
PG Conversion Probability	The probability of the graphemic representation for each phoneme in each real word. Values ranged from 0.01 to 96.47 and were extracted from a digitized version of the Hanna (1966) study, developed on-site.
Orthographic Neighborhood Density	The number of words that differed from a target word by changing one letter while preserving the identity and positions of the rest (e.g., *pork – fork*). Values ranged from 1 to 19 and were taken from the English Lexicon Project database (Balota et al., 2007).
Phonological Neighborhood Density	The number of words that differ from a target word by changing one *phoneme* while preserving the identity and positions of the rest, e.g.,/**f**iːt/(feet) -/**h**iːt/(heat). Values ranged from 1 to 39 and were taken from the English Lexicon Project database (Balota et al., 2007).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

---

[7]Juhasz et al. (2015) scores were multiplied by a hundred and rounded to the nearest integer so that ratings could be presented as integers on a scale from 100 to 700, and therefore be on a similar scale as used in the other two databases.

Table 2

Mean, standard deviation and ranges of coefficient values from the LMEM's for the nine variables of interest per variant. Underlined variables correspond to the variables subsequently selected in Step 2 and reported in Table 4. For lexicality, positive coefficients indicate better performance for words, as opposed to pseudowords. For the continuous variables (i.e., length, position, imageability, frequency, PG probability, Orthographic N., Phonological N.) positive coefficients indicate better performance for higher values on those variables. For position-quadratic, positive coefficients indicate better performance on the two ends of a word compared to the middle positions.

Variant/Variable		Lexicality	Length	Position	Position-Quadratic	Imageability	Frequency	PG Probability	Orthographic N.	Phonological N.
lvPPA (N=17)	Mean (SD)	0.52 (1.53)	-0.22 (0.46)	-0.96 (0.94)	1.29 (1.04)	0.22 (0.69)	0.37 (0.38)	0.43 (0.29)	0.64 (0.57)	0.11 (0.94)
	Min	-1.41	-0.96	-3.35	-0.2	-0.46	-0.12	0.00	-0.07	-1.65
	Max	4.45	0.81	0.47	3.23	2.59	1.26	1.10	2.08	2.72
nfvPPA (N=10)	Mean (SD)	2.24 (2.45)	-0.34 (0.26)	-1.31 (0.74)	0.97 (0.90)	0.49 (0.46)	0.31 (0.50)	0.56 (0.32)	0.84 (0.94)	0.06 (0.82)
	Min	-1.36	-0.62	-2.18	-0.23	-0.17	-0.58	0.02	-0.42	-0.9
	Max	6.45	0.18	-0.08	2.86	1.41	1.28	1.03	2.47	1.77
svPPA (N=6)	Mean (SD)	-1.11 (1.38)	0.22 (0.26)	0.01 (0.95)	1.52 (1.18)	-0.31 (0.20)	0.54 (0.46)	0.81 (0.19)	0.33 (0.99)	0.05 (1.40)
	Min	-3.12	-0.10	-1.24	-0.05	-0.59	-0.17	0.51	-1.17	-0.79
	Max	0.85	0.65	1.50	2.74	0.00	1.13	1.00	1.54	2.88

Table 3

Mean and standard deviation values, as well as ranges of percentages for *word-pseudoword grapheme accuracy* and each error type for each variant group. Underlined variables correspond to the variables subsequently selected in Step 2 and reported in Table 4.

Variant/Error Type		Word-Pseudoword Gr. Acc.	Lexical Substitution	PPE	Pseudoword for word	Pseudoword for pseudoword
lvPPA (N=17)	*Mean (SD)*	0.02 (0.08)	0.06 (0.05)	0.1 (0.1)	0.21 (0.21)	0.44 (0.18)
	Min	0.18	0	0	0.01	0.26
	Max	−0.11	0.16	0.32	0.71	0.79
nfvPPA (N=10)	*Mean (SD)*	0.15 (0.14)	0.11 (0.08)	0.05 (0.04)	0.16 (0.15)	0.6 (0.29)
	Min	0.45	0.01	0	0	0.11
	Max	−0.04	0.26	0.14	0.42	0.89
svPPA (N=6)	*Mean (SD)*	−0.05(0.09)	0.04 (0.03)	0.23 (0.07)	0.15 (0.17)	0.31 (0.28)
	Min	0.09	0	0.13	0.04	0.05
	Max	−0.19	0.07	0.32	0.48	0.68

Table 4

Variables that differed across the three variants, based on the p-values obtained from variant pairwise ANOVAs after application of the (arbitrary) threshold of p < .05.

Variable/Pairwise comparison	Informative variables
Lexicality	x
Length	x
Position	x
Position-Quadratic	
Imageability	x
Frequency	
Orthographic Neighborhood Density	
Phonological Neighborhood Density	
Word-Pseudoword Grapheme Accuracy	x
Lexical Substitution Errors	
Phonologically Plausible Errors	x
Pseudoword for Word Errors	
Pseudoword for Pseudoword Errors	

Table 5

Classification accuracy based on binomial leave-one-out cross classification (see text for details).

A: nfvPPA vs svPPA (N=16)		B: lvPPA vs svPPA (N=23)		C: lvPPA vs nfvPPA (N=27)	
Overall	94%	*Overall*	74%	*Overall*	77%
nfvPAA	100%	**lvPPA**	76%	**lvPPA**	82%
svPAA	83%	**svPAA**	67%	**nfvPAA**	70%

Table 6

Confusion matrix with classification accuracy from Step 4 analysis - *white*: accurately classified cases; *light-grey*: misclassified cases, *dark-grey*:unclassified cases.

		TARGET LABEL		
		nfvPPA (n=10)	*svPPA (n=6)*	*lvPPA (n=17)*
PREDICTED LABEL	*nfvPPA*	70% (n=7)	17% (n=1)	12% (n=2)
	svPPA	0% (n=0)	66% (n=4)	18% (n=3)
	lvPPA	30% (n=3)	17% (n=1)	59% (n=10)
	Unclassified	0% (n=0)	0% (n=0)	12% (n=2)