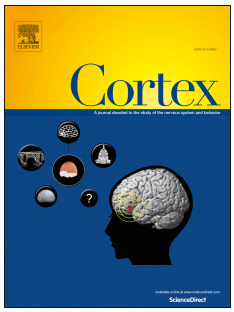
Accepted Manuscript



Machine learning in the clinical and language characterisation of primary progressive aphasia variants

Jordi A. Matias-Guiu, Josefa Díaz-Álvarez, Fernando Cuetos, María Nieves Cabrera-Martín, Ignacio Segovia-Ríos, Vanesa Pytel, Teresa Moreno-Ramos, José Luis Carreras, Jorge Matías-Guiu, José Luis Ayala

PII: S0010-9452(19)30201-1

DOI: https://doi.org/10.1016/j.cortex.2019.05.007

Reference: CORTEX 2646

To appear in: *Cortex*

Received Date: 10 January 2019

Revised Date: 17 March 2019

Accepted Date: 7 May 2019

Please cite this article as: Matias-Guiu JA, Díaz-Álvarez J, Cuetos F, Cabrera-Martín MN, Segovia-Ríos I, Pytel V, Moreno-Ramos T, Carreras JL, Matías-Guiu J, Ayala JL, Machine learning in the clinical and language characterisation of primary progressive aphasia variants, *CORTEX*, https://doi.org/10.1016/j.cortex.2019.05.007.

Title:

Machine learning in the clinical and language characterisation of primary progressive aphasia variants

Authors:

-Jordi A Matias-Guiu [a]

-Josefa Díaz-Álvarez [b]

-Fernando Cuetos [c,d]

-María Nieves Cabrera-Martín [e]

-Ignacio Segovia-Ríos [b]

-Vanesa Pytel [a]

-Teresa Moreno-Ramos [a]

-José Luis Carreras [e]

-Jorge Matías-Guiu [a]

-José Luis Ayala [f]

a) Department of Neurology, Hospital Clínico San Carlos, San Carlos Research Health Institute (IdISSC), Universidad Complutense. Madrid, Spain.
b) Department of Computer Architecture and Communications, Centro Universitario de Mérida, Universidad de Extremadura. Mérida, Spain.
c) Department of Psychology, University of Oviedo. Oviedo, Spain.
d) Department of Psychology, University of Malaga. Málaga, Spain.
e) Department of Nuclear Medicine, Hospital Clínico San Carlos, San Carlos Research Health Institute (IdISSC), Universidad Complutense. Madrid, Spain.
f) Department of Computer Architecture and Automation, Universidad Complutense. Madrid, Spain.

Corresponding author:

Jordi A Matías-Guiu. Department of Neurology. Hospital Clínico San Carlos. Calle Prof. Martín Lagos. 28040 Madrid (Spain). Telephone: +34913303511,

+34676933312. E-mail: jordimatiasguiu@hotmail.com, jordi.matias-guiu@salud.madrid.org

Running title: Machine learning in primary progressive aphasia

ABSTRACT

Introduction. Primary progressive aphasia (PPA) is a clinical syndrome of neurodegenerative origin with 3 main variants: non-fluent, semantic, and logopenic. However, there is some controversy about the existence of additional subtypes. Our aim was to study the language and cognitive features associated with a new proposed classification for PPA.

Material and methods. Sixty-eight patients with PPA in early stages of the disease and 20 healthy controls were assessed with a comprehensive language and cognitive protocol. They were also evaluated with ^{18}F-FDG PET. Patients were classified according to FDG PET regional metabolism, using our previously developed algorithm based on a hierarchical agglomerative cluster analysis with Ward's linkage method. Five variants were found, with both the non-fluent and logopenic variants being split into 2 subtypes. Machine learning techniques were used to predict each variant according to language assessment results.

Results. Non-fluent type 1 was associated with poorer performance in repetition of sentences and reading of irregular words than non-fluent type 2. Conversely, the second group showed a higher degree of apraxia of speech. Patients with logopenic variant type 1 performed more poorly on action naming than patients with logopenic type 2. Language assessments were predictive of PET-based subtypes in 86%-89% of cases using clustering analysis and principal components analysis.

Conclusions. Our study supports the existence of 5 variants of PPA. These variants show some differences in language and FDG PET imaging characteristics. Machine learning algorithms using language test data were able to predict each of the 5 PPA variants with a relatively high degree of accuracy, and enable the possibility of automated, machine-aided diagnosis of PPA variants.

Keywords: primary progressive aphasia; apraxia of speech; positron emission tomography; neuropsychological assessment.

Declarations of interest:

None.

1. INTRODUCTION

Primary progressive aphasia (PPA) is a clinical syndrome that can present as the initial manifestation of several neurodegenerative disorders, mainly tauopathies, TDP-43 proteinopathies, and Alzheimer disease (Matias-Guiu & García-Ramos 2013; Marshall et al. 2018). The classification of distinct variants of PPA has evolved in recent years. In 1998, Neary et al. distinguished "progressive non-fluent aphasia" from "semantic dementia with associative agnosia" in fronto-temporal degeneration criteria (Neary et al., 1998). Similarly, Mesulam (2001) classified PPA as being "with agrammatism" or "with comprehension deficits." However, a major breakthrough was made in 2011 with the publication of the current consensus criteria by Gorno-Tempini et al. (2011). PPA was categorised into 3 clinical variants, the non-fluent, semantic, and logopenic subtypes, with each variant defined according to certain language and neuroimaging features.

Application of these criteria has shown a high degree of correlation between clinical diagnosis and neuroimaging findings (Matias-Guiu et al., 2014; Leyton et al., 2011) and, most importantly, improved the ability to predict the underlying pathology. However, several issues in PPA classification have been emphasised since the publication of the consensus criteria. In this regard, some groups have observed that a large proportion of patients remain unclassifiable (Sajjadi et al., 2012; Wicklund et al., 2014), or have experienced various difficulties in differential diagnosis between variants, especially regarding the logopenic subtype (Sajjadi et al., 2014; Hoffman et al., 2017). Furthermore, controversy exists as to whether the non-fluent and logopenic subtypes are unitary disorders (Leyton et al., 2015). In this regard, Botha et al. (2015) suggest separating the non-fluent variant into agrammatic PPA and apraxia of speech; this could be especially relevant given the specificity of apraxia of speech to tau pathology (Josephs et al., 2012), while non-fluent PPA as a whole group has also been associated with TDP-43 proteinopathies. Similarly, an amyloid-negative group has been detected within the logopenic variant, which has been linked to TDP-43 proteinopathies (Matias-

Guiu et al., 2015; Josephs et al., 2014; Rohrer et al., 2010). Heterogeneous clinical and neuroimaging findings have even been described within amyloid-positive logopenic aphasia (Krishnan et al., 2017). Furthermore, prediction of underlying pathology is still incomplete, as there are more pathological entities than the 3 main clinical syndromes recognised (Spinelli et al., 2017; Harris et al., 2013).

In a previous study, we used an unsupervised clustering algorithm to examine alternative classification of PPA based on the analysis of regional brain metabolism in a large cohort of patients with PPA. The use of FDG PET favoured the classification of PPA patients into 6 subtypes, rather than 3. The expansion from 3 to 6 variants arose from splitting non-fluent PPA into 3 variants (denominated non-fluent type 1A, *k0*; type 1B, *k5*; and type 2, *k2*), and the logopenic variant into 2 subtypes (logopenic type 1, *k1;* and type 2, *k4*). Conversely, the semantic variant (*k3*) remained unchanged. Interestingly, the new grouping was more predictive of clinical course and amyloid imaging results (Matias-Guiu et al., 2018). This new classification system supports the current consensus criteria, because 3 main variants were respected, but it does suggest that classification may be improved.

In the present study, we aimed to define the clinical, language, and cognitive features associated with each of the variants previously identified through analysis of brain metabolism. We also sought to develop a machine-learning algorithm to predict the anatomical subtype of PPA based on the results of language and cognitive testing.

2. MATERIAL AND METHODS

2.1. Study population

From June 2014 to July 2018, we prospectively evaluated a large cohort of patients with PPA with a common protocol including a comprehensive battery of neuropsychological and language tests. All patients met the current diagnostic consensus criteria for PPA (Gorno-Tempini et al., 2011).

For this study, we selected only those patients in early or mild stages, defined as: *1)* independent in the activities of daily living (Functional Activities Questionnaire score < 7), and *2)* scoring 3-5 on the Boston Aphasia Severity Rating Scale. Healthy controls were recruited from among the spouses of patients from our centre's Department of Neurology, and volunteers. In all subjects in the study, Spanish was their mother tongue. We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study. All inclusion and exclusion criteria were determined prior to the analysis.

A total of 68 patients with PPA and 20 healthy controls were included in this study. Fifty-six participants were also included in a previous study (Matias-Guiu et al., 2018), and 32 were new (18 patients and 14 healthy controls). Patients were classified as having non-fluent (n = 23), semantic (n = 9), and logopenic (n = 36) PPA, according to the consensus criteria. Mean age in the PPA group was 72.62 ± 8.00 years; 39 patients (57.40%) were women. Our sample had received a mean of 11.84 ± 4.85 years of formal education. This sample size is considered appropriate for machine learning algorithms.

2.2. FDG PET image acquisition and metabolism-based classification

FDG PET images were acquired following the European guidelines (Varrone et al. 2009), according to the procedure described elsewhere (Matias-Guiu et al., 2018). Images were pre-processed using the Statistical Parametric Mapping software, version 8 (http://www.fil.ion.ucl.ac.uk/spm/), and normalised to a specific FDG PET template for cognitive neurodegenerative disorders (della Rosa et al., 2014). The MarsBaR toolbox was used to conduct a region of interest analysis over the 116 areas of the Automated Anatomical Labeling atlas.

The automated classification methodology was generated by applying machine learning techniques to the dataset acquired. The *WeKa* framework was used to evaluate the performance of several classification techniques, listed in **Table 1**.

We chose several classifiers, which implement different classification methods as Figure 1 shows, in order to find those that perform best. All classifiers used the

previous model as training set and the new dataset as test. For each algorithm, a brief methodological explanation is given below.

BayesNet and Naives were chosen from the Bayesian network algorithms. Bayesian networks consist of 2 stages, the first stage learns the structure of the network, the second one learns the probability tables. Bayes Network algorithm is customized with K2 as the local search algorithm.

Naive Bayes uses the Bayes Theorem for predicting the class a data point belongs to. A new data point is more likely to be assigned to a class with the highest probability.

LBK, Kstar and LWL classifiers are under the lazy algorithms. They apply k-nearest neighbour algorithm to find out values closest to a new data point to make the prediction.

LBK classifier uses the Euclidean distance for classifying a training set. Kstar distributes N data point into k clusters according to the closest mean . The entropy distance is applied for placing new data point to the nearest class. LWL assigns a weighting factor to each data point to indicate the influence on the new data point. Data points nearest the new one obtain higher weights. LinearNN with Euclidean distance is used as the nearest neighbour search algorithm

Decision Table, OneR and Part belong to the rules algorithms. Decision Table shapes decision trees as an ordered and understandable set of If-Then rules to give the prediction. The result is a simpler and less computing-intensive algorithm. OneR is a simple and effective classification algorithm. Each attribute is evaluated and a new branch for each different value is created. Every value generates a new rule. Part algorithm generates a decision list applying separate-and-conquer. A partial C4.5 decision tree is build and for each iteration the best leaf is become into a rule.

Regarding trees algorithms, we chose DecisionStump, J48 and Reptree. DecisionStump is one level decision tree and it applies entropy for classification and the mean-squared error for regression. J48 is an open source implementation of the C4.5 algorithm. J48 builds the tree and places the data points into those

values below a threshold and those above it. Reptree is based on C4.5 algorithm and uses gain and variance information for building regression and decision trees, respectively.

As a first step, we evaluated the capability of the model generated in our previous study (Matias-Guiu et al., 2018) to classify the new patients. Next, we used the set of 32 new participants as the test set and evaluated the classification performance of several classifiers. The best performance was achieved by *LBK*, with a true positive rate of 0.931, and those based on rules, such as *PART*, or decision trees, such as *RepTree* and *J48*, which exhibit high accuracy, with a true positive rate of 0.920. Results of the classification are shown in **Figure 2**. Not all algorithms were tested (only those with better performance were chosen); however, classification results for *ZeroR*, considered as a baseline, were worse with only 24 instances correctly classified.

2.3. Cognitive and language assessment

All participants were assessed with the same protocol. Language assessment comprised the following tests: Cookie Theft picture description from the Boston Diagnostic Aphasia Examination; picture naming task; action naming task; initial phoneme deletion; word spelling; digit span forward and backward; semantic association task; word-picture matching; verb-action picture matching; synonyms judgements; reading of regular words; reading of foreign words; reading of words without stress marks; reading of non-words; verbal repetition (syllables, pairs of syllables, words, pairs of words, non-words, and sentences); category fluency (animals in one minute); action fluency; and letter verbal fluency (words beginning with “p”). Severity of apraxia of speech was graded according to the Apraxia of Speech Severity Rating Scale (Strand et al., 2014). Buccofacial apraxia was also assessed. Some of these tasks were developed by our group (Matias-Guiu et al., 2017), and the other tests came mainly from Boston Diagnostic Aphasia Examination and Test Barcelona (Quintana et al., 2011).

General neuropsychological assessment was performed using the following tests: Mini–Mental State Examination; Addenbrooke's Cognitive Examination (Matias-Guiu et al., 2016); Corsi block-tapping test; Trail Making Test parts A and B; Symbol Digit Modalities Test (SDMT); Rey-Osterrieth Complex Figure Test (copy and recall at 3 and 30 minutes); Visual Object and Space Perception Battery (object decision, progressive silhouettes, position discrimination, and number location subtests), Stroop Color-Word Interference test, and Tower of London (Drexel version). These tests mainly belong to the NEURONORMA battery, the main study providing normative data for neuropsychological assessment in Spain (Peña-Casanova et al., 2009).

2.4. Data analysis

Statistical analysis was performed using the IBM® SPSS Statistics software, version 20.0. Descriptive results are shown as mean ± standard deviation or median (interquartile range). The Kruskall-Wallis test with subsequent pairwise comparison was used to evaluate differences between groups. Groups k6 and k7 (female and male controls, respectively) were combined for statistical analysis. A p-value < 0.05 was considered statistically significant. Pairwise comparisons were considered statistically significant after adjusting for multiple comparisons using the false-discovery rate.

Our automated classification methodology was extended to support classification based on cognitive and language assessments. The analysis of new patients included the application of hierarchical clustering with Ward's method as the clustering algorithm (Everitt et al., 2011). The working dataset contains 2 well defined sets of features. The first set is composed of patients' responses in the cognitive and language tests, while the second includes the PET images. We applied the 8 clusters considered the gold standard in our previous work with the Davies-Bouldin index (Davies et al., 1979), and performed a series of tests, detailed below.

We applied hierarchical clustering to the features belonging to language function, and analysed how instances were assigned to each cluster. Age, years of education, and sex were also included in the analyses. Nevertheless, the number of features in each group in our dataset is high, and not all provide relevant information. We addressed this by applying principal components analysis (PCA) (Jolliffe et al., 2016) to reduce the high dimensionality and extract the set of the most relevant features with minimal loss of accuracy. PCA is a powerful mathematical algorithm for data analysis, and is able to identify patterns in data based on their similarities and differences. Thus, PCA returns a set of principal components as a linear combination of features ranked by relevance.

In this study, PCA was applied to extract the most relevant features in the dataset resulting from the language assessment. Once PCA was performed, we selected the first 5 principal components, which accounted for 66% of the cumulative value. The features selected were saved and used as a new dataset to perform clustering and to compare results with those obtained previously. **Figure 3** shows the clustering results for all features related with language function, and the results obtained after extracting the most relevant features with PCA.

2.5. Brain metabolism analysis

The Statistical Parametric Mapping software was used to conduct a two-sample T-test to compare brain metabolism between each patient group and a group of 40 age- and sex-matched healthy controls. Images were previously normalised to the Montreal Neurological Institute space and smoothed at 12 mm full width at half maximum. Age and sex were added as covariates to the statistical model. A family-wise error–corrected *p*-value of < 0.05 was used as threshold. Only clusters with at least 50 voxels were considered statistically significant.

2.6. Standard protocol approvals, registrations, and patient consents

The study was conducted with the approval of our hospital's Ethics Committee and all participants (or their legally authorized representative) gave written informed consent. . The conditions of our ethics approval do not permit public archiving of anonymized study data. Readers seeking access to the data should contact the corresponding author or the local ethics committee of Hospital Clinico San Carlos, Madrid. Access can be granted only to named individuals in accordance with ethical procedures governing the reuse of sensitive clinical data. No part of the study procedures or analyses was pre-registered in a time-stamped, institutional registry prior to the research being conducted, although decisions regarding design and analysis were decided a priori. The machine learning algorithms used to process the data and obtain the presented conclusions have been obtained from the framework WeKa. This framework can be downloaded from https://www.cs.waikato.ac.nz/ml/weka/ and, after proper configuration, used to execute the mentioned algorithms that included in the tool.

3. RESULTS

3.1. Demographic characteristics according to each subtype

According to the FDG PET cluster-based classification, participants were categorised as: 9 k0 (non-fluent PPA type 1A), 24 k1 (logopenic PPA type 1), 15 k2 (non-fluent PPA type 2), 9 k3 (semantic PPA), 11 k4 (logopenic PPA type 2), and 20 k6/k7 (controls). No patient was classified within the k5 group (non-fluent PPA type 1B). There were no differences between groups regarding age or schooling. The percentage of women was particularly high in the logopenic type 1 group (k1), with a greater proportion of men in the non-fluent type 1 (k0) and logopenic type 2 (k4) groups (**Table 2**).

3.2. Language assessment

Language and speech assessment results are shown in **Table S1 (Supplementary Material)**. **Table S1** and **Figure 3** summarise the specific tests showing statistically significant differences in each variant of PPA in comparison to healthy controls. The decision tree for classification based on the *J48* algorithm is shown in **Figure 4**, which includes Information Gain for each non-terminal node.

Patients with non-fluent PPA type 1 (k0) showed greater impairment in sentence repetition and reading of irregular words in comparison to non-fluent type 2 (k2). Conversely, the latter group showed a higher degree of apraxia of speech. Features of apraxia of speech (Apraxia of Speech Rating Scale score ≥ 1) were present in 6 of the 15 patients (40%) with type 2 (k2). Non-fluent type 1 (k0) differed from both logopenic subtypes only in buccofacial apraxia, which was more severe in non-fluent type 1 (k0).

Patients with logopenic variant type 1 (k1) performed more poorly on action naming than those with logopenic type 2 (k4). A comparison of non-fluent type 1 (k0) and the logopenic variants showed poorer performance on picture naming and sentence repetition in both logopenic subtypes, and also on action naming in logopenic type 1 (k1). Apraxia of speech was present in non-fluent type 2 (k2), but not in the logopenic subtypes.

The comparison between the semantic and other PPA variants identified different behaviour in several tests: results for initial phoneme deletion, word spelling, digit span backward, repetition of pairs of syllables and non-words, and buccofacial apraxia were poorer in non-fluent type 1 (k0); apraxia of speech was more severe and reading of non-words was poorer in non-fluent type 2 (k2), while picture and action naming results were poorer in the semantic variant; in comparison to logopenic type 1 (k1), significant differences were observed in picture naming, word-picture matching, reading of irregular words, repetition of pairs of syllables and sentences, and reading of non-words, all of which were more impaired in the semantic variant. In comparison to logopenic variant type 2 (k4), significant differences were present in picture and action naming, word-picture matching,

synonyms judgement, and reading of irregular words (all more impaired in the semantic variant).

3.3. General cognitive assessment

Results of neuropsychological testing are shown in **Table S2 (Supplementary Material)**; **Figure 5** summarises the statistically significant differences between each variant and the healthy control group. Few tests showed differences between each variant of PPA (**Figure 5**). In this regard, patients with logopenic variant type 1 (k1) showed poorer Mini–Mental State Examination and Rey-Osterrieth Complex Figure scores than those with non-fluent type 2 (k2). No other differences were observed between the non-fluent and logopenic subtypes.

Non-fluent type 1 (k0) was associated with poorer scores on the SDMT and the Visual Object and Space Perception battery (object decision) in comparison to the semantic variant, which scored lower in Stroop Color-Word Interference parts A and B. Patients with non-fluent PPA type 2 (k2) scored worse in SDMT and Stroop Color-Word interference part A but better in Addenbrooke's Cognitive Examination than those with the semantic variant.

Patients with logopenic variant type 2 (k4) showed greater impairment in Trail-Making Test parts A and B, SDMT, and Rey-Osterrieth Complex Figure (copy and recall) in comparison to the semantic variant. In contrast, those with logopenic variant type 2 only scored worse in Trail-Making Test part A and SDMT in comparison to semantic PPA.

3.4. Hierarchical clustering analysis of language assessments

A percentage of agreement of 86.36% was found between language-based and metabolism-based clustering, with a kappa index of 0.832 (95% CI, 0.744-0.920). The percentage of agreement broken down by subtype was: 44% for k0, 92% for k1, 50% for k2, 66.7% for k3, 91% for k4, and 100% for k6/k7. Misclassifications were mainly associated with group k0 (non-fluent type 1). Eight of the 9 patients classified as non-fluent type 2 (k2) and all 6 classified as semantic (k3) according

to language-based clustering were also categorised as such according to FDG PET findings. However, 6 of the patients classified as non-fluent type 2 (k2) and 3 of those classified as semantic (k3) according to brain metabolism findings were classified as k0 with language-based clustering.

3.5. Principal components analysis

PCA of language task data obtained 5 components with the following tests: component 1 included initial phoneme deletion, word spelling, category fluency, non-word repetition, and action fluency; component 2 comprised word-picture matching, verb-action picture matching, the semantic association task, synonyms judgement, and word repetition; component 3 included the picture naming task, Apraxia of Speech Rating Scale score, action naming task, reading of non-words (time), and buccofacial apraxia; component 4 contained sex, reading of words without stress marks (time), reading of words (accuracy and time), and dysarthria; and component 5 consisted of repetition of pairs of words, repetition of non-words, repetition of pairs of syllables, letter verbal fluency, and animals fluency (**Table 3**).

Using this classification model, the kappa index with hierarchical clustering analysis of language tests was 0.958. In comparison to FDG PET–based classification, kappa index was 0.874 (95% CI, 0.79-0.95), with a percentage of agreement of 89.77%. Broken down by PPA subtype, the percentage of agreement was: 53.33% for non-fluent type 1 (k0), 92% for logopenic type 1 (k1), 50% for non-fluent type 2 (k2), 100% for semantic (k3), 91.67% for logopenic type 2 (k4), and 100% for controls (k6/k7). Eight of the 9 patients classified as non-fluent type 2 (k2) according to PCA for language tests were also classified as k2 according brain metabolism findings.

3.6. Brain metabolism

In comparison to the healthy control group, patients with non-fluent PPA type 1 (k0) showed lower metabolism in 3 clusters: a first cluster, with a peak of

significance in the left inferior frontal gyrus with extension to the left superior frontal gyrus, insula, and anterior cingulate; a second cluster of voxels in the left caudate; and a third cluster involving the left middle and medial frontal gyri. Non-fluent PPA type 2 (k2) displayed lower metabolism in a cluster of voxels involving the left superior, middle, medial, and inferior frontal gyri, as well as the left precentral and right superior frontal gyri (**Figure 6**). Logopenic type 1 (k1) showed lower metabolism in the left parieto-temporal lobe, extending to the left frontal lobe and anterior temporal lobe. In turn, logopenic type 2 (k4) showed lower metabolism in the left parieto-temporal lobe extending to the right angular and middle temporal gyri **(Figure 7).** Finally, the semantic variant showed lower metabolism in the anterior temporal lobe bilaterally, especially in the left hemisphere, and some small clusters including the gyrus rectus and anterior cingulate gyrus (**Figure 6)**. Specific statistics regarding voxel-based brain mapping analysis are shown in **Table S3** (Supplementary Material).

4. DISCUSSION

Our study demonstrates the existence of 5 subtypes of PPA in early stages, as we observed in a previous study analysing only neuroimaging data. These variants may be detected in patients in early stages of the disease using an automated algorithm based on regional brain metabolism. These subtypes result from splitting the non-fluent and logopenic variants into 2 additional subtypes, which may be associated with improved prediction of outcome and amyloid biomarkers (Matias-Guiu et al., 2018). Furthermore, from a conceptual perspective, the definition of PPA subtypes according to regional metabolism is supported by the tendency of each neurodegenerative disorder to affect different brain regions and networks (Fu et al., 2018).

One of the most striking results of our research is the separation of 2 clusters within the non-fluent variant. Non-fluent PPA may be diagnosed if patients have agrammatism and/or apraxia of speech. The convenience of dividing the non-fluent variant into 2 subtypes (agrammatic PPA vs apraxia of speech) is an open question in the literature (Josephs et al., 2013; Botha et al., 2015). Our data suggest

that 2 variants can be differentiated according to FDG PET imaging findings. The first variant (k0) showed involvement of the left frontal lobe (Broca's area, but also the anterior cingulate and superior and middle frontal gyri), extending to other regions of the left hemisphere; in contrast, the second variant (k2) also involved the inferior frontal gyrus but tended to affect more medial regions, as well as the right frontal lobe (Matias-Guiu et al., 2018). In this study, however, there was considerable overlap in the brain regions associated with each subtype in comparison to healthy controls. Apraxia of speech was present only in the second type, but affected only 40% of patients. Differences were also observed in performance in sentence repetition and reading of irregular words, which were poorer in non-fluent type 1. Overall, language distinction between these 2 subtypes was incomplete, probably because of substantial overlap between both variants in the impaired brain regions. This supports the need to incorporate neuroimaging biomarkers (such as FDG PET, among others) in this differential diagnosis. No patient was classified as non-fluent type 1B (k5), which is a subtype probably associated with more advanced stages of the disease.

Furthermore, non-fluent type 1 showed some similarities to logopenic subtypes in language impairment. On the one hand, this is consistent with previous research showing some difficulties in discriminating between logopenic variant and other variants of PPA (Sajjadi et al., 2012). On the other hand, a mixed non-fluent/logopenic aphasia phenotype has been associated with progranulin mutations (Rohrer et al., 2012); this fact, as well as the trend toward left hemisphere involvement, suggests that this variant could be associated with TDP-43 proteinopathies. In contrast, non-fluent type 2, in which apraxia of speech is often present and which tends to involve the medial frontal lobes bilaterally, could more probably be associated with tau deposition. Buccofacial apraxia was more frequent in non-fluent type 1 (k0) than in logopenic variants, representing one of the distinctive characteristics of the former. This is consistent with the reported association between buccofacial apraxia and atrophy of the prefrontal cortex (Botha et al., 2014).

The logopenic variant was separated into 2 subtypes, with each displaying clear differences in sex distribution: almost all patients with logopenic type 1 (k1) were women, while all patients with logopenic type 2 (k4) were men. Both types involve the left parieto-temporal junction, but type 1 tends to extend to the left frontal lobe, whereas type 2 involves a more posterior region and the right parieto-temporal lobe (Matias-Guiu et al., 2018). There were no differences in language characteristics, except for action fluency, which was more impaired in type 1; this task has been associated with left frontal lobe function (Beber and Chaves, 2014). The existence of some differences in regional metabolism and the differing representation of each sex by subtype contributes to our understanding of differences between men and women in regional and genetic vulnerability to Alzheimer pathology, as has recently been suggested (Liesinger et al., 2018; Deming et al., 2018).

Interestingly, language tests were able to predict with a relatively high degree of accuracy the FDG PET–based variant of PPA, according to the different algorithms used in this study. This is especially relevant given the challenge of differentiating between PPA variants in previous studies (Sajjadi et al., 2012; Harris et al., 2018). We obtained 5 language test components using PCA analysis. Component 1 included several tests mainly associated with phonology; component 2 was particularly related to semantics; component 3 to various tasks (including picture naming, apraxia of speech, action fluency, reading of non-words [time], and buccofacial praxis); component 4 to several reading tasks; and component 5 comprised several repetition and fluency tasks. These results suggest that language performance in PPA is described by multiple components, including phonology, semantics, motor production, reading, repetition, and verbal fluency. Thus, language assessment needs to be comprehensive, and machine learning approaches are probably necessary for optimised classification of patients. Conversely, differences between subtypes according to cognitive tests examining other neuropsychological functions were minimal, which may be explained by the fact that we only included patients in early stages of the disease.

Our study does have some limitations. Firstly, our results are based on the battery of language and cognitive assessments used in this study. Although we included a comprehensive battery analysing the main aspects of language and general cognition needed for PPA diagnosis, we cannot discount the possibility that a deeper study of some aspects (for instance, grammar or social cognition) might improve the discrimination between each subtype. Similarly, connected speech was not included in the analysis, which could have improved the classification of patients with PPA (Boschi et al. 2017; Fraser et al. 2014). Second, our study was performed in Spanish speakers. Spanish is less complex than English in such aspects as motor speech (with fewer consonants, vowels, and consonant clusters, and a majority of words consisting of 2 syllables) and reading (Matias-Guiu et al., 2017). This may hypothetically produce some differences in language characteristics, which may be less evident in language examination. However, we used specific tasks developed in Spanish to reduce the effect of this limitation. Third, our study is based on the existence of several separated variants of PPA in the early stages. Because of the probable existence of some cases overlapping two or more variants (clinically or regarding regions of brain atrophy or hypometabolism), a multi-functional linguistic classification could be of interest in future works. This may be especially interesting when considering cases with PPA in more advanced stages, where overlapping in language features of some variants is more frequent.

5. CONCLUSIONS

In conclusion, our study supports the existence of 5 variants of PPA based on regional metabolism. These variants show some differences in terms of language characteristics. The application of machine learning algorithms to language tests enabled relatively accurate prediction of each of the 5 PPA variants; this gives rise to the possibility of automated, machine-aided diagnosis of PPA variants. Longitudinal data and pathological confirmation are necessary to further validate this categorisation of PPA.

6. ACKNOWLEDGEMENTS

We acknowledge support from the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (ERDF) under project TIN2017-85727-C4-4-P and projects IB16035, of the Regional Government of Extremadura's Department of Commerce and Economy, conceded by the European Regional 282 Development Fund, "A way to build Europe."

7. REFERENCES

-Beber BC, Chaves MLF (2014). The basis and applications of the action fluency and action naming tasks. *Dementia e Neuropsychologia* 8:47-57.

-Boschi V, Catricalà E, Consonni M, Chesi C, Moro A, Cappa SF (2017). Connected speech in neurodegenerative language disorders: a review. *Frontiers in Psychology* 8:269.

-Botha H, Duffy JR, Strand EA, Machulda MM, Whitwell JL, Josephs KA (2014). Nonverbal oral apraxia in primary progressive aphasia and apraxia of speech. *Neurology* 82:1729-1735.

-Botha H, Duffy JR, Whitwell JL, Strand EA, Machulda MM, Schwarz CG, Reid RI, Spychalla AJ, Senjem ML, Jones DT, Lowe V, Jack CR, Josephs KA (2015). Classification and clinicoradiologic features of primary progressive aphasia (PPA) and apraxia of speech. *Cortex* 69:220-236.

-Davies DL, Bouldin W (1979). A cluster separation measure. *IEEE Trans Pattern Anal Mach Intell* 1:224-227.

- Della Rosa PA, Cerami C, Gallivanone F, Prestia A, Caroli A, Castiglioni I, Gilardi MC, Frisoni G, Friston K, Ashburner J, Perani D, the EADC-PET Consortium (2014). A standardized [18F]-FDG-PET template for spatial normalization in statistical parametric mapping of dementia. *Neuroinformatics* 12:575-593.

-Deming Y, Dumitrescu L, Barnes LL, Thambisetty M, Kunkle B, Gifford KA, et al (2018). Sex-specific genetic predictors of Alzheimer's disease biomarkers. *Acta Neuropathologica* 2018. doi:10.1007/s00401-018-1881-4.

-Everitt BS, Landau S, Leese M, Stahl D. Hierarchical clustering, in Cluster Analysis, 5th Ed. Chichester 2011: John Wiley & Son, Ltd.

-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.

-Fu H, Hardy J, Duff KE (2018). Selective vulnerability in neurodegenerative diseases. *Nature Neuroscience* 21:1350-1358.

- 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, Vandenbergue 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.

-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.

-Harris JM, Saxon JA, Jones M, Snowden JS, Thompson JC (2018). Neuropsychological differentiation of progressive aphasic disorders. *Journal of Neuropsychology*. doi:10.111/jnp.12149.

-Hoffman P, Sajjadi SA, Patterson K, Nestor PJ (2017). Data-driven classification of patients with primary progressive aphasia. *Brain & Language* 174:86-93.

-Jolliffe IT, J Cadima J (2016). Principal component analysis: a review and recent developments. *Phil Trans a Math Phys Eng Sci* 374:20150202.

-Josephs KA, Duffy JR, Strand EA, Machulda MM, Senjem ML, Master AV, Lowe VJ, Jack CR Jr, Whitwell JL (2012). Characterizing a neurodegenerative syndrome: primary progressive apraxia of speech. *Brain* 135:1522-1536.

-Josephs KA, Duffy JR, Strand EA, Machulda MM, Senjem ML, Lowe VJ, Jack CR Jr, Whitwell JL (2013). Syndromes dominated by apraxia of speech show distinct characteristics from agrammatic PPA. *Neurology* 81:337-345.

-Josephs KA, Duffy JR, Strand EA, Machulda MM, Vemuri P, Senjem ML, Perkerson RB, Baker MC, Lowe V, Jack CR Jr, Rademakers R, Whitwell JL (2014). Progranulin-associated PiB-negative logopenic progressive aphasia. *Journal of Neurology* 261:604-614.

-Krishnan K, Machulda MM, Whitwell JL, Butts AM, Duffy JR, Strand EA, Senjem ML, Spychalla AJ, Jack CR, Lowe VJ, Josephs KA (2017). Varying degrees of temporoparietal hypometabolism on FDG-PET reveal amyloid-positive logopenic

primary progressive aphasia is not a homogeneous clinical entity. *Journal of Alzheimers Disease* 55:1019-1029.

-Leyton CE, Villemagne VL, Savage S, Pike KE, Ballard KJ, Piguet O, Burrell JR, Rowe CC, Hodges JR (2011). Subtypes of progressive aphasia: application of the International Consensus Criteria and validation using beta-amyloid imaging. *Brain* 134:3030-3043.

-Leyton C, Hodges JR, McLean CA, Kril JJ, Piguet O, Ballard KJ (2015). Is the logopenic-variant of primary progressive aphasia a unitary disorder? *Cortex* 67:122-133.

-Liesinger AM, Graff-Radford NR, Duara R, Carter RE, Al-Shaikh FSH, Koga S, Hinkle KM, DiLello SK, Johnson MF, Aziz A, Ertekin-Taner N, Ross OA, Dickson DW, Murray ME (2018). Sex and age interact to determine clinicopathologic differences in Alzheimer's disease. *Acta Neuropathologica.* doi:10.1007/s00401-018-1908-x.

-Marshall CR, Hardy CJD, Volkmer A, Russell LL, Bond RL, Fletcher PD, Clark CN, Mummery CJ, Schott JM, Rossor MN, Fox NC, Crutch SJ, Rohrer JD, Warren JD (2018). Primary progressive aphasia: a clinical approach. *Journal of Neurology*, 265:1474-1490.

- Matías-Guiu JA & García-Ramos R (2013). Primary progressive aphasia: from syndrome to disease. *Neurologia* 28:366-374.

- Matías-Guiu JA, Cabrera-Martín MN, García-Ramos R, Moreno-Ramos T, Valles-Salgado M, Carreras JL, Matías-Guiu J (2014). Evaluation of the new consensus criteria for the diagnosis of primary progressive aphasia using fluorodeoxyglucose positrón emission tomography. *Dementia and Geriatric Cognitive Disorders* 38:147-152.

-Matias-Guiu JA, Cabrera-Martín MN, Moreno-Ramos T, Valles-Salgado M, Fernández-Matarrubia M, Carreras JL, Matías-Guiu J (2015). Amyloid and FDG-PET study of logopenic primary progressive aphasia: evidence for the existence of two subtypes. *Journal of Neurology* 262:1463-1472.

-Matias-Guiu JA, Fernández-Bobadilla R, Fernández-Oliveira A, Valles-Salgado M, Rognoni T, Cortés-Martínez A, Moreno-Ramos T, Kulisevsky J, Matías-Guiu J (2016). Normative data for the Spanish version of the Addenbrooke's Cognitive Examination III. *Dementia and Geriatric Cognitive Disorders* 41:243-250.

-Matias-Guiu JA, Cuetos F, Cabrera-Martín MN, Valles-Salgado M, Moreno-Ramos T, Carreras JL, Matías-Guiu J (2017). Reading difficulties in primary progressive aphasia in a regular language-speaking cohort of patients. *Neuropsychologia* 101:132-140.

-Matias-Guiu JA, Díaz-Álvarez J, Ayala JL, Risco-Martín JL, Moreno-Ramos T, Pytel V, Matías-Guiu J, Carreras JL, Cabrera-Martín MN (2018). Clustering analysis of FDG-PET imaging in primary progressive aphasia. *Frontiers in Aging Neuroscience* 10:230.

-Mesulam M (2001). Primary progressive aphasia. *Annals of Neurology* 49:425-432.

-Neary D, Snowden J, Gustafson L, Passant U, Stuss D, Black S, Freedman M, Kertesz A, Robert PH, Albert M, Boone K, Miller BL, Cummings J, Benson DF (1998). Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* 51:1546-1554.

- Peña-Casanova J, Blesa R, Aguilar M, Gramunt-Fombuena N, Gómez-Ansón B, Oliva R et al., for the Neuronorma Study Team (2009). Multicenter Normative Studies (NEURONORMA Project: methods and sample characteristics. *Archives of Clinical Neuropsychology* 24:307-319.

-Quintana M, Peña-Casanova J, Sánchez-Benavides G, Langohr K, Manero RM, Aguilar M, et al; Neuronorma Study Team (2011). Spanish multicenter normative studies (Neuronorma project): norms for the abbreviated Barcelona test. *Archives of Clinical Neuropsychology* 26:144-157.

-Rohrer JD, Crutch SJ, Warrington EK, Warren JD (2010). Progranulin-associated primary progressive aphasia: a distinct phenotype? *Neuropsychologia* 48:288-297.

-Sajjadi SA, Patterson K, Arnold RJ, Watson PC, Nestor PJ (2012). Primary progressive aphasia: a tale of two syndromes and the rest. *Neurology* 78:1670-1677.

-Sajjadi SA, Patterson K, Nestor PJ (2014). Logopenic, mixed, or Alzheimer-related aphasia? *Neurology* 82:1127-1131.

-Spinelli EG, Mandelli ML, Santos-Santos MA, Wilson SM, Agosta F, Grinberg LT, Huang EJ, Trojanowski JQ, Meyer M, Henry ML, Comi G, Rabinovici G, Rosen HJ, Filippi M, Miller BL, Seeley WW, Gorno-Tempini ML (2017). Typical and atypical pathology in primary progressive aphasia variants. *Annals of Neurology* 81:430-443.

-Strand E, Duffy JR, Clark HM, Josephs KA (2014). The apraxia of speech rating scale: a tool for diagnosis and description of apraxia of speech. *Journal of Communication Disorders* 51:43-50.

- Varrone A, Asenbaum S, Vander Borght T, Booij J, Nobili F, Nagren K, et al.; European Association of Nuclear Medicine Neuroimaging Committee (2009). EANM procedure guidelines for PET brain imaging using [18F]]FDG, version 2. *European Journal of Nuclear Medicine and Molecular Imaging* 36:2103-2110.

Figure and Table legends.

Figure 1. Classification methods used.

Figure 2. Accuracy of each classifier. PET features – classification. Previous model as training set and new set as test.

Figure 3. Language tasks showing statistically significant differences between PPA subtypes. Specific tests showing differences between two subtypes are shown over the line connecting those subtypes. For instance, between K0 (non-fluent type 1) and K2 (non-fluent type 2): repetition of sentences, reading irregular words, and apraxia of speech; or between K0 (non-fluent type 1) and k4 (logopenic type 2): buccofacial apraxia (in orange, as the line connecting K0 and K4).

Figure 4. *J48* decision tree algorithm for classifying PPA subtypes using language tasks.

Figure 5. Cognitive tests showing statistically significant differences between PPA subtypes. Specific tests showing differences between two subtypes are shown over the line connecting those subtypes. For instance, between K0 (non-fluent type 1) and k4 (logopenic type 2): TMT-A, TMT-B, SDMT, ROCF, and ROCF memory (in orange, as the line connecting K0 and K4).

Figure 6. Statistical parametric map overlaid onto an MRI template (axial view), showing the brain regions with lower metabolism in non-fluent PPA type 1 (k0, in red), non-fluent PPA type 2 (k2, in yellow), and semantic PPA (k3, in cyan).

Figure 7. Statistical parametric map overlaid onto an MRI template (axial view), showing the brain regions with lower metabolism in logopenic PPA type 1 (k1, in blue) and logopenic PPA type 2 (k4, in green).

Table 1. Machine learning classification techniques.

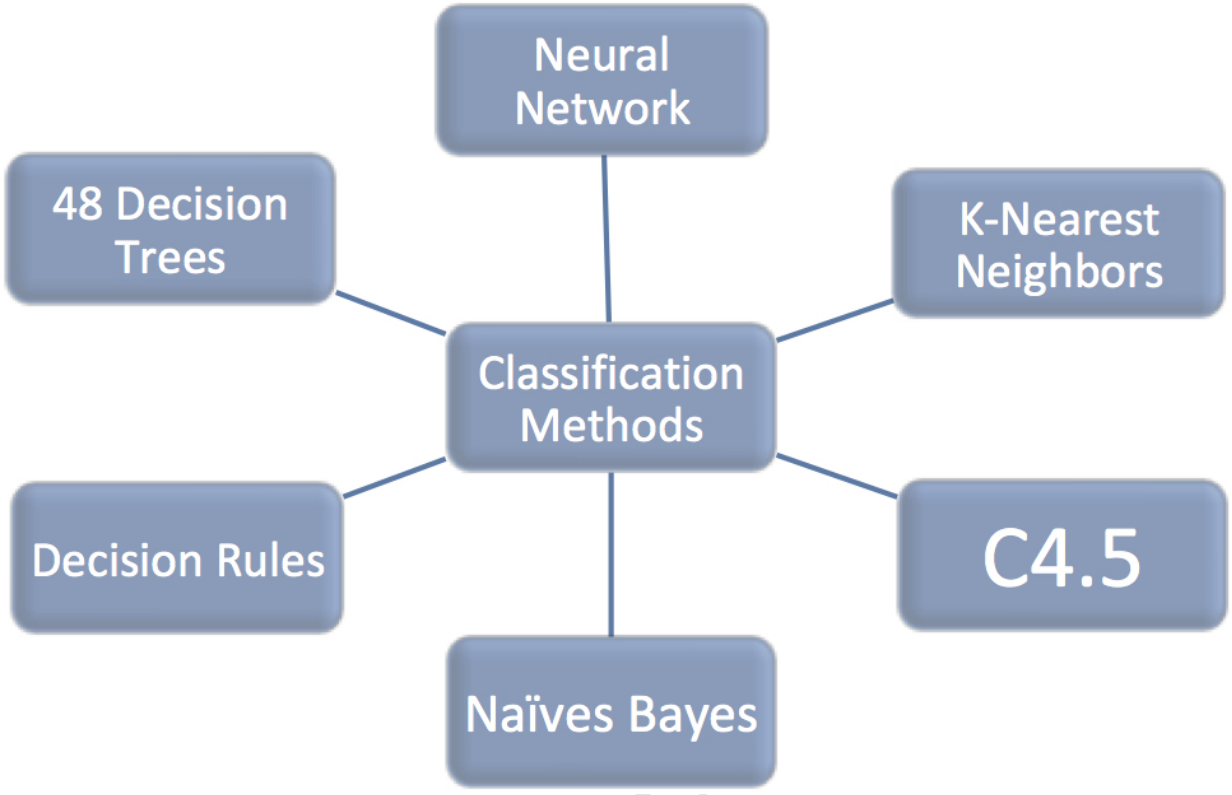
Supplementary Material

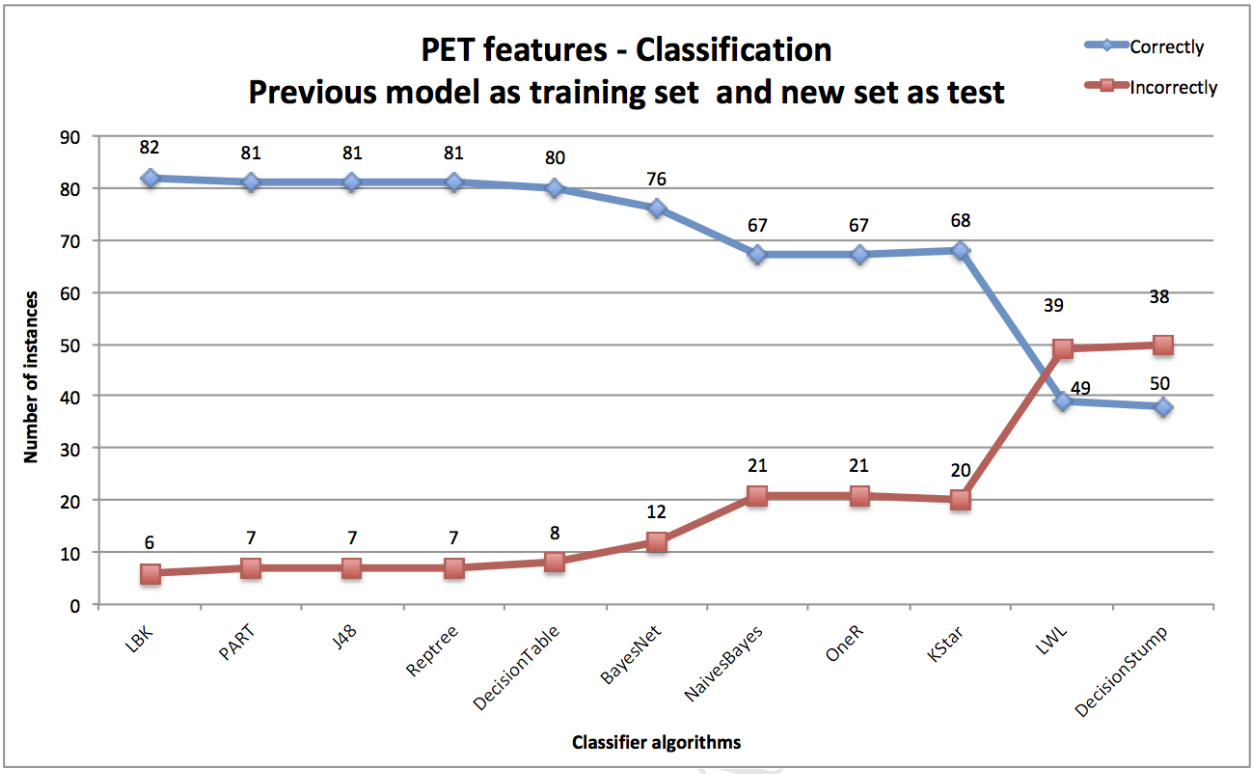
Table 1. Machine learning classification techniques.			
Bayes	**Lazy**	**Rules**	**Trees**
BayesNet	LBK	OneR	J48 (pruned, unpruned)
NaiveBayes	Kstar	PART	Reptree
	LWL	DecisionTable	DecisionStump

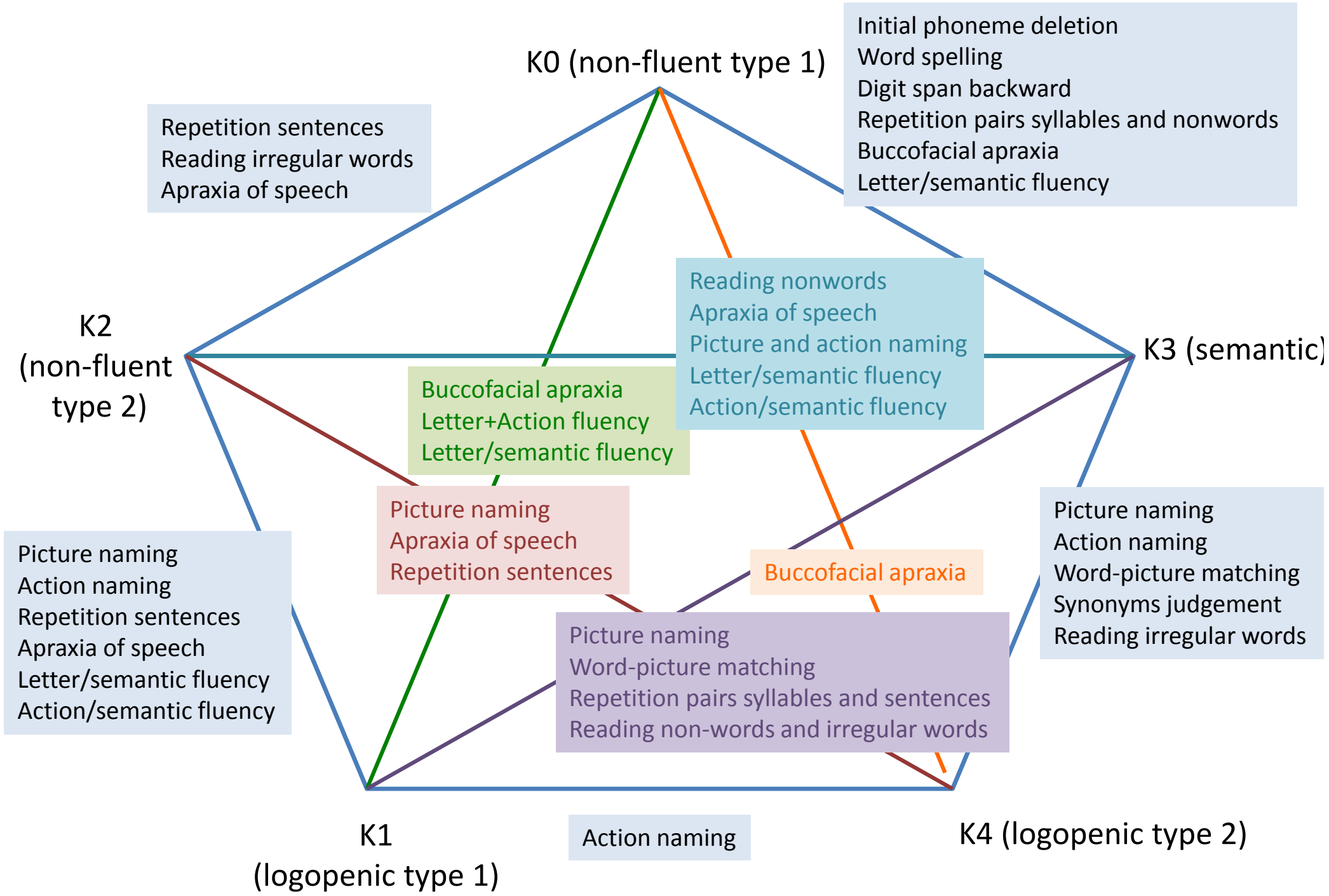
Table 2.
Demographic characteristics between groups.
NF-1: non-fluent type 1
L-1: logopenic type 1
NF: non-fluent type 2
S: semantic
L-2: logopenic type 2
HC: healthy controls.

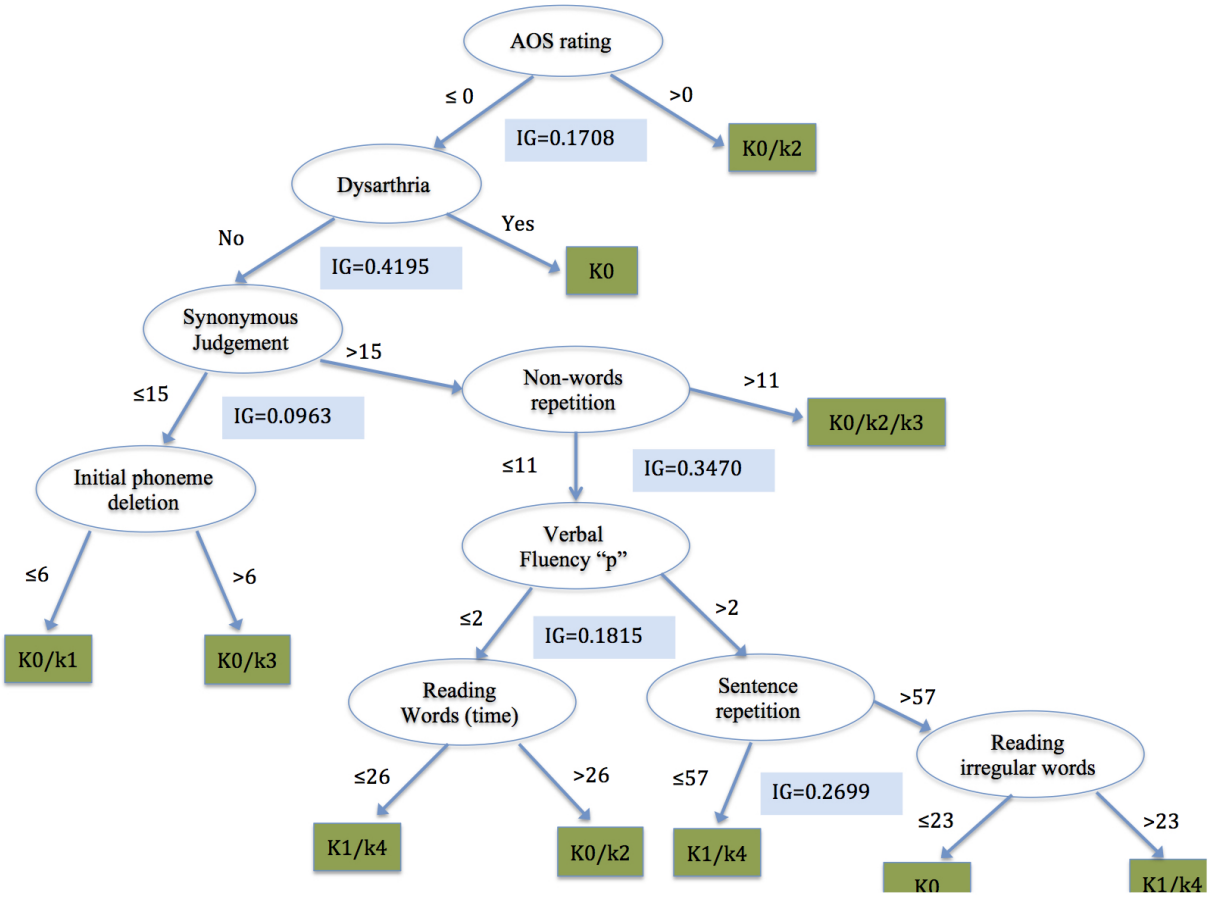
Groups	K0 (NF-1)	K1 (L-1)	K2 (NF-2)	K3 (S)	K4 (L-2)	K6/K7 (HC)	p-value
Number of patients	9	24	15	9	11	20	-
Age	73.11±8.55	73.79±6.87	70.53±7.67	68.44±11.60	75.82±6.33	68.05±6.75	0.112
Gender (females) n (%)	1 (11.1%)	23 (95.8%)	9 (60%)	6 (66.7%)	0 (0%)	16 (80%)	0.000
Years of education	8.89±4.22	11.38±4.81	11.87±5.50	12.44±3.74	14.73±4.33	13.00±4.60	0.146
FAQ	3.67±2.59	1.46±2.28	1.80±2.85	4.0±2.78	2.18±3.15	0±0	

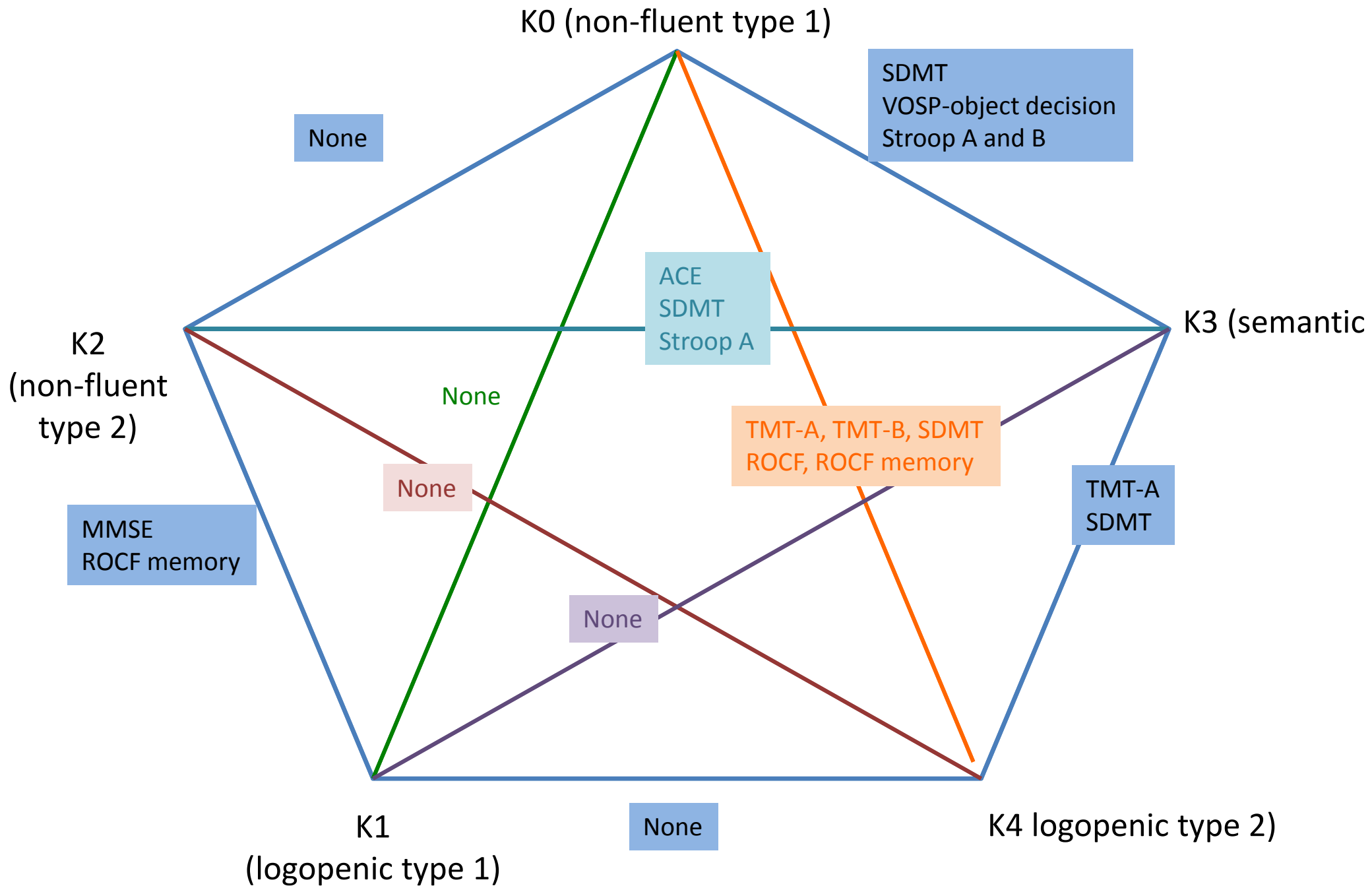
Table 3. PCA analysis.	
Component 1	–0.242 initial phoneme deletion –0.226 word spelling –0.222 category fluency –0.217 non-words repetition –0.213 action fluency
Component 2	–0.458 word-picture matching –0.396 verb-action picture matching –0.386 semantic association task –0.313 synonyms judgement –0.304 word repetition
Component 3	–0.371 picture naming task –0.341 Apraxia of Speech Rating Scale score –0.334 action naming task –0.307 reading of non-words (time) +0.273 buccofacial apraxia
Component 4	+0.342 sex (women = 2) +0.291 reading of words without stress mark (time) –0.271 reading of words (accuracy) +0.267 reading of words (time) –0.22 dysarthria
Component 5	–0.371 repetition of pairs of words –0.31 repetition of non-words –0.307 repetition of pairs of syllables +0.278 letter verbal fluency +0.267 animals fluency

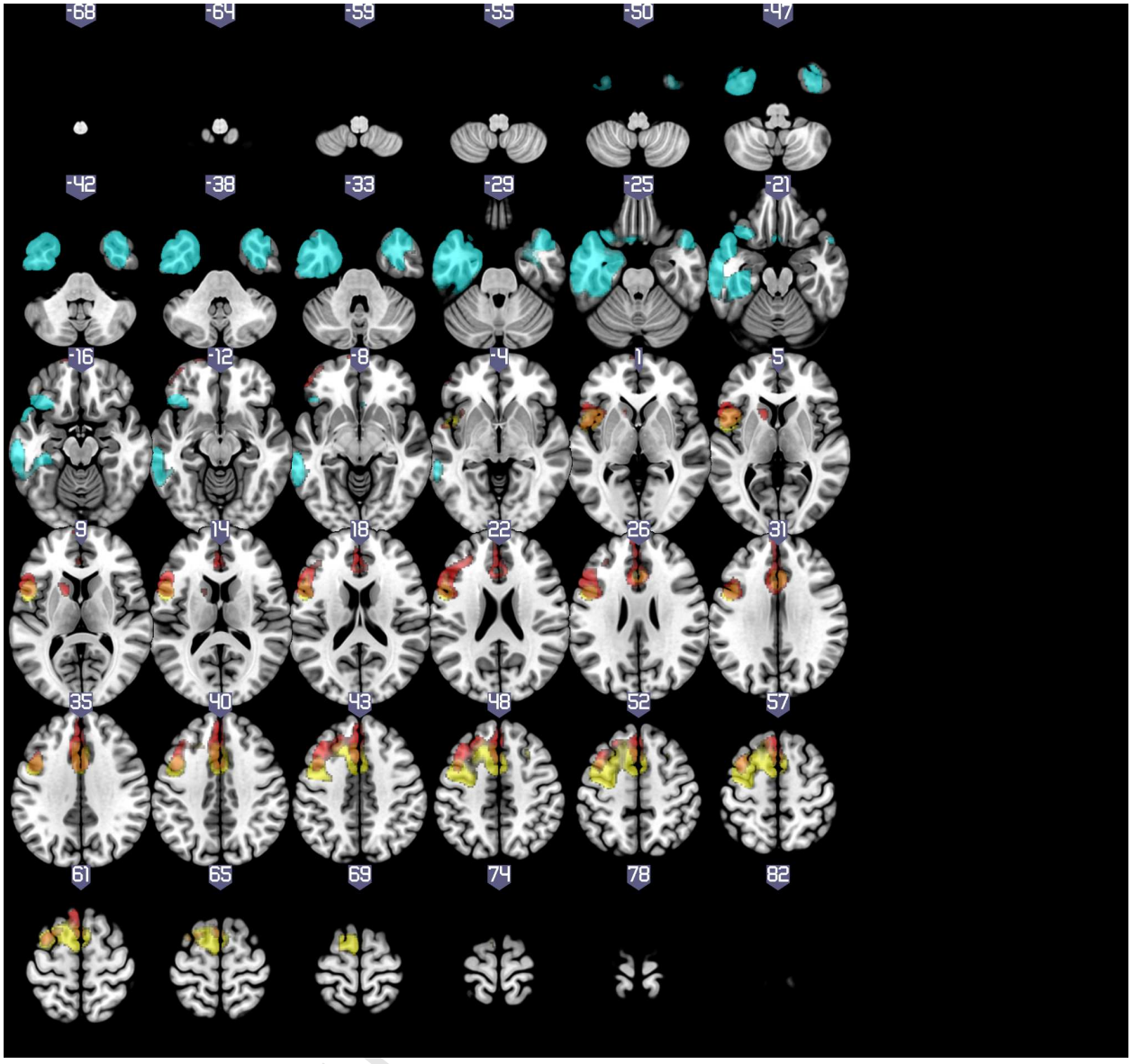


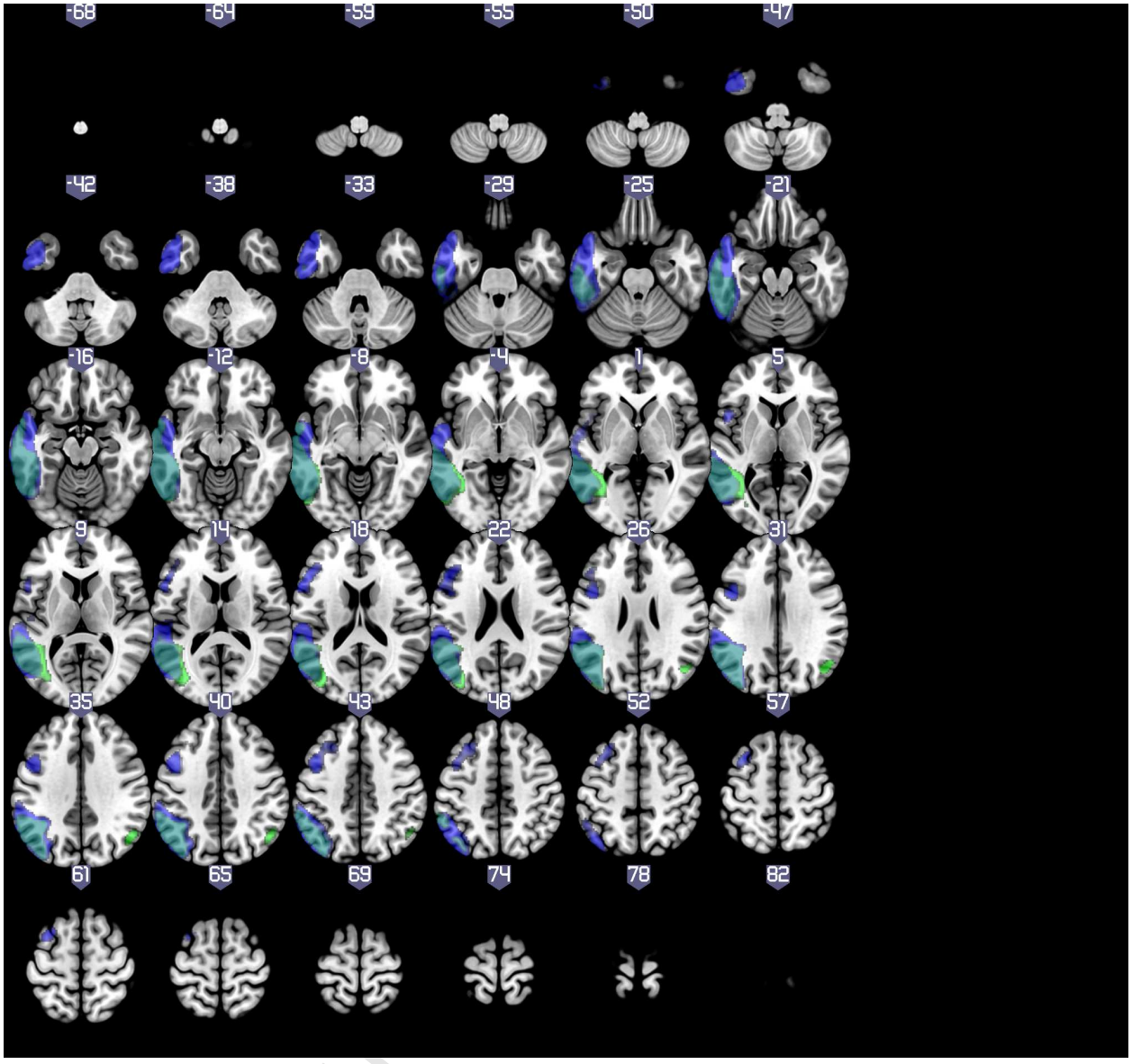



K0 (non-fluent type 1)
Initial phoneme deletion
Word spelling
Digit span backward
Repetition pairs syllables and nonwords
Buccofacial apraxia
Letter/semantic fluency
Repetition sentences
Reading irregular words
Apraxia of speech
K2
(non-fluent
type 2)
K3 (semantic)
Reading nonwords
Apraxia of speech
Picture and action naming
Letter/semantic fluency
Action/semantic fluency
Buccofacial apraxia
Letter+Action fluency
Letter/semantic fluency
Picture naming
Apraxia of speech
Repetition sentences
Picture naming
Action naming
Repetition sentences
Apraxia of speech
Letter/semantic fluency
Action/semantic fluency
Buccofacial apraxia
Picture naming
Action naming
Word-picture matching
Synonyms judgement
Reading irregular words
Picture naming
Word-picture matching
Repetition pairs syllables and sentences
Reading non-words and irregular words
K1
(logopenic type 1)
Action naming
K4 (logopenic type 2)




K0 (non-fluent type 1)
SDMT
VOSP-object decision
Stroop A and B
None
ACE
SDMT
Stroop A
K3 (semantic
K2
(non-fluent
type 2)
None
TMT-A, TMT-B, SDMT
ROCF, ROCF memory
None
TMT-A
SDMT
MMSE
ROCF memory
None
K1
(logopenic type 1)
None
K4 logopenic type 2)


-68
-64
-59
-55
-50
-47
-42
-38
-33
-29
-25
-21
-16
-12
-8
-4
1
5
9
14
18
22
26
31
35
40
43
48
52
57
61
65
69
74
78
82


-68
-64
-59
-55
-50
-47
-42
-38
-33
-29
-25
-21
-16
-12
-8
-4
1
5
9
14
18
22
26
31
35
40
43
48
52
57
61
65
69
74
78
82

CRediT author statement:

-Jordi A Matias-Guiu: conceptualization; investigation; formal analysis; writing original draft;

-Josefa Díaz-Álvarez: methodology; formal analysis; writing original draft;

-Fernando Cuetos: writing original draft review and editing; resources.

-María Nieves Cabrera-Martín: formal analysis; visualization; resources; investigation.

-Ignacio Segovia-Ríos: formal analysis; visualization.

-Vanesa Pytel: data acquisition; resources; investigation.

-Teresa Moreno-Ramos: data acquisition; resources

-José Luis Carreras: supervision; resources

-Jorge Matías-Guiu; writing original draft review and editing; supervision.

-José Luis Ayala: conceptualization; formal analysis; writing original draft.