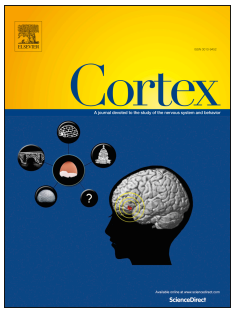
Accepted Manuscript



The language profile of progressive supranuclear palsy

E. Catricalà, V. Boschi, S. Cuoco, F. Galiano, M. Picillo, E. Gobbi, A. Miozzo, C. Chesi, V. Esposito, G. Santangelo, M.T. Pellecchia, V.M. Borsa, P. Barone, P. Garrard, S. Iannaccone, S.F. Cappa

PII: S0010-9452(19)30065-6

DOI: https://doi.org/10.1016/j.cortex.2019.02.013

Reference: CORTEX 2570

To appear in: *Cortex*

Received Date: 3 June 2018

Revised Date: 12 December 2018

Accepted Date: 14 February 2019

Please cite this article as: Catricalà E, Boschi V, Cuoco S, Galiano F, Picillo M, Gobbi E, Miozzo A, Chesi C, Esposito V, Santangelo G, Pellecchia MT, Borsa VM, Barone P, Garrard P, Iannaccone S, Cappa SF, The language profile of progressive supranuclear palsy, *CORTEX*, https://doi.org/10.1016/j.cortex.2019.02.013.

The language profile of progressive supranuclear palsy

Catricalà E. [1]*, Boschi V. [2]*, Cuoco S.[3], Galiano F.[4], Picillo M. [3], Gobbi E.[5], Miozzo A.[6], Chesi C.[1], Esposito V.[4,7], Santangelo G.[3,8], Pellecchia M.T.[3], Borsa V.M. [1,9], Barone P.[3], Garrard P. [10], Iannaccone S.[7], Cappa S.F.[1,5]

[1]NEtS Center, School of Advanced Studies IUSS Pavia, Pavia, Italy;

[2]Accademia della Crusca, Florence, Italy;

[3]Department of Medicine, Surgery, and Dentistry "Scuola Medica Salernitana", Neuroscience Section, University of Salerno, Italy;

[4]Vita-Salute San Raffaele University, Milan, Italy;

[5]IRCCS San Giovanni di Dio Fatebenefratelli, Brescia, Italy;

[6]Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy;

[7]Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Milan, Italy;

[8]Department of Psychology, University of Campania Luigi Vanvitelli, Caserta, Italy;

[9] NEUROFARBA – Dipartimento di Neuroscienze, Psicologia, Area del Farmaco e Salute del Bambino, Università di Firenze, Florence, Italy;

[10]Neuroscience Research Centre, St George's—University of London, London, UK.

*=equal contribution

Corresponding author:

Prof. Stefano F. Cappa

Institute for Advanced Study-IUSS Pavia, Palazzo del Broletto, Piazza Vittoria 15, 27100 Pavia, Italy.

e-mail address: stefano.cappa@iusspavia.it

Abstract

A progressive speech/language disorder, such as the non fluent/agrammatic variant of primary progressive aphasia and progressive apraxia of speech, can be due to neuropathologically verified Progressive Supranuclear Palsy (PSP). The prevalence of linguistic deficits and the linguistic profile in PSP patients who present primarily with a movement disorder is unknown. In the present study, we investigated speech and language performance in a sample of clinically diagnosed PSP patients using a comprehensive language battery, including, besides traditional language tests, a detailed analysis of connected speech (picture description task assessing 26 linguistic features). The aim was to identify the most affected linguistic levels in seventeen PSP with a movement disorder presentation, compared to 21 patients with Parkinson's disease and 27 healthy controls. Machine learning methods were used to detect the most relevant language tests and linguistic features characterizing the language profile of PSP patients. Our results indicate that even non-clinically aphasic PSP patients have subtle language deficits, in particular involving the lexical-semantic and discourse levels. Patients with the Richardson's syndrome showed a lower performance in the word comprehension task with respect to the other PSP phenotypes with predominant frontal presentation, parkinsonism and progressive gait freezing. The present findings support the usefulness of a detailed language assessment in all patients in the PSP spectrum.

Keywords: language; progressive supranuclear palsy; connected speech; machine learning; Richardson's syndrome.

1. Introduction

In addition to motor symptoms, patients with Progressive Supranuclear Palsy (PSP) commonly show behavioural and cognitive disorders (Lee et al., 2016; Monza et al., 1998; Robbins et al., 1994; Soliveri et al., 2000; for a review see Burrell et al., 2014). The recent Movement Disorder Society (MDS) criteria (Hoglinger et al., 2017) include cognitive dysfunction as one of the core diagnostic criteria, together with oculomotor dysfunction, postural instability and akinesia. Different variants have been identified, including, in addition to the classical Richardson's syndrome, initial predominance of ocular motor dysfunction, postural instability, Parkinsonism resembling idiopathic Parkinson's disease, frontal lobe cognitive or behavioral presentations, including behavioral variant frontotemporal dementia, progressive gait freezing, corticobasal syndrome, primary lateral sclerosis, cerebellar ataxia, and speech/language disorders, including nonfluent/agrammatic primary progressive aphasia (PNFA) and progressive apraxia of speech (AOS). The presence of a speech/language disorder is considered as a core clinical feature, with the highest level of diagnostic certainty in the cognitive domain (C1).

The neurolinguistic features of PSP patients presenting with an aphasic phenotype have been extensively described (Perkin et al., 1978; Esmonde et al., 1996; Della Sala & Spinnler, 1998; Kertesz and McMonagle, 2010; Josephs et al., 2005; Robinson et al., 2006; Boeve et al., 2003; Joseph et al., 2006; Roher et al., 2010). Patients with PSP-PNFA when compared with controls show an impairment in repetition, naming, semantic and phonemic fluency, single word and sentence comprehension and non-word reading (Rohrer et al., 2010; Santos et al., 2016). Connected speech in picture description is characterised by phonetic errors, reduced speech rate (Santos et al., 2016; Rohrer et al., 2016), low mean length of utterance (Santos et al., 2016), and an increase in syntactic errors (Santos et al., 2016; Rohrer et al., 2016), indicating an impairment at the phonetic and syntactic levels. When PSP-PNFA is compared with non-PSP-PNFA, a more severe reduction in spontaneous speech, with fewer speech errors and less marked impairment of literacy skills were reported in the former (Rohrer et al., 2010).

Mild language disorders have been reported also in patients with the classical Richardson's syndrome. The first studies (Maher et al, 1985; Podoll et al., 1991; for a review, see Kim and McCann, 2015) interpreted language impairment as 'secondary to other neurological and neuropsychological disturbances' (Podoll et al., 1991). Defective performance in lexico-semantic tasks has been successively reported by several studies (Daniele et al., 2013; Bak et al., 2006), in

particular for action-verbs (vs. nouns) (Bak et al., 2006; Daniele, 1994; Daniele et al., 2013; Cotelli et al., 2006). More recently, an attempt to better characterize the language profile reported impairments in naming, word comprehension, semantic association and syntactic comprehension (Burrell et al., 2018). When compared to PNFA, a comparable impairment in naming, word comprehension, semantic association and syntactic comprehension tasks has been reported (Burrell et al., 2018).

In the present study, we analyse the language impairment in a sample of patients recruited from movement disorders clinics, in order to assess the prevalence of linguistic deficits and to characterize the linguistic profile in PSP patients with a primarily non-cognitive presentation. To this aim, we used a comprehensive language battery, including an analysis of a sample of connected speech obtained through a picture description task, to identify the most affected linguistic level in PSP patients presenting with different phenotypes, excluding only patients with an aphasic/speech apraxia variant. Using machine-learning algorithms, we aimed at capturing the language tests and linguistic features best describing the linguistic profile of the disorder when compared to healthy subjects. As a second aim, in order to assess the specificity of language impairment in PSP with respect to other movement disorders, we compared the performance of PSP patients with a sample of patients with Parkinson disease (PD).

2. Materials and methods

2.1 Participants

From 2016 to 2017, fifty-nine patients were consecutively included in the study: 32 patients with a clinical diagnosis progressive supranuclear palsy (PSP) and 27 with Parkinson's disease (PD). Participants were diagnosed by experienced movement disorder neurologists, in accordance with the MDS criteria (Hoglinger et al., 2017) for PSP patients and according to the UK Brain Bank Criteria (Gibb and Lee, 1988) for PD patients. Patients were recruited at the Movement Disorders Center of the University of Salerno, and at the Neuroscience Division of the San Raffaele's Hospital, Milano. All participants were Italian native speakers. All patients underwent to a comprehensive neuropsychological and language assessment (see below for details). Twenty-one participants were excluded from the study, as 7 PSP and 1 PD had unintelligible speech and 3 PSP and 1 PD refused to complete the picture description task. Additional inclusion criteria were: a corrected score on Mini Mental State Examination (MMSE) of at least 18, leading to the exclusion of 4 PSP, and a disease duration lesser than 10 years, leading to the exclusion of 4 PD. Finally, we

excluded a PSP patient fulfilling the criteria for a non fluent/agrammatic PPA phenotype (Gorno Tempini et al., 2011). The patient had been referred to the movement disorder outpatient clinic, but showed a prominent speech and language impairment, assessed through a comprehensive language evaluation, and characterized by a severe AOS and agrammatism in production, and by a deficit in sentence comprehension but not in single word comprehension and semantic association tasks. The demographic information for the final sample of thirty-eight patients, 21 PD and 17 PSP (including 10 presenting with Richardson' syndrome, 4 with predominant frontal presentation, 2 with predominant parkinsonism and 1 predominant progressive gait freezing), are summarized in Table 1. Disease duration was calculated in years from the onset of the first motor symptoms. Disease severity was assessed with the PSP Rating Scale (PSPRS) (Payan et al., 2011) and the Unified Parkinson Disease Rating Scale (UPDRS, part III motor subscale) (Fahn and Elton, 1987).

The control sample included 27 healthy controls (15 males) matched for age (p=.92) and education (p=.8) with the 38 patients. None of the controls had a history of neurological illness or mental decline, and all had an adjusted score on the MMSE of at least 24.

The 3 groups (PD, PSP and controls) were not matched for age (p=.056) and education (p=.028). Post-hoc comparisons revealed differences only between the PD and PSP groups. Patients in the PSP group were significantly older than those in the PD group (p=.05) and had fewer years of formal education (p<.05). Both groups of patients were matched with controls for both age (p at least=.551) and education (p a least=.343).

The study was approved by the local research ethics committees.

Table 1: Demographic data of PSP, PD and controls.

Demographics	**PSP (n = 17)**	**PD (n = 21)**	**Controls (n = 27)**
Gender (M/F)	8/9	15/6	15/12
Age (years)	71,29 (63-84)	64,53 (52-76)	67,78 (45-84)
Education (years)	8,82 (3-17)	12,9 (5-20)	10,78 (5-18)
Handedness	17 R	20 R	25 R

F= female; M= male; R= right hand

2.2 Neuropsychological assessment

All patients underwent a complete standardized neuropsychological examination, administered by experienced neuropsychologists. It included the Mini Mental State Examination (MMSE, Measso et al., 1993) and Montreal Cognitive Assessment (MoCA, Nasreddine et al., 2005; Santangelo et al.,

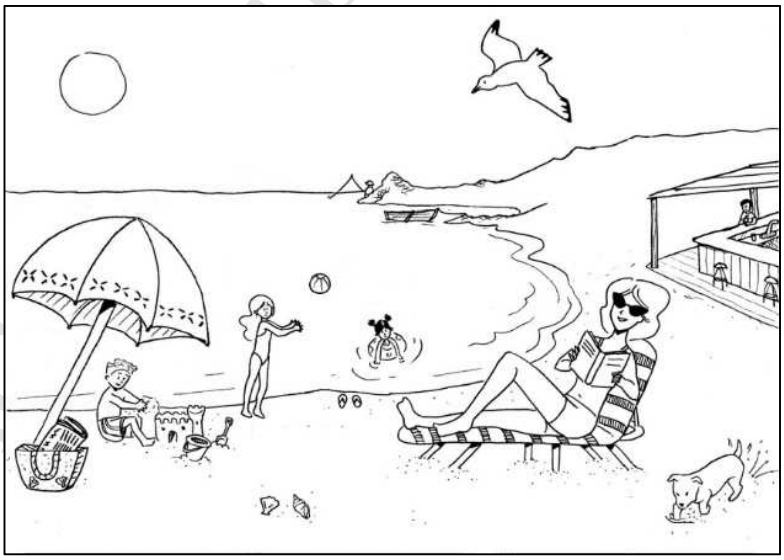
2014) as measures of the global cognitive functioning. Short term memory was assessed using the digit span forward test and the Corsi visuo-spatial span test (Monaco et al., 2013). The immediate and delayed recall scores of the Rey auditory verbal learning test (RAVLT, Carlesimo et al., 1996), the prose memory test (Spinnler and Tognoni, 1987) and the recall of Rey Osterrieth figure (Caffarra et al., 2002) were used to assess episodic memory. Visuo-spatial abilities were evaluated with the Benton Judgement of Line Orientation test (Benton et al., 1978). Visuo-constructive abilities were assessed with the Copy of the Rey Osterrieth figure (Caffarra et al., 2002), the constructional apraxia test (Spinnler and Tognoni, 1987) and Clock drawing test (Mondini et al., 2003). Praxis was assessed using the orofacial praxis test (Spinnler and Tognoni, 1987) and the imitation of gestures (De Renzi et al., 1980, 1986). Attention and executive functions were assessed with the digit span backward test (Monaco et al., 2013), the letter (P-F-L) fluency test (Novelli et al., 1986), the short version of the Stroop Interference Test (Caffarra et al, 2002b) and the Trail making test (Giovagnoli et al., 1996).
The Frontal behavioural inventory (FBI, Alberici et al, 2007), the apathy evaluation scale (AES, Santangelo et al., 2014) and the Beck Depression Inventory (BDI, Visser et al., 2006) were used as measures of affect, personality and social behaviour.

2.3 Speech and language assessment. In order to assess the severity of motor speech disorders an expert speech therapist (A.M.) used a 3-point scale in order to classify the presence and in case the severity of the AOS. The evaluation was based on the oral production obtained at picture description, picture naming, repetition and reading tasks. In each of these tasks we evaluated: 1) articulation: presence of consonant and vowel errors; production of elongated phonemes; errors related to the manner (voiceless and voiced tracts) or to the place of articulation; difficulties in repetition and reading of non-words; error consistency; 2) rhythm and prosody: reduction of the spontaneous speech rate, presence of lengthening phenomena; 3) fluency: alteration of the fluency with self-correction; repetition of syllable sounds; difficulty in starting articulatory sequences. According to the 3-points scale, 0 represents normal speech, 1 is associated to mild articulatory distortions and prosody alterations and 2 represents a production with moderate articulatory disorders. Patients with unintelligible speech were excluded from the study. On this basis, patients were classified in three different groups, namely patients with moderate AOS, patients with mild AOS and patients without AOS.
Language was evaluated with the category fluency test (Novelli et al., 1986) and the SAND battery (Catricalà et al., 2017; Battista et al., 2018). The SAND provides a brief but comprehensive language assessment, including:
- Picture naming: The subject is asked to name 14 black and white object drawings;

- Sentence comprehension: The subject is asked to choose which of two pictures matches the meaning of the sentence read by the examiner. The sentences included two short active, two short passive, two coordinates and two embedded structures;
- Word comprehension: The subject is asked to point at the target among four object pictures in response to a spoken word;
- Repetition: The subject is asked to repeat words and non-words read by the examiner;
- Sentence repetition: The subject is asked to repeat the sentences read by the examiner;
- Reading: The subject is asked to read regular and irregular words and non-words;
- Semantic association: the subject is asked to point at the two semantically related images out of three;
- Writing: The subject is asked to describe how to brush their teeth;
- Picture description task: The subject is asked to describe a complex picture (see below).

Picture description task. To elicit connected speech, we used the Summer Time picture of the SAND battery, depicting a seaside scene (Fig. 1). This picture includes 36 information units (IU), subdivided in four different types: 8 subjects, 10 actions, 5 places and 13 objects. For a detailed description of this task, see Catricalà et al. (2017).

Figure 1: Summer time picture.



The examiner shows the picture to the participant and asks her/him to *"Look carefully at this picture and describe aloud all you can see, trying to use sentences"*. If the subject stops spontaneously the production before the end of two minutes (e.g. pausing longer than 20 seconds,

or claiming to have nothing more to say), the experimenter can intervene, suggesting to look more carefully at the picture and asking if there is something else to describe, in order to encourage the subject to continue. The oral description for each participant was recorded using a stationary microphone attached to a laptop or a digital recorder. Recordings ranging from 25 seconds to 4 min were manually transcribed into Italian orthography, with the exception of punctuation and sentence initial capitalization that were not used. Pauses were marked with dots, with each dot indicating a second of silence. Transcriptions were segmented into utterances, i.e. a sequence of words not interrupted by pause lasting more than 2 seconds. Utterance boundaries were identified using semantic, syntactic and prosodic features, mostly coinciding with sentences, a grammatically complete string of words expressing a complete thought, or a group of words that forms an independent grammatical unit. Fillers such as *ehm* and *mh* were also transcribed. Productions considered as non-descriptive, such as questions addressed to the experimenter (i.e. *What should I say?*), interjections (i.e. *so, I don't know*) and meta-linguistic comments (i.e. *how do you say that? I can't remember the name*) were transcribed, but excluded from the analysis.

All transcriptions were analyzed according to 26 features belonging to four linguistic levels (phonetic-phonological, lexico-semantic, morpho-syntactic and discourse-pragmatic), and selected on the basis of a review of the relevant literature (Rusz et al., 2011; Ash et al., 2012b; Ash and Grossman, 2015; Ash et al., 2011; Murray, 2000; Ellis et al., 2015; Robinson et al., 2015; Santos et al., 2016; Rohrer et al., 2010). The 26 features and their transcription modality are described in Table 2.

Table 2: Summary of the 26 features used

Linguistic feature	Definition/how to measure
Phonetic and Phonological (5)	
Speech rate	Total words uttered/total time in minutes
Total locution time	The amount of time in the sample containing both speech and pauses
Number of pauses	Number of pauses produced / total locution time
Between –utterance pause duration	Total duration of pauses between utterances in seconds/ total locution time
Phonemic errors	Well-articulated phoneme substitutions, additions, transpositions, and deletions / total words
Lexico-semantic (8)	
Noun rate	Total number of nouns / total number of words
Verb rate	Total number of verbs / total number of words
Pronoun rate	Total number of pronouns / total number of words
Noun-verb ratio	Total number of nouns / total number of verbs
Pronoun-noun ratio	Total number of pronouns / total number of nouns
Quantifiers	Total number of quantifiers / total number of nouns
Repaired sequences	Sequences of one or more complete words, resulting in redundancies by

	subsequent repetitions, elaborations or alternative expressions. Repaired sequences / total words
Semantic errors	Total number of errors occurring when a target word is replaced by a term that could, from the context, be identified as a semantically related item; this feature includes semantic (semantically erroneous substitutions) and visual paraphasias (substitutions that are visually similar to the target object). Semantic errors / total words
Morpho-syntactic (5)	
Mean length of sentence (MLS)	The average number of words per sentence
Sentences	Total number of sentences
Incomplete sentences	Total number of sentences that are abandoned after producing a verb. Total number of incomplete sentences / total sentences
Dependent clauses	Total number of clauses that do not form a sentence on their own. Total number of dependent clauses / total sentences
Morpho-Syntactic errors	Erroneous uses of grammatical rules involving sentence structure or ungrammatical sentences; errors in inflection and morphological derivations of words. Morpho-Syntactic errors / total words
Discourse and Pragmatic (8)	
Total words	Total number of words uttered
Information Units	Total number of correct information units; information units are usually subdivided in subjects, places, objects, and actions
Microproposition	Number of utterances which provide details given in addition to the central topic. Microproposition / total sentences
Implausible or irrelevant details	Total number of utterances which provide implausible or irrelevant information given in addition to the central topic. Implausible or irrelevant details / total sentences
Index of discourse effectiveness (IDE)	The ratio of the total number of recalled words divided by the number of information units. Index of discourse effectiveness / total sentences
Errors in content elements	Total number of utterances containing factually inaccurate elements. Errors in content elements / total sentences
Referential cohesion errors	Total number of referential cohesive ties (pronouns), used in an ambiguous or erroneous way. Referential cohesion errors / total pronouns
Efficiency	Total number of information units / total locution time

2.4 Statistical Analyses

Performances obtained by each patient in all neuropsychological and language (SAND) tests were first classified as impaired or unimpaired on the basis of the respective normative data. We then calculated the percentage of impaired cases within each group (PD and PSP).

Subsequently, all the measures derived from neuropsychological and language assessment (using corrected scores) were checked for normality of distribution. For differences between the two patient groups in neuropsychological tests, nonparametric data were analyzed using the Mann–Whitney U test, parametric data using t-test for independent sample. Corrected scores were used as PD and PSP were not balanced for age and education.

For measures derived from the language tests and connected speech production task, the control group was also included in analyses. Bootstrap one-way ANOVA analyses with 1000 bootstrap equally-sized samples obtained by randomly resampling with replacement from the original data

were conducted to assess differences between the three groups (controls, PD and PSP) separately for each test. Corrected scores were used as PD and PSP were not balanced for age and education. Post-hoc analyses were then conducted using Bonferroni correction for differences between the three groups.

Bootstrap ANCOVA analyses with 1000 bootstrap equally-sized samples obtained by randomly resampling with replacement from the original patient data were conducted to assess differences between the three groups (controls, PD and PSP), separately for each linguistic feature. Age and education were used as covariates, as PD and PSP were not balanced for these variables. Post-hoc analyses were then conducted using Bonferroni correction for differences among the three groups.

In order to further check for a possible influence of age and education, in particular on the comparison between PD and PSP, we selected a subsample of 15 PD patients (mean age=65,4; sd=6,58; mean education= 10,67; sd=3,7) matched with controls and PSP patients for both education (F=1,216; p=.304) and age (F=1,198; p=.156). Bootstrap one-way ANOVA analyses with 1000 bootstrap equally-sized samples obtained by randomly resampling with replacement from the original data were conducted to assess differences among the three groups (controls, PD and PSP), separately for each test and feature using the raw scores.

Explorative correlation matrices showing, respectively, the relation between language tests and linguistic features, between language tests and neuropsychological tests, and between linguistic features and neuropsychological tests separately in PSP and PD patients are reported in Supplementary tables 2s - 7s.

In addition, to better qualify the language profile of the PSP patients with different levels of AOS severity, namely PSP without AOS, PSP with mild AOS, and PSP with moderate AOS, we considered the number of patients showing a pathological performance in language tests and linguistic features requiring a verbal output, which were impaired in PSP when compared to controls. In order to classify the performance of each PSP patient as pathological or ‘normal’ we used the corrected score and the cut-off value of the respective normative data for the language tests. For the linguistic features, in the absence of normative data, we used the Crawford & Garthwaite test (2002), in which a patient’s performance is compared to a matched control sample (in our study N=27).

Finally, non parametric analyses were also carried out to investigate differences in language impairments between Richardson’s syndrome (N=10) and the other phenotypes (N=7; 4 with predominant frontal presentation, 2 with predominant parkinsonism and 1 predominant progressive gait freezing) using Bonferroni adjustment (0,05/50).

2.5 Machine learning classification

The goal of a machine learning classification task is to take a feature vector as input, and to produce as output a class label (in this case, either PSP, PD or controls). In order to obtain a classification, it is necessary to train a classifier to predict participants' diagnoses. The features used to populate the vectors[1] are the following:

1) the corrected scores obtained in the 24 specific language tests of the SAND battery (see Table 3);
2) the 26 linguistic features reported above;
3) 1 and 2, i.e. the corrected scores obtained in the 24 specific language tests and the 26 linguistic features.

Age and education were used as extra features for vectors 2 and 3, as we used raw (uncorrected) scores for age and education (note that PD and PSP were not matched for age and education).

Three group vectors were created:

- Group classification task 1: controls vs. PSP; composed by 27 vectors for the 27 instances belonging to the control group and 17 vectors for PSP (Tot: 44 vectors between controls and PSP groups);
- Group classification task 2: PD vs. PSP; composed by 21 vectors for the 21 instances belonging to PD group and 17 vectors for PSP (Tot: 38 vectors between PD and PSP groups).

All combinations of groups were used for the language tests, the linguistic features and the combination of language tests and linguistic features. These vectors were used for training the classifiers, comparing different machine learning algorithms. We used Weka environment (Witten et al., 2016) to test the most appropriate algorithms for the current classification tasks, such as One R, NaïveBayes, NaïveBayesMultinomial, Random Forest and 3 different Support Vector Machines (SVMs), libSVM, libLINEAR and SMO (see supplementary materials for a description).

Our analysis included two steps for all 3 models (language tests, linguistic features and the combination of features and tests): the selection of relevant features and the classification step.

[1] A vector is a multidimensional object whose dimensions/components are usually numerical values representing each feature included in the representation of this object: for instance, a patient A could be represented with a three components vector such as <76, 344, control> where the first number corresponds to his age, the second to the length, in term of number of words, of the description he provided for the picture s/he has been asked to describe, the last component corresponds to his/her clinical classification, i.e. control; the first two components are numerical, the second is nominal and is the class to be learned by the machine learning algorithm after the processing of a number of patients all represented, in this simple case, as 3-components vectors.

As reported in several studies (Fraser et al., 2014; Garrard et al., 2014), the inclusion of too many features could lead to an overfitting of the classifier to idiosyncrasies in the training set, resulting in poor generalization to new data points. For this reason, it is important to include only those features that are non-redundant and highly informative with respect to the classification task. Different feature selection algorithms have been proposed. To select features for training the classifiers, we used two Attribute Selection algorithms: Information Gain and CfsSubsetEleval with Ranking Search as searching method, and mRMR (minimum-Redundancy Maximum-Relevance) as re-ranking method, across all the vectors. Subsequently, we trained the learning algorithms using either the whole set of features or a subset of them, properly selected using both algorithmic procedures mentioned above. We then tested each classifier for accuracy, evaluating each learning algorithm with a percentage split validation, leaving some completely unseen data aside for testing (80% training; 20% testing), for 5 iterative repetition runs to reduce the error rate of the model and to estimate the most accurate learning performance.

Using Experimenter Weka interface, we compared the performances of each learning algorithm using t-test statistic to evaluate if the attribute selection enhanced classification accuracy. Feature selection improved only some of the classification performances, as reported in the results section.

For each model, we reported results in term of classification accuracy of the best three algorithms, i.e. AUC, True Positive rate and precision values. Precision, also known as Positive Predictive Value, is the number of True Positives divided by the number of True Positives and False Positives, i.e. the number of positive predictions divided by the total number of positive values predicted. It can be considered as a measure of a classifier's exactness.

3. Results

3.1 Neuropsychological assessment

The percentage of PSP and PD impaired on each test, on the basis of normative data, as well as the means of the corrected scores (and the p values for differences) for both patient groups are reported in Table 1S in the Supplementary Materials. A high percentage of PSP patients showed an impairment in several cognitive domains, including attentional-executive, visuo-constructional abilities, orofacial praxis and immediate recall tasks, with a mean score significantly lower than PD patients.

3.2 Speech and Language assessment

According to the respective normative data, more than 50% of PSP patients were impaired in picture naming, semantic fluency and in sentence and single word comprehension. More than 40% of PSP were also impaired in non-word repetition, reading and in the number of sentences. The direct comparison between PSP patients and controls showed a lower performance in all tasks, with the exception of word repetition and number of IU, nouns, verbs and semantic errors in writing (see table 3).

PSP were more impaired than PD patients in naming, single word and sentence comprehension, sentence repetition, reading, semantic association, semantic fluency and number of orthographic errors, see table 3. Note, however, that considering a subsample of 15 PD patients matched for both age and education with PSP patients, single word comprehension and sentence repetition were not significantly different. Correlation matrices showing the relation between the neuropsychological tasks and the language tests separately for PSP and PD patients are included in Supplementary Tables 3s and 6s.

Table 3: Speech and Language data of PSP and PD patients

Measure	PSP (n = 17)	% of impaired PSP (n. PSP assessed)	PD (n = 21)	% of impaired PD (n. PD assessed)	controls	P	C vs PSP 95% CI	C vs PD 95% CI	PSP vs PD 95% CI
Apraxia of speech scale:									
Unimpaired		23,53%		95,24%					
Mild		64,71%		4,76%					
Moderate		11,76%		0%					
Semantic Fluency °	25,76 (8,23)	58,8%	37,48 (9,93)	4,8%	-	-	-	-	.000
Screening for Aphasia in NeuroDegeneration - SAND									
Naming:									
Total	9,98 (2,39)	52,9%	12,57 (2,1)	14,3%	13,32 (1,14)	.000	1.5/4.16	ns	-3.58/-0.45
Living	4,75 (1,66)	29,4%	6,33 (1,51)	9,5%	6,64 (0,79)	.000	0.9/2,7	ns	-2.57/-0.31
Non-living	5,21 (1,45)	58,8%	6,38 (0,80)	9,5%	6,7 (0,61)	.000	0.5/1.77	ns	-1.44/-.14
Sentence Comprehension	6,32 (1,44)	52,9%	7,68 (1,12)	9,5%	7,77 (0,48)	.000	0.57/2.21	ns	-2.23/-0.28
Single word comprehension:									
Total	9,96 (1,94)	52,9%	11,4 (1,61)	9,5%	11,83 (0,44)	.001	0.76/2.98	ns	-2.62/-0.04
Living	4,84 (1,38)	41,2%	5,72 (0,88)	9,5%	5,96 (0,18)	.001	0.41/1.96	ns	-1.75/-0.04

Non-living	5,13 (0,68)	29,4%	5,68 (0,77)	9,5%	5,86 (0,43)	**.007**	**0.29/1.12**	ns	Ns
Repetition:									
Total	6,93 (1,95)	23,5%	8 (1,41)	9,5%	8,69 (1,17)	**.006**	**0.53/2.78**	ns	Ns
Words	5,37 (1,1)	17,6%	5,77 (0,61)	9,5%	5,8 (0,54)	ns			.
Non-words	1,4 (1,43)	41,2%	2,21 (1,21)	9,5%	2,93 (0,92)	**.002**	**0.53/2.2**	**0.13/1.43**	Ns
Sentence Repetition:									
Total	3,37 (1,37)	23,5%	4,46 (1,41)	14,3%	4,9 (1,14)	**.003**	**0.68/2.28**	ns	**-1.9/-0.11**
Predictable	1,95 (0,79)	11,8%	2,23 (0,83)	9,5%	2,7 (0,58)	**.01**	**0.17/1.13**	**0.07/0.95**	Ns
Unpredictable	1,43 (0,87)	11,8%	2,27 (0,9)	4,8%	2,19 (0,9)	**.007**	**0.28/1.34**	ns	**-1.44/-0.37**
Reading:									
Total	12,48 (4,39)	41,2%	15,11 (1,36)	4,5%	15,85 (0,35)	**.000**	**1.17/6.08**	**0.22/1.45**	**-5.51/-0.28**
Words (regular and irregular)	9,39 (3,14)	41,2%	11,43 (0,98)	9,5%	11, 94 (0,22)	**.000**	**0.1/4.54**	**0.14/0.99**	**-4.15/-0.41**
Non-words	3,03 (1,36)	23,5%	3,67 (0,63)	4,5%	3,89 (0,31)	**.008**	**0.17/1.66**	ns	Ns
Semantic Association	2,54 (0,78)	7%	3,28 (0,87)	0%	3,59 (0,64)	**.001**	**0.56/1.59**	ns	**-1.29/-0.02**
Writing:									
Information Units	3,87 (1,71)	7% (14)	3,94 (1,77)	25% (20)	4,64 (1,17)	Ns	ns	ns	Ns
Total words	17,75 (11,39)	21,4% (14)	18,22 (10,5)	10% (20)	27,14 (14,19)	**.027**	**1.24/17.02**	**1.16/15.59**	Ns
Noun/total words	0,56 (0,94)	14,3% (14)	0,25 (0,09)	15% (20)	0,26 (0,07)	Ns	ns	ns	ns
Verb/total words	0,22 (0,08)	7% (14)	0,27 (0,12)	5% (20)	0,2 (0,08)	.064	ns	-0.12/-0.006	Ns
Sentences	0,64 (0,44)	50% (14)	0,89 (0,26)	20% (20)	1,19 (0,49)	**.000**	**0.28/0.85**	**0.11/0.54**	Ns
Orthographic Errors	4,56 (6,8)	28,6% (14)	0,46 (0,79)	0% (20)	0,52 (1,07)	**.001**	**-7.91/-1.0**	ns	**1.08/8**
Semantic Errors	0,01 (0,04)	0% (14)	0,01 (0,02)	0% (20)	3,59 (0,64)	Ns	ns	ns	Ns

Mean (standard deviation) and percentage of impaired subjects (out of the number of patients assessed) at each scale and test. C=controls; CI= Confidence Intervals; IU= Information Units; ns=not significant. Bold p values and CI denote significant group differences.

3.3 Picture description task

The means of the corrected scores (and the p values for differences) for the patient and control groups are reported in Table 4. Several linguistic features, belonging to the phonological, lexico-semantic and discourse levels, were significantly different between PSP and controls. PSP showed a lower speech rate, characterized by a lower number of sentences and a higher number of pronouns. At the discourse–pragmatic level, PSP produced a lower number of total words, IUs, index of discourse effectiveness (IDE), and had decreased efficiency, as well a higher number of errors in content elements than controls. In contrast, only efficiency and IDE distinguished between PD and controls, with PD showing a significant impairment.

PSP and PD differed on several linguistic features, belonging to all the considered domains. PSP had a lower speech rate, a lower number of sentences, of total words and IUs. In contrast, PD were lower in noun rate, and produced more incomplete sentences than PSP. When a subsample of 15 PD patients matched for both age and education was compared with PSP patients, the number of errors in content was also significantly different, with PSP producing more errors than PD.

Correlation matrices showing the relation between the neuropsychological tasks, the language tests and the linguistic features separately for PSP and PD patients are included in Supplementary Tables 2s-7s.

Table 4: Connected speech results

Connected speech measures	PSP (n = 17)	PD (n = 21)	Controls (n = 27)	p	PSP vs C 95% CI	PD vs C 95% CI	PSP vs PD 95% CI
Total locution time	74,35 (50,89)	101,43 (39,36)	78,93 (33,87)	ns			
Speech rate	60,73 (22,91)	85,59 (28,65)	91,63 (20,85)	.001	-44,59/-17,21	ns	-44.8/-2.39
Number of pauses	0,11 (0,04)	0,13 (0,07)	0,14 (0,06)	ns			
BUPD	0,27 (0,13)	0,19 (0,13)	0,17 (0,12)	ns			
Phonemic and phonetic errors	0,002 (0,007)	0,001 (0,003)	0,001 (0,003)	ns			
Lexico semantic level							
Noun rate	0,28 (0,05)	0,24 (0,05)	0,27 (0,04)	p=.029	Ns	ns	.003/.08
Verb rate	0,17 (0,03)	0,17 (0,03)	0,16 (0,03)	ns			
Pronoun rate	0,09 (0,06)	0,06 (0,04)	0,04 (0,03)	p=.009	.008/.07	ns	Ns
Noun-verb ratio	1,76 (0,58)	1,53 (0,49)	1,7 (0,47)	ns			
Pronoun-Noun ratio	0,33 (0,24)	0,29 (0,27)	0,18 (0,12)	p=.043	.004/.28	ns	Ns
Quantifiers	0,004 (0,007)	0,008 (0,013)	0,008 (0,008)	ns			
Repaired sequences	0,06 (0,04)	0,07 (0,03)	0,05 (0,03)	ns			
Semantic errors	0,02 (0,02)	0,009 (0,02)	0,008 (0,01)	ns			
Morpho-syntactic level							
Morpho-syntactic errors	0,005 (0,01)	0,003 (0,007)	0,002 (0,006)	ns			
Sentences	8,41 (4,2)	14,9 (6,43)	12,7 (6,06)	p=.018	-7.3/-.58	ns	-9.55/-2.39
MLS	7,86 (1,37)	9,66 (2,24)	9,67 (2,81)	p=.067	-3,17/-.46	ns	-2,91/-,31
Tot. Incomplete sentences	0,029 (0,06)	0,076 (0,09)	0,037 (0,1)	p=.059	Ns	ns	-.13/-.01
Tot. Dependent clauses	0,45 (0,36)	0,54 (0,28)	0,6 (0,36)	ns			
Discourse-pragmatic level							
Total words	66,53 (37,23)	141,05 (63,58)	117,81 (53,55)	p=.004	-76.51/-16.77	ns	-97.51/-31,51
Micropropositions	0	0,022 (0,05)	0,014 (0,041)	ns			
Implausible or irrelevant details	0,05 (0,09)	0,11 (0,13)	0,08 (0,12)	ns			

Errors in content elements	0,23 (0,25)	0,09 (0,13)	0,07 (0,09)	p=.023	.04/.27	ns	Ns
IU	13,76 (6,7)	24,57 (11,32)	25,78 (8,45)	p=.002	-15.06/-5.58	ns	-14.03/-.89
IDE	5,32 (2,6)	6,45 (3,02)	4,5 (1,16)	p=.004	Ns	.93/3.77	Ns
Efficiency	0,23 (0,13)	0,25 (0,11)	0,35 (0,1)	p=.001	-.19/-.03	-.18/-.05	Ns
Cohesion referential errors	0,06 (0,1)	0,06 (0,11)	0,03 (0,13)	ns			

Values shown are mean and standard deviation. C= controls; CI= confidence Intervals; BUPD= between utterance pause duration; MLS=mean length of sentence; IU=information Units; IDE=Index of discourse effectiveness.

3.4 Classification

We consider separate classifications for the language tests of SAND, for the linguistic features and for the combination of language tests and linguistic features to distinguish between: (1) controls and PSP, and (2) PD and PSP. Table 5 reports accuracy, AUC, precision and the True positive rate values of the three best algorithms for the different measures used. The most relevant language tests and linguistic features are reported in Table 6. Language tests obtained the best performance using Random Forest algorithm, showing 88,33% of accuracy in classifying controls versus PSP and 85,12% in classifying PSP versus PD. The NaïveBayes Multinomial algorithm showed the best accuracy for the controls vs. PSP (83,06%) and PD-PSP (83,21%) comparisons using linguistic features. The combination of both language tests and linguistic features obtained the best classifications using the NaiveBayes algorithm, namely an accuracy of 90,65% in discriminating PSP from controls and of 87,34 in discriminating PSP from PD.

Table 5: results of the best three algorithms based on 24 SAND tests, the 26 linguistic features and the combination of the SAND tests and linguistic features, used in distinguishing controls and PSP and PD and PSP. Accuracy, True Positive (TP) rate, precision, and Area Under the Curve (AUC) values, with and without features selection, are reported for each algorithm.

	Controls – PSP				**PSP – PD**			
	Accuracy	**TP Rate**	**Precision**	**AUC**	**Accuracy**	**TP Rate**	**Precision**	**AUC**
SAND TESTS								
NaiveBayes	86,11	0,87	0,86	0,92	74,68	0,6	0,83	0,82
NaiveBayes + Attr.Sel.	79,17	0,87	0,75	0,9	66,47	0,43	0,77	0,74
LibLINEAR	81,11	0,65	0,85	0,79	69,96	0,55	0,73	0,69
LibLINEAR + Attr.Sel.	88,06	0,82	0,9	0,87	66,75	0,48	0,73	0,65
Random Forest	**88,33**	0,87	0,91	0,93	**85,12**	0,83	0,87	0,91
RandomForest + Attr.Sel.	86,11	0,87	0,86	0,91	76,9	0,67	0,82	0,83

LINGUISTIC FEATURES								
NaiveBayes	73,78	0,78	0,82	0,68	78,77	0,83	0,75	0,85
NaiveBayes + Attr. Sel.	78,89	0,74	0,92	0,84	73,77	0,9	0,67	0,82
NaïveBayes Multinomial	**83,06**	0,89	0,87	0,83	**83,21**	0,88	0,8	0,96
NaïveBayes Multinomial + Attr. Sel.	69,44	0,78	0,76	0,74	73,41	0,65	0,8	0,79
libLINEAR	78,33	0,81	0,86	0,77	75,91	0,77	0,74	0,76
libLINEAR + Attr. Sel.	81,11	0,81	0,88	0,81	68,97	0,67	0,64	0,69
SAND TESTS + LINGUISTIC FEATURES								
NaiveBayes	**90,65**	0,87	0,93	0,85	**87,34**	0,83	0,9	0,92
NaiveBayes + Attr.Sel.	83,33	0,87	0,77	0,89	76,27	0,78	0,75	0,85
libLINEAR	83,61	0,82	0,82	0,83	78,77	0,83	0,79	0,8
libLINEAR + Attr.Sel.	76,11	0,68	0,72	0,75	64,6	0,57	0,62	0,64
RandomForest	88,33	0,87	0,88	0,94	82,26	0,78	0,82	0,89
RandomForest + Attr.Sel.	90,56	0,87	0,91	0,92	71,55	0,68	0,74	0,77

Table 6: The most relevant attributes selected using Information Gain (IG) and minimum-Redundancy Maximum-Relevance (mRMR) methods for each classification and for each measure, i.e. language tests, linguistic features and the combination of language tests and linguistic features.

	Controls – PSP		PSP – PD	
	IG	mRMR	IG	mRMR
SAND TESTS				
	0,413	Semantic association	0,316	Semantic association
	0,394	Naming (total)	0,308	Sentence comprehension
	0,299	Non-word repetition	0,204	Writing - n. orthographic errors
	0,299	Sentence comprehension		
	0,254	Word comprehension (living)		
	0,234	Sentence repetition (predictable)		
	0,168	Writing – sentences		
LINGUISTIC FEATURES				
	0,333	Information units	0,343	Total words
	0,328	Speech rate	0,28	Information Units
	0,254	Efficiency	0,213	Mean lenght of sentences
	0,232	Pronoun rate	0,212	Semantic errors
	0,192	Errors in content elements	0,21	Index of discourse effectiveness
SAND TESTS + LINGUISTIC FEATURES				
	0,413	Semantic association	0,343	Total Words
	0,388	Naming (total)	0,316	Semantic association
	0,328	Speech Rate	0,308	Sentence comprehension

0,299	Reading (total)	0,28	Information Units
0,28	Reading (words)	0,213	Mean length of sentences
0,28	Repetition (total)	0,212	Semantic errors
0,279	Word comprehension (non living)	0,21	Index of discourse effectiveness
0,254	Efficiency	0,128	Writing - n. orthographic errors
0,232	Pronoun Rate		
0,192	Errors in Content Elements		
0,168	Writing – number of words		
0,168	Writing – sentences		

3.5 Differences according to AOS severity and PSP phenotypes

Table 7 shows the number of PSP patients, divided according to the level of AOS severity, impaired at each language task and linguistic features significantly different between PSP and controls. Qualitatively, it can be observed that the two patients with moderate AOS did not show an impairment profile that can be completely explained by the presence of AOS.

While semantic fluency and speech rate were reduced, neither of these patients showed an impairment in the total number of words produced, in the number of sentences and in the mean length of sentences.

Table 7: number of PSP patients, divided according to the level of AOS severity (namely no AOS, mild AOS and moderate AOS), resulting impaired at each linguistic variable considered (see text for details).

AOS	Number of PSP	Semantic fluency	Naming (total)	Repetition (total)	Sentence repetition (total)	Reading (total)	Speech rate	Sentences	MLS	Total words	Errors in content elements	IU	Efficiency
Unimpaired	4	0	1	0	0	1	1	0	0	0	1	1	1
Mild	11	8	6	2	3	5	3	0	0	0	8	5	5
Moderate	2	2	2	2	1	1	2	0	0	0	0	0	1

Patients with Richardson's syndrome and other PSP phenotypes (NON RS) were matched for age (p=0,364), education (p=0,364) and disease duration (p=0,635). Only picture naming (RS mean=8,6, ds=1,9; non RS mean=11,43, ds=2,3; p=0,025) and word comprehension (RS mean=8,7, ds=1,57; nonRS mean=11,71, ds=0,49; p<0,000) tests were significantly different between PSP groups, with a more severe impairment in RS patients (see figure 2). The number of information units was the only significant linguistic feature, with RS patients producing lesser information (mean RS=10,5; ds=5,3; mean nonRS=18,4, ds=5,3; p=0,002). Adjusting with Bonferroni

correction, considering the number of comparison for 24 language tests and 26 linguistic features (0,05/50=0,001), only the word comprehension score was significant.

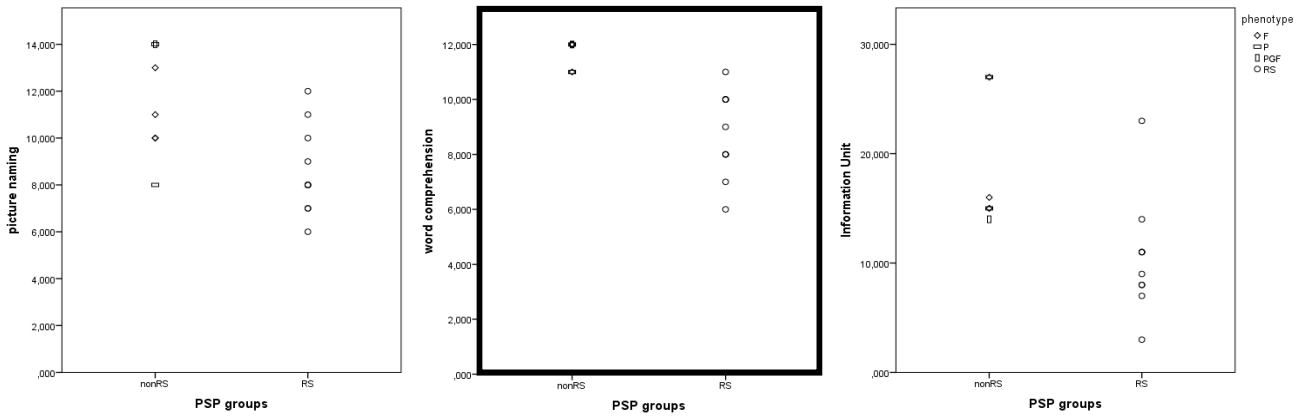


Figure 2: performance of PSP groups (RS and NON RS) at picture naming and word comprehension tasks and the number of information units

4. Discussion

The aim of this study is to identify the linguistic profile of PSP patients presenting primarily with a movement disorder, through a comprehensive assessment including language tests and the linguistic features provided by connected speech analysis.

The first result is that subtle language deficits can be observed also in PSP patients with non aphasic/speech apraxia presentation. Machine learning classifications, using separately language tests and linguistic features, resulted in a high performance in discriminating between PSP and controls and between PSP and PD, on the basis of either language tests or linguistic features. The combination of both types of measures further improved the classification accuracy in distinguishing PSP from controls and PD.

4.1 Qualitative speech and language impairment in non-aphasic PSP

Neuropsychological profile. At least 50% of PSP patients showed an impairment in attentional-executive abilities, including set shifting, inhibitory control and cognitive flexibility, and in immediate recall tasks, visuo-constructional abilities, and orofacial praxis, in line with previous findings (Lee et al., 2016; Monza et al., 1998; Robbins et al., 1994; Soliveri et al., 2000; Santangelo et al., 2018). The prominent executive impairment is attributed to the frontal-subcortical

dysfunction present in PSP, and has been considered to be responsible also for memory disorders in these patients (Pillon et al., 1994; Santangelo et al., 2018).

Speech and language profile.

Machine learning classification showed that language tests correctly classified 88,33% of controls and PSP. The most relevant tests included semantic association, picture naming, sentence and word comprehension, i.e. tasks already reported to be impaired in PSP when compared to controls (Burrell et al., 2018). Non-word and sentence repetition and the number of sentences produced in a written description task were also found to discriminate PSP vs. controls (Burrell et al., 2018).

We found that even using relatively short samples of descriptive speech, classifiers were able to achieve a high degree of accuracy in distinguishing PSP patients from controls (83,06%). The best distinguishing features were speech rate, IUs, the number of pronouns, efficiency, and the number of error in content elements, with PSP patients obtaining a worse performance than controls.

These results suggest the presence of an impairment mainly at lexical-semantic level, as characterized by a lower performance in measures like naming, semantic association, word comprehension, speech rate, IUs, and number of pronouns. A milder deficit at the pragmatic-discourse level is also detected on the basis of measures such as efficiency and number of errors in content elements.

Naming impairment, in particular, is characterized by a high number of semantic errors (55,56%), followed by visual (19%), visual/semantic and phonological/semantic (7,9%), and phonological (4%) errors. Correlation analyses showed a significant association between picture naming, semantic fluency, and word comprehension.

The presence of a lower speech rate in PSP can be attributed to different factors. No correlation was found for specific neuropsychological test scores. The slower speech rate in PSP patients may include a motor component, reflected by articulatory alterations in the AOS scale. Although we have excluded patients with unintelligible speech, 65% of PSP presented with a mild articulatory impairment and 12% (namely 2 patients) with a moderate impairment (see Table 3). The performance of the two patients with moderate AOS, however, does not support an exclusive role of articulatory disorders. While both patients were impaired at speech rate, the number of words or sentences produced in the picture description task was not significantly reduced. In addition, exclusion of these two patients from the analyses comparing PSP with controls did not change the results.

This feature correlated with semantic fluency, naming, number of total words and of sentences in the written description task, and with IU, verb rate and efficiency of picture description task. Further investigations, using a larger number of participants with different levels of AOS severity, are needed

to clarify the role of the articulatory component.

Impaired IU could be considered as reflecting impaired global coherence, which indicates utterances closely associated with the general topic, and in the case of IU could refer to the items depicted in the picture (Croisile et al., 1996). This finding may thus be also consistent with the naming impairment (Burrell et al., 2018). In fact, IUs correlated with semantic fluency, naming, word comprehension, reading, speech rate, number of words and sentences on the picture description task, suggesting the presence of a lexico-semantic impairment resulting in a simplified description (number of words and sentences). The increased usage of demonstrative pronouns such as *this, that*, or of personal pronouns, such as *he, she,* used deictically, rather than of more specific nouns i.e. *child, woman,* in PSP patients may also be compatible with word-finding difficulties. Pronoun rate correlated with naming scores and with the number of dependent clauses in the picture description task.

Errors in naming content elements, together with other impaired features, result in discourse, which is weakly cohesive (pronoun misuse) and coherent (IU underuse).

To summarize, the most relevant tests and features that distinguish PSP patients' language involve mainly the lexical-semantic and discourse-pragmatic levels. It is noteworthy that PSP patients with the Richardson's syndrome showed a greater impairment on word comprehension when compared to PSP patients with the other phenotypes. A semantic memory impairment, assessed with picture naming, written word synonym and judgment task and an association task including both word and picture versions has been already reported in these patients (van der Huk and Hodges, 1995). It is noteworthy that a focal, bilateral cortical thinning involving not only the prefrontal/precentral cortex but also the temporal pole, a crucial area for semantic cognition severely affected in the semantic variant of PPA (Iaccarino et al., 2015), has been recently reported in PSP patients with the Richardson's variant (Caso et al., 2016). With the exception of semantic fluency, all these tasks use pictorial stimuli, where performance can be affected by the visual problems (blurred vision and diplopia) frequently reported by these patients (Kim et al., 2015, Podoll et al., 1991). The correlation analyses, however, did not show a significant association between picture naming and visuospatial performance.

Distinguishing our patients from PSP patients with a prominent language impairment (PSP-PNFA), a prominent syntactic or phonetic impairment was not present in our sample. A direct comparison of PSP-PNFA, patients with PNFA and PSP with a prominent movement disorder based on the same comprehensive assessment is beyond the aims of the present paper, but future studies could provide further important insights about differences in site and extent of brain pathology and their relationship with the heterogeneity of the language phenotype in these conditions.

4.2 Progressive supranuclear palsy versus Parkinson's disease

Neuropsychological profile. PSP patients were more impaired in comparison with PD patients on attentional-executive and memory tasks, visuo-constructional abilities and praxis. A greater cognitive impairment in PSP with respect to PD patients has been reported in other studies, in particular in the case of executive functions (Santangelo et al., 2018).

Speech and language profile. Machine learning algorithms resulted in a high accuracy (87%) in detecting differences in language performance between PSP and PD. The most discriminant language tests were semantic association, sentence comprehension and writing (orthographic errors), with PSP more impaired than PD. 'The most discriminating linguistic features were the number of total words, IU, MLS, and IDE, lower in PSP than in PD, as well as the number of semantic errors, higher in PSP. Taken together, these data confirm the presence of a distinctive and more severe language impairment in PSP when compared to PD patients, mainly involving semantic abilities. Tests such as semantic association tests and language features including number of total words in oral production, IU and semantic errors are in fact ascribable at the lexico-semantic level, as described in the previous section.

In addition, PSP were more impaired than PD at syntactic level, producing a lower mean length of sentence, according to the significant correlation of this measure with the number of incomplete sentences in connected speech and the number of sentences in the writing task.

The reason for the lower score obtained by PSP patients on sentence comprehension is less clear and does not seem associated to a pure syntactic impairment. In fact, sentences with the more complex syntactic structures were not the most impaired. The short passive sentences were well understood by at least the 82% of PSP patients, the two coordinates sentences were well performed by at least 52%, the two longer embedded sentences by at least 59%, and the short active sentences were correctly comprehended by at least the 94% of PSP patients. Correlation analyses in addition showed an association between the sentence comprehension performance and some lexico-semantic

and executive tasks, and only with the number of sentences produced as a syntactic measure. Syntactic comprehension disturbances have been already reported in PSP (Burrell et al., 2018), but not consistently (Cotelli et al., 2007).

At the pragmatic and discourse level, PSP patients had a distinctive disorder characterized by a lower performance on efficiency and produced a greater number of errors in content elements when compared with controls, but not with PD. In fact, PD patients were worse than PSP patients on discourse effectiveness, as they used more words to describe the same content element identified by PSP patients, perhaps to compensate to the lower number of nouns. In fact, the index of discourse effectiveness negatively correlated with the number of nouns in PD. Deficits at pragmatic discourse measures have been previously reported in PD (Ash et al., 2012; Boschi et al., 2017). As shown by correlational analyses, the index of discourse effectiveness was related with tasks assessing working memory and mental flexibility, suggesting an association between difficulties on discourse measures and executive dysfunction in PD, likely reflecting prefrontal dysfunction (Ash et al., 2012). Differences in discourse measures emerging from the comparison of PSP and PD patients should be further investigated for a better characterization of the pragmatic impairment in patients with movement disorders.

5. Conclusion

We have demonstrated high classification accuracy provided by machine learning in discriminating PSP from healthy subjects and PSP from PD, in particular combining language tests and linguistic features provided by a connected speech task. Although PSP with prevalent movement disorders is not typically associated with language deficits, our analyses indicate the presence of a subtle language impairment, involving mainly lexical-semantic and discourse-pragmatic levels. While lexical-semantic impairment characterizes the linguistic profile of PSP patients when compared with controls and PD, deficits at discourse level are a common feature of PD.

The language profile of PSP described by our studies comparing patients with PSP to those with PD and to controls characterizes a unique profile of language impairment. Detailed information about language performance in patients with PSP may contribute to categorizing and distinguishing among the different PSP phenotypes. Further, the findings of our study, which identify unique

linguistic deficits in PSP may prove valuable in constructing sensitive neuropsychological tests as part of a diagnostic evaluation.

Competing interests

None declared.

Acknowledgements

This study was supported by the MRC Research Grant: Ref MR/N025881/1, and by the "AIRAlzh Onlus" and "ANCC-COOP" Italia issued to V.M. Borsa.

References

Alberici, A., Geroldi, C., Cotelli, M., Adorni, A., Calabria, M., Rossi, G., ... & Kertesz, A. (2007). The Frontal Behavioural Inventory (Italian version) differentiates frontotemporal lobar degeneration variants from Alzheimer's disease. Neurological Sciences, 28(2), 80-86.

Ash, S. and Grossman, M. (2015) Why study connected speech production? In Willems, R. M. (Ed.). Cognitive neuroscience of natural language use. Cambridge University Press, pp. 29-58.

Ash, S., Xie, S. X., Gross, R. G., Dreyfuss, M., Boller, A., Camp, E., et al. (2012a). The organization and anatomy of narrative comprehension and expression in Lewy body spectrum disorders. Neuropsychology 26:368.

Ash, S., McMillan, C., Gross, R. G., Cook, P., Gunawardena, D., Morgan, B., ... & Grossman, M. (2012b). Impairments of speech fluency in Lewy body spectrum disorder. Brain and language, 120(3), 290-302.

Ash, S., McMillan, C., Gross, R. G., Cook, P., Morgan, B., Boller, A., ... & Grossman, M. (2011). The organization of narrative discourse in Lewy body spectrum disorder. Brain and language, 119(1), 30-41.

Bak, T. H., Yancopoulou, D., Nestor, P. J., Xuereb, J. H., Spillantini, M. G., Pulvermüller, F., & Hodges, J. R. (2005). Clinical, imaging and pathological correlates of a hereditary deficit in verb and action processing. *Brain*, *129*(2), 321-332.

Battista, P., Catricalà, E., Piccininni, M., Copetti, M., Esposito, V., Polito, C., ... & Picillo, M. (2018). Screening for Aphasia in NeuroDegeneration for the Diagnosis of Patients with Primary Progressive Aphasia: Clinical Validity and Psychometric Properties. *Dementia and Geriatric Cognitive Disorders*, *46*(3-4), 243-252.

Benton, A. L., Varney, N. R., & deS Hamsher, K. (1978). Visuospatial judgment: A clinical test. Archives of Neurology, 35(6), 364-367.

Boeve B, Dickson D, Duffy J, Bartleson J, Trenerry M, Petersen R (2003) Progressive nonfluent aphasia and subsequent aphasic dementia associated with atypical progressive supranuclear palsy pathology. Eur Neurol 49, 72-78.

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

Burrell, J. R., Ballard, K. J., Halliday, G. M., & Hodges, J. R. (2017). Aphasia in Progressive Supranuclear Palsy: As Severe as Progressive Non-Fluent Aphasia. Journal of Alzheimer's Disease, (Preprint), 1-11.

Burrell, J. R., Hodges, J. R., & Rowe, J. B. (2014). Cognition in corticobasal syndrome and progressive supranuclear palsy: a review. Movement Disorders, 29(5), 684-693.

Carlesimo, G. A., Caltagirone, C., Gainotti, G. U. I. D., Fadda, L., Gallassi, R., Lorusso, S., ... & Parnetti, L. (1996). The mental deterioration battery: normative data, diagnostic reliability and qualitative analyses of cognitive impairment. European neurology, 36(6), 378-384.

Caffarra, P., Vezzadini, G., Dieci, F., Zonato, F., & Venneri, A. (2002a). Rey-Osterrieth complex figure: normative values in an Italian population sample. Neurological Sciences, 22(6), 443-447.

Caffarra, P., Vezzadini, G., Dieci, F., Zonato, F., & Venneri, A. (2002b). Una versione abbreviata del test di Stroop. Nuova Rivista di Neurologia, 12(4), 111-115.

Caso, F., Agosta, F., Volonté, M. A., Ferraro, P. M., Tiraboschi, P., Copetti, M., ... & Filippi, M. (2016). Cognitive impairment in progressive supranuclear palsy-Richardson's syndrome is related to white matter damage. Parkinsonism & Related Disorders, 31, 65-71.

Catricalà, E., Gobbi, E., Battista, P., Miozzo, A., Polito, C., Boschi, V., ... & Garrard, P. (2017). SAND: a Screening for Aphasia in NeuroDegeneration. Development and normative data. Neurological Sciences, 38(8), 1469-1483.

Crawford, J. R., & Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia*, *40*(8), 1196-1208.

Cotelli, M., Borroni, B., Manenti, R., Alberici, A., Calabria, M., Agosti, C., ... & Zanetti, O. (2006). Action and object naming in frontotemporal dementia, progressive supranuclear palsy, and corticobasal degeneration. Neuropsychology, 20(5), 558.

Cotelli, M., Borroni, B., Manenti, R., Ginex, V., Calabria, M., Moro, A., ... & Padovani, A. (2007). Universal grammar in the frontotemporal dementia spectrum: evidence of a selective disorder in the corticobasal degeneration syndrome. *Neuropsychologia*, *45*(13), 3015-3023.

Croisile, B., Ska, B., Brabant, M. J., Duchene, A., Lepage, Y., Aimard, G., & Trillet, M. (1996). Comparative study of oral and written picture description in patients with Alzheimer's disease. Brain and language, 53(1), 1-19.

Daniele A, Barbier A, Di Giuda D, Vita MG, Piccininni C, Spinelli P, Tondo G, Fasano A, Colosimo C, Giordano A, Gainotti G (2013) Selective impairment of action-verb naming and comprehension in progressive supranuclear palsy. Cortex 49, 948-960.

De Renzi, E., Motti, F., & Nichelli, P. (1980). Imitating gestures: a quantitative approach to ideomotor apraxia. Archives of Neurology, 37(1), 6-10.

De Renzi, E., Faglioni, P., Scarpa, M., & Crisi, G. (1986). Limb apraxia in patients with damage confined to the left basal ganglia and thalamus. Journal of Neurology, Neurosurgery & Psychiatry, 49(9), 1030-1038.

Sala, S. D., & Spinnler, H. (1998). Echolalia in a case of progressive supranuclear palsy. Neurocase, 4(2), 155-165.

Ellis, C., Crosson, B., Gonzalez Rothi, L. J., Okun, M. S., & Rosenbek, J. C. (2015). Narrative discourse cohesion in early stage Parkinson's disease. Journal of Parkinson's disease, 5(2), 403-411.

Esmonde, T., Giles, E., Xuereb, J., & Hodges, J. (1996). Progressive supranuclear palsy presenting with dynamic aphasia. Journal of Neurology, Neurosurgery & Psychiatry, 60(4), 403-410.

Fahn S, Elton R. Unified Parkinson's Disease Rating Scale. In: Recent developments in Parkinson's disease. 1987. p. 153–63.

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

Garrard, P., Rentoumi, V., Gesierich, B., Miller, B., & Gorno-Tempini, M. L. (2014). Machine learning approaches to diagnosis and laterality effects in semantic dementia discourse. Cortex, 55, 122-129.

Gibb, W. R., & Lees, A. J. (1988). The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. Journal of Neurology, Neurosurgery & Psychiatry, 51(6), 745-752.

Giovagnoli, A. R., Del Pesce, M., Mascheroni, S., Simoncelli, M., Laiacona, M., & Capitani, E. (1996). Trail making test: normative values from 287 normal adult controls. The Italian journal of neurological sciences, 17(4), 305-309.

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

Iaccarino, L., Crespi, C., Della Rosa, P. A., Catricalà, E., Guidi, L., Marcone, A., ... & Perani, D. (2015). The semantic variant of primary progressive aphasia: clinical and neuroimaging evidence in single subjects. *PLoS One*, *10*(3), e0120197.

Höglinger, G. U., Respondek, G., Stamelou, M., Kurz, C., Josephs, K. A., Lang, A. E., ... & Arzberger, T. (2017). Clinical diagnosis of progressive supranuclear palsy: the movement disorder society criteria. Movement Disorders, 32(6), 853-864.

Josephs, K. A., Duffy, J. R., Strand, E. A., Whitwell, J. L., Layton, K. F., Parisi, J. E., ... & Dickson, D. W. (2006). Clinicopathological and imaging correlates of progressive aphasia and apraxia of speech. Brain, 129(6), 1385-1398.

Josephs, K. A., Boeve, B. F., Duffy, J. R., Smith, G. E., Knopman, D. S., Parisi, J. E., ... & Dickson, D. W. (2005). Atypical progressive supranuclear palsy underlying progressive apraxia of speech and nonfluent aphasia. Neurocase, 11(4), 283-296.

Kertesz, A., & McMonagle, P. (2010). Behavior and cognition in corticobasal degeneration and progressive supranuclear palsy. Journal of the neurological sciences, 289(1), 138-143.

Kim, J. H., & McCann, C. M. (2015). Communication impairments in people with progressive supranuclear palsy: A tutorial. Journal of communication disorders, 56, 76-87.

Lee, Y. E. C., Williams, D. R., & Anderson, J. F. (2016). Frontal deficits differentiate progressive supranuclear palsy from Parkinson's disease. Journal of neuropsychology (10), 1–14.

Maher, E. R., Smith, E. M., & Lees, A. J. (1985). Cognitive deficits in the Steele-Richardson-Olszewski syndrome (progressive supranuclear palsy). Journal of Neurology, Neurosurgery & Psychiatry, 48(12), 1234-1239.

Measso, G., Cavarzeran, F., Zappalà, G., Lebowitz, B. D., Crook, T. H., Pirozzolo, F. J., ... & Grigoletto, F. (1993). The mini-mental state examination: Normative study of an Italian random sample. Developmental Neuropsychology, 9(2), 77-85.

Monaco, M., Costa, A., Caltagirone, C., & Carlesimo, G. A. (2013). Forward and backward span for verbal and visuo-spatial data: standardization and normative data from an Italian adult population. Neurological Sciences, 34(5), 749-754.

Mondini, S., Mapelli, D., Vestri, A., & Bisiacchi, P. S. (2003). Esame neuropsicologico breve. Milano: Raffaello Cortina Editore, 160.

Monza, D., Soliveri, P., Radice, D., Fetoni, V., Testa, D., Caffarra, P., ... & Girotti, F. (1998). Cognitive dysfunction and impaired organization of complex motility in degenerative parkinsonian syndromes. Archives of neurology, 55(3), 372-378.

Murray, L. L. (2000). Spoken language production in Huntington's and Parkinson's diseases. Journal of Speech, Language, and Hearing Research, 43(6), 1350-1366.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. Journal of the American Geriatrics Society, 53(4), 695-699.

Novelli, G., Papagno, C., Capitani, E., & Laiacona, M. (1986). Tre test clinici di ricerca e produzione lessicale. Taratura su soggetti normali. Archivio di psicologia, neurologia e psichiatria.

Payan, C. A., Viallet, F., Landwehrmeyer, B. G., Bonnet, A. M., Borg, M., Durif, F., ... & Agid, Y. (2011). Disease severity and progression in progressive supranuclear palsy and multiple system atrophy: validation of the NNIPPS–Parkinson Plus Scale. PLoS One, 6(8), e22293.

Perkin, G. D., Lees, A. J., Stern, G. M., & Kocen, R. S. (1978). Problems in the Diagnosis of Progressive Supranuclear Palsy (Steele—Richardson—Olszewski Syndrome). Canadian Journal of Neurological Sciences, 5(2), 167-173.

Pillon, B., Deweer, B., Michon, A., Malapani, C., Agid, Y., & Dubois, B. (1994). Are explicit memory disorders of progressive supranuclear palsy related to damage to striatofrontal circuits? Comparison with Alzheimer's, Parkinson's, and Huntington's diseases. *Neurology*, *44*(7), 1264-1264.

Podoll, K., Schwarz, M., & Noth, J. (1991). Language functions in progressive supranuclear palsy. Brain, 114(3), 1457-1472.

Robbins, T. W., James, M., Owen, A. M., Lange, K. W., Lees, A. J., Leigh, P. N., ... & Summers, B. A. (1994). Cognitive deficits in progressive supranuclear palsy, Parkinson's disease, and multiple system atrophy in tests sensitive to frontal lobe dysfunction. Journal of Neurology, Neurosurgery & Psychiatry, 57(1), 79-88.

Robinson, G. A., Spooner, D., & Harrison, W. J. (2015). Frontal dynamic aphasia in progressive supranuclear palsy: distinguishing between generation and fluent sequencing of novel thoughts. Neuropsychologia, 77, 62-75.

Robinson, G., Shallice, T., & Cipolotti, L. (2006). Dynamic aphasia in progressive supranuclear palsy: A deficit in generating a fluent sequence of novel thought. Neuropsychologia, 44(8), 1344-1360.

Rohrer, J. D., Paviour, D., Bronstein, A. M., O'sullivan, S. S., Lees, A., & Warren, J. D. (2010). Progressive supranuclear palsy syndrome presenting as progressive nonfluent aphasia: a neuropsychological and neuroimaging analysis. Movement Disorders, 25(2), 179-188.

Rusz, J., Cmejla, R., Ruzickova, H., & Ruzicka, E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. The Journal of the Acoustical Society of America, 129(1), 350-367.

Santos-Santos, M. A., Mandelli, M. L., Binney, R. J., Ogar, J., Wilson, S. M., Henry, M. L., ... & Pakvasa, M. (2016). Features of Patients with Nonfluent/Agrammatic Primary Progressive Aphasia With Underlying Progressive Supranuclear Palsy Pathology or Corticobasal Degeneration. JAMA neurology, 73(6), 733-742.

Santangelo, G., Barone, P., Cuoco, S., Raimo, S., Pezzella, D., Picillo, M., ... & Franco, S. (2014). Apathy in untreated, de novo patients with Parkinson's disease: validation study of Apathy Evaluation Scale. Journal of neurology, 261(12), 2319-2328.

Santangelo, G., Cuoco, S., Pellecchia, M. T., Erro, R., Barone, P., & Picillo, M. (2018). Comparative cognitive and neuropsychiatric profiles between Parkinson's disease, multiple system atrophy and progressive supranuclear palsy. *Journal of neurology*, *265*(11), 2602-2613.

Soliveri, P., Monza, D., Paridi, D., Carella, F., Genitrini, S., Testa, D., & Girotti, F. (2000). Neuropsychological follow up in patients with Parkinson's disease, striatonigral degeneration-type multisystem atrophy, and progressive supranuclear palsy. Journal of Neurology, Neurosurgery & Psychiatry, 69(3), 313-318.

Spinnler, H., & Tognoni, G. (1987). Italian Group on the Neuropsychological Study of Ageing: Italian standardization and classification of neuropsychological tests. Ital J Neurol Sci, 6(suppl 8), 1-120.

Van der Hurk, P. R., & Hodges, J. R. (1995). Episodic and semantic memory in Alzheimer's disease and progressive supranuclear palsy: a comparative study. *Journal of clinical and experimental neuropsychology*, *17*(3), 459-471.

Visser, M., Leentjens, A. F., Marinus, J., Stiggelbout, A. M., & van Hilten, J. J. (2006). Reliability and validity of the Beck depression inventory in patients with Parkinson's disease. Movement Disorders, 21(5), 668-672.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

Eleonora Catricalà: Design and conceptualization of study; Acquisition, analysis and interpretation of data; Drafted and revised the manuscript for intellectual content

Veronica Boschi: Design and conceptualization of study; Acquisition, analysis and interpretation of data; Drafted and revised the manuscript for intellectual content.

Sofia Cuoco: Acquisition, analysis and interpretation of data; revised the manuscript for intellectual content.

Francesco Galiano: Analysis and interpretation of data; revised the manuscript for intellectual content.

Marina Picillo: Acquisition and interpretation of data; revised the manuscript for intellectual content.

Elena Gobbi: Acquisition and interpretation of data; revised the manuscript for intellectual content.

Antonio Miozzo: Acquisition and interpretation of data; revised the manuscript for intellectual content.

Cristiano Chesi: Analysis and interpretation of data; revised the manuscript for intellectual content

Valentina Esposito: Acquisition and interpretation of data; revised the manuscript for intellectual content.

Gabriella Santangelo: Acquisition and interpretation of data; revised the manuscript for intellectual content.

Maria Teresa Pellecchia: Acquisition and interpretation of data; revised the manuscript for intellectual content.

Virginia M. Borsa: Analysis and interpretation of data; revised the manuscript for intellectual content.

Paolo Barone: Design and conceptualization of study; interpretation of data; revised the manuscript for intellectual content.

Peter Garrard: Design and conceptualization of study; interpretation of data; revised the manuscript for intellectual content.

Sandro Iannaccone: Design and conceptualization of study; interpretation of data; revised the manuscript for intellectual content.

Stefano F. Cappa: Design and conceptualization of study; interpretation of data; drafted and revised the manuscript for intellectual content.