

1 **Language Uncovers Visuospatial Dysfunction in Posterior Cortical Atrophy:**
2 **A Natural Language Processing Approach**

3
4

5 Neguine Rezaii^{1*}, Daisy Hochberg¹, Megan Quimby¹, Bonnie Wong¹, Scott McGinnis², Bradford C
6 Dickerson^{1,3,4}, Deepti Putcha¹

7

8 ¹Frontotemporal Disorders Unit, Department of Neurology, Massachusetts General Hospital and Harvard Medical
9 School, Boston, MA 02114

10 ²Center for Brain Mind Medicine, Department of Neurology, Brigham & Women's Hospital, Boston, MA 02115

11 ³Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA 02129

12 ⁴Alzheimer's Disease Research Center, Massachusetts General Hospital, Charlestown, MA 02129

13

14 ***Correspondence:**

15 Neguine Rezaii

16 nrezaii@mgh.harvard.edu

17

18

19 **keywords:** posterior cortical atrophy, language indicators, visuospatial impairment, semantic
20 processing, natural language processing

21 **Abstract**

22 **Introduction:** Posterior Cortical Atrophy (PCA) is a syndrome characterized by a progressive decline in
23 higher-order visuospatial processing, leading to symptoms such as space perception deficit,
24 simultanagnosia, and object perception impairment. While PCA is primarily known for its impact on
25 visuospatial abilities, recent studies have documented language abnormalities in PCA patients. This study
26 aims to delineate the nature and origin of language impairments in PCA, hypothesizing that language
27 deficits reflect the visuospatial processing impairments of the disease.

28 **Methods:** We compared the language samples of 25 patients with PCA with age-matched cognitively
29 normal (CN) individuals across two distinct tasks: a visually-dependent picture description and a visually-
30 independent job description task. We extracted word frequency, word utterance latency, and spatial
31 relational words for this comparison. We then conducted an in-depth analysis of the language used in the
32 picture description task to identify specific linguistic indicators that reflect the visuospatial processing
33 deficits of PCA.

34 **Results:** Patients with PCA showed significant language deficits in the visually-dependent task,
35 characterized by higher word frequency, prolonged utterance latency, and fewer spatial relational words,
36 but not in the visually-independent task. An in-depth analysis of the picture description task further
37 showed that PCA patients struggled to identify certain visual elements as well as the overall theme of the
38 picture. A predictive model based on these language features distinguished PCA patients from CN
39 individuals with high classification accuracy.

40 **Discussion:** The findings indicate that language is a sensitive behavioral construct to detect visuospatial
41 processing abnormalities of PCA. These insights offer theoretical and clinical avenues for understanding
42 and managing PCA, underscoring language as a crucial marker for the visuospatial deficits of this atypical
43 variant of Alzheimer's disease.

44 Introduction

45 Posterior cortical atrophy (PCA) is a clinico-radiological syndrome characterized by a
46 progressive decline in higher-order visuospatial processing with relative preservation in other cognitive
47 domains at initial presentation (1–3). Common visuospatial symptoms of the syndrome include impaired
48 object and space perception, simultanagnosia, environmental agnosia, and visual field defects (2). From
49 the neuroimaging perspective, the syndrome is associated with atrophy, hypometabolism, and usually tau
50 deposition in posterior parietal, occipital, and temporo-occipital cortices (4,5). As the majority of PCA
51 cases are due to underlying Alzheimer’s pathology, PCA is also referred to as the visual variant of
52 Alzheimer’s disease (AD) (6,7). While the diagnostic criteria for PCA indicate preserved functions in
53 cognitive domains outside of visuospatial processing at symptom onset, a growing literature has
54 documented language abnormalities in PCA emerging early in the course of the illness. Specifically,
55 impaired category fluency and confrontation naming have been documented on formal
56 neuropsychological assessments (8–11). Language abnormalities are also evident during spontaneous
57 speech, such as using higher frequency words and slowed speech rate (number of words per minute) (12).
58 These emerging observations suggest that there is still much to be understood about the nature and origin
59 of language impairments in PCA.

60 The specific types of language abnormalities observed in PCA may be related to the network
61 dysfunction that supports lexicosemantic retrieval, as has been previously postulated across the
62 phenotypic spectrum of AD (10). Another possible explanation is that the language abnormalities
63 observed in PCA may stem from the visuospatial impairments central to the syndrome, rather than
64 representing a primary language deficit. A large body of research supports close relationships between the
65 visual processing of objects and the amodal semantic processing required for retrieving the names of
66 those objects (13–15). Recently, it has been shown that anterior to each region that is selective for the
67 visual processing of a given category in the visual cortex, there is a corresponding area selective to its
68 linguistic processing (16). This anatomical and functional configuration suggests that the anterior border
69 of the visual cortex acts as a convergence zone where information from the unimodal visual system enters
70 the amodal linguistic systems involved in linguistic retrieval. Therefore, the successful production of a
71 word that has visual attributes requires intact visual processing, essential for providing the information
72 needed to retrieve its corresponding linguistic representation (i.e., its name). Therefore, deficits in visual
73 processing would theoretically impede the production of words with visual attributes. If the
74 pathophysiology of language abnormalities in PCA involves disrupted visual processing, then it stands to
75 reason that tasks heavily dependent on visual processing will exhibit significant language impairments.
76 Conversely, tasks with minimal reliance on visual input should result in relatively intact language
77 performance.

78 In the current study, we sought to test this hypothesis by contrasting the language used in two
79 different speech samples as PCA participants described the Picnic scene from the Western Aphasia
80 Battery (17) (visually dependent) and their prior jobs (visually independent). In each speech sample, we
81 measured word frequency, word utterance latency, and the use of spatial relational words. For the picture
82 description task, we hypothesized that PCA patients would use higher frequency words (e.g., replacing
83 specific names of pictured items with superordinate words potentially including “thing”), have increased
84 word utterance latency due to object recognition difficulty, and use fewer spatial relational words such as
85 “into” or “underneath” compared to healthy individuals. For the non-visually dependent job description
86 task, we expected these linguistic features to be comparable between PCA patients and healthy
87 individuals.

88 Building on the hypothesis that speech patterns in visually dependent tasks reflect visuospatial
89 processing deficits, we next sought to identify linguistic markers of these challenges. Specifically, we
90 investigated which elements in the picnic scene presented particular retrieval difficulties and whether
91 PCA patients could intuitively grasp and articulate the overall theme of the scene, such as using the word
92 “picnic”. Due to the difficulty with visually integrating a scene (simultanagnosia) (9,18,19), we

93 hypothesized that PCA patients were less likely to verbalize the term “picnic” compared to healthy
94 individuals. Lastly, to address the clinical significance of this work, we used the language features derived
95 from the picture description task to develop a classifier aimed at distinguishing PCA patients from healthy
96 individuals and hypothesized a high degree of classification accuracy.

97

98 **Methods**

99 **Participants**

100 *PCA patients*

101 Twenty-five patients diagnosed with PCA were recruited from the Massachusetts General Hospital
102 (MGH) Frontotemporal Disorders Unit PCA program for this study (20). All but one was confirmed
103 amyloid positive (A+) and tau positive (T+) by either CSF analysis or amyloid and tau PET. The
104 remaining participant’s biomarker status is unknown due to a failed lumbar puncture. Each patient had
105 posterior cortical atrophy and/or hypometabolism, consistent with the typical neurodegeneration (N+) of
106 PCA. See Table 1 for demographic and clinical data. All participants received a standard clinical
107 evaluation comprising a structured history obtained from both participant and informant, comprehensive
108 neurological and psychiatric history, as well as neuropsychological assessment. See Table S1 for
109 neuropsychological profiles of the PCA cohort included in this study. The clinical formulation was
110 performed through a consensus conference by our multidisciplinary team of neurologists, psychiatrists,
111 neuropsychologists, and speech and language pathologists, with each patient classified based on all
112 available clinical information as having a 3-step diagnostic formulation of mild cognitive impairment or
113 dementia (Cognitive Functional Status), a specific Cognitive-Behavioral Syndrome, and a likely etiologic
114 neuropathologic diagnosis (21). Patients underwent neuroimaging sessions involving structural MRI, FTP
115 PET, and amyloid (PiB or FBB) PET scans of CSF analysis for A β and tau. A β positivity was determined
116 by a combination of visual read and mean amyloid PET signal extracted from a cortical composite region
117 of interest according to previously published procedures (22–24). Determination of tau and
118 neurodegeneration positivity was conducted by visual read using internal methods similar to published
119 work (25–27). This work was carried out according to The Code of Ethics of the World Medical
120 Association (Declaration of Helsinki) for experiments involving humans. All participants and their
121 caregivers provided informed consent in accordance with the protocol approved by the Mass General
122 Brigham Human Research Committee Institutional Review Board in Boston, Massachusetts. A speech
123 sample for the picture description task was acquired from all twenty-five PCA participants. Twenty-three
124 PCA participants also took part in the job description task.

125

126

127 *Cognitively normal (CN) individuals*

128 Twenty-nine CN participants (CN1) were enrolled through the Speech and Feeding Disorders Laboratory
129 at the MGH Institute of Health Professions to participate in the picture description task. These participants
130 passed a cognitive screen, were native English speakers, and had no history of neurologic injury or
131 developmental speech/language disorders. Twenty-six CN participants (CN2) were additionally recruited
132 through Amazon’s Mechanical Turk (MTurk) to describe their jobs. MTurk participants filled out the
133 short and validated version of the 12-item Everyday Cognition questionnaire, a questionnaire designed to
134 detect cognitive and functional decline (28). Only language samples from participants who were native
135 English speakers with no self-reported history of brain injury or speech/language disorder, either
136 developmental or acquired, were included in the analyses.

137

138 **Speech samples and data analysis**

139 Speech samples were collected under two conditions. For the visually dependent task, participants
140 described the Western Aphasia Battery—Revised (WAB-R) (17) Picnic Scene with the instruction to use
141 full sentences. For the visually independent task, participants were asked to describe what they did for
142 work. There were no time limits applied to either task. Autotranscription was done using Google Cloud

143 Speech-to-Text API for audio transcription (29), and manually verified by a research staff blinded to the
144 diagnosis.

145
146 *Speech sample analysis*

147 All feature extraction was performed automatically using Quantitext, a speech and text analysis toolbox
148 we developed in the Frontotemporal Disorders Unit of MGH, to produce a set of quantitative speech and
149 language metrics. The goal of developing this package is to increase the precision and objectivity of
150 language assessments along the goals described previously (30). The toolbox incorporates several natural
151 language processing toolkits and software such as Stanza (31) as well as text analysis libraries in R.
152 Quantitext receives audio or transcribed language samples as input and generates a number of metrics
153 such as word frequency, syntax frequency (32), content units (33), total units, efficiency of words (34),
154 and part-of-speech tags.

155
156 *Word frequency*

157 To measure word frequency, we used the Switchboard corpus (35), which consists of spontaneous
158 telephone conversations averaging six minutes in length spoken by more than 500 speakers of both sexes
159 from a variety of dialects of American English. We use this corpus to estimate word frequency in spoken
160 English, independently of the patient and control sample. The corpus contains 2,345,269 words. Here,
161 word frequency denotes the log frequency of content words (comprised of nouns, verbs, adjectives, and
162 adverbs).

163
164 *Word utterance latency and articulation rate*

165 Our analysis employed the Google Cloud Speech-to-Text API to ascertain word timestamps, pinpointing
166 the onset and offset for each spoken word within the audio recordings. Speech rate—often quantified as
167 the number of words spoken per minute—can vary based on factors such as word utterance latency and
168 the individual articulation rate of each word. To ensure a more granular and accurate interpretation of the
169 underlying phenomena, we sidestepped aggregated metrics like speech rate, focusing instead on
170 separately evaluating its constituent components. Word utterance latency is defined as the time interval
171 preceding the articulation of a word. This method was applied on all except for the very first word in each
172 sample, as the time to start the description task depends on multiple factors. Articulation rate measures the
173 number of syllables per second (36).

174
175 *Spatial relational words*

176 Relational words are automatically tagged by Stanza as “case”. For most words, the relational words are
177 spatial, for example, the word “under” in the sentence “I found the gem under my bed”. In our analysis,
178 we divided the number of relational words by the total words.

179
180 *Content units*

181 To determine which items within the picture posed greater challenges for PCA compared to CN
182 participants, we coded the visual items using content units. Content units are words with correct
183 information units that are intelligible in context and accurate about the picture or topic. Words do not
184 have to be used in a grammatically correct manner to be counted as content units (37). Each content unit
185 is only counted once, regardless of how many times it is mentioned in a sample. The morphological
186 variants were grouped within one single content unit. For example, the nouns ‘girl’ and ‘daughter’ are
187 considered the same content unit. Therefore, if one participant used both words (girl and daughter), they
188 would only be counted as one content unit. To specify content units, Quantitext first generates a Python
189 dictionary using a predefined set of words as previously described and then uses this dictionary to
190 automatically identify all content units in new texts it receives. Previously, we showed that the program
191 has an accuracy of 99.7% in identifying content units (33).

192
193 *Statistical analysis*

194 We used Welch Two Sample t-tests to compare the language features across the two groups. We
195 performed a pairwise Pearson correlation analysis on our dataset to investigate the relationships between
196 the likelihood of reporting each content unit and the group designation (with PCA dummy coded as 1 and
197 healthy controls as 0). For classification, we used a binary logistic regression model. We employed a
198 leave-one-out cross-validation (LOOCV) approach on our dataset to validate the model's performance. In
199 each iteration of the LOOCV, a single observation was set aside as the test data, and the remaining
200 observations were used to train the model. Alpha was set at 0.05.

201

202

203 Results

204 Language abnormalities in PCA are observed during picture description but not job description.

205 We first compared the speech samples of PCA participants describing the WAB Picnic Scene to the CN1
206 group to determine language abnormalities in this task. We also compared the speech sample from the job
207 description task between the PCA and CN2 groups. We used Welch Two Sample t-tests to compare the
208 means of the following features across the two groups (Figure 1).

209

210 *Picture description task*

211 Patients with PCA used higher frequency words (i.e., more commonly used words) (mean = 6.50 ± 0.53)
212 compared to healthy controls (mean = 5.85 ± 0.40) ($t(44.28) = -5.02$, $p < 0.001$). The time latency to the
213 onset of words was longer for patients with PCA (mean = 0.38 ± 0.54) compared to healthy controls
214 (mean = 0.03 ± 0.05) ($t(24.34) = -3.20$, $p = 0.004$). We found that patients with PCA used fewer spatial
215 relational terms (mean = 0.08 ± 0.03) compared to healthy controls (mean = 0.09 ± 0.02) ($t(45.55) = 2.35$,
216 $p = 0.023$).

217 There was a trend toward a slower articulation rate in PCA patients (mean = 2.92 ± 0.60) compared to
218 healthy controls (mean = 3.14 ± 0.35) ($t(33.88) = 1.616$, $p = 0.115$), suggesting that slower speech rate
219 may be primarily due to an increased word utterance latency rather than articulation rate.

220

221 *Job description task*

222 The frequency of content words in PCA patients (mean = 6.15 ± 0.72) was not different from healthy
223 controls (mean = 6.31 ± 0.72) ($t(46.27) = 0.79$, $p = 0.43$). Similarly, there was no significant difference in
224 word utterance latency between PCA patients (mean = 0.11 ± 0.22) and healthy controls (mean = $0.06 \pm$
225 0.06) ($t(25.24) = -1.11$, $p = 0.28$). No statistical difference was found in the use of relational words
226 between PCA patients (mean = 0.10 ± 0.04) and healthy individuals (mean = 0.09 ± 0.02) ($t(33.36) = -$
227 0.74 , $p = 0.47$). There was no difference in articulation rate in PCA patients (mean = 2.93 ± 0.60)
228 compared to healthy controls (mean = 3.11 ± 0.35) ($t(39.76) = 1.12$, $p = 0.269$).

229

230 Specific language indicators of visuospatial processing deficits of PCA can be extracted from the 231 picture description task.

232 We next probed the samples obtained from the picture description task to extract the specific language
233 features that reflect visuospatial impairment in PCA compared to healthy controls. First, we determined
234 the likelihood of mentioning each content unit by each diagnostic group. The picture consists of 32
235 content units, as shown in Figure 2. We performed a pairwise Pearson correlation analysis on our dataset
236 to investigate the relationships between the likelihood of reporting each content unit and the group
237 designation. As shown in Figure 2, we did not observe a uniform reduction in the likelihood of
238 mentioning each content unit in PCA. Instead, certain content units had a much lower chance of being
239 verbalized. Of all content units, "fisherman", a small, central feature of the WAB Picnic scene, was the
240 least likely to be mentioned by a patient with PCA compared with healthy controls ($r = -0.85$, $p < 0.001$).

241 A few content units had a numerically higher, though not statistically significant, likelihood of being
242 mentioned by patients with PCA compared to healthy individuals, such as "clouds" ($r = 0.18$, $p = 0.186$).

243 Figure 3 is the artistic rendering we developed to show the rate at which patients with PCA mention each

244 content unit. We then compared the total number of content units retrieved across the two groups.
245 Overall, PCA patients retrieved fewer content units (mean = 7.20 ± 5.63) compared to healthy individuals
246 (mean = 16.28 ± 4.41) ($t(45.21) = 6.52$, $p < 0.001$). Lastly, patients with PCA had a lower likelihood of
247 reporting the overall theme of the picture (i.e., mentioning the word “picnic”) (mean = 0.16 ± 0.37), than
248 healthy individuals did (mean = 0.90 ± 0.31) ($t(46.77)$, $p < 0.001$).

249 **Diagnostic Classification**

251 We used binary logistic regression to classify PCA and healthy individuals. Our predictor variables
252 consisted of word frequency, word utterance latency, relational words, the total number of content units,
253 and the probability of mentioning “picnic”. As the sixth variable, we included the probability of
254 mentioning “fisherman” because this content unit had the highest correlation with the group designation,
255 likely due to its visuospatial processing demands. The average accuracy of the model was 98.15% after
256 leave-one-out cross-validation. The average precision across all iterations was found to be 0.96, which
257 means that, on average, 96% of the predicted positive cases were actual positive cases. Moreover, the
258 model demonstrated an average recall of 1, indicating that it successfully identified all the positive cases
259 from the test data in each iteration. We also evaluated the performance of the model using a Receiver
260 Operating Characteristic (ROC) curve. The Area Under the Curve (AUC) was 1, indicating the perfect
261 discrimination ability of the model (Figure 4).

262 **Discussion**

264 Using computational linguistic methods, our study illuminated a distinction in PCA patients between
265 compromised language abilities in visually-dependent tasks and preserved language skills in visually-
266 independent contexts. At the theoretical level, we used this method to show that at least some of the
267 language abnormalities increasingly being identified in PCA are byproducts of visuospatial deficits
268 characteristic of this atypical AD syndrome. On the visually-independent job description task, the
269 characteristics of language production we measured here were not impaired in patients with PCA.
270 Translating our observations to clinical practice, we showed that computational linguistic analysis of a
271 simple picture description task robustly classified nearly all PCA patients as distinct from healthy
272 controls, supporting its value in clinical diagnostic evaluation.

274 Our work is consistent with studies showing rich connections between networks representing
275 information directly received from senses and information conveyed through spoken language (16,38,39).
276 Unimodal sensory information and abstract language information are combined at multiple points across
277 the cortex, such as inferior parietal lobule (comprising the angular and supramarginal gyri) and large
278 swaths of posterolateral temporal cortex (40,41), many of which can be affected in PCA. Therefore,
279 language abnormalities in PCA may arise for at least two potential reasons. First, the neurodegeneration
280 of PCA may extend beyond visuospatial areas to encompass regions involved in abstract language
281 processing. In support of this hypothesis, evidence suggests that the brain regions affected in PCA overlap
282 with those critical for word retrieval (42–46). Alternatively, language anomalies may arise as a
283 consequence of visuospatial deficits hindering the transfer of necessary sensory information for amodal
284 language processing.

285 While these two possibilities are not mutually exclusive, our results suggest that language
286 impairments might be largely secondary to visuospatial dysfunction. In our analysis comparing a variety
287 of speech and language properties of the narratives produced when PCA patients describe a complex
288 visual scene versus a recounting of their primary occupation from memory, we observed speech and
289 language impairments in only the visually dependent picture description task. We believe the increased
290 word frequency and word utterance latency in the picture description task reflect visual deficits in object
291 recognition. Similarly, the reduced use of spatial relational words may reflect the patients’ difficulty
292 processing spatial relations between elements of the picnic scene. The absence of abnormalities in word

293 frequency, word utterance latency, and relational words in the job description task provides evidence that
294 these language impairments do not stem from an intrinsic deficit in the language system in PCA. Since
295 most daily communication is a blend of visuospatial cognition, episodic memory, and other cognitive
296 domains, we anticipate that an analysis of everyday speech would reveal linguistic deficiencies
297 proportionate to the visuospatial load of its content. This expectation aligns with prior research reporting
298 linguistic anomalies in participants recounting their recent holiday, an account that naturally encompasses
299 the visuospatial processing of a recent event, such as where they went and what they saw (12). Another
300 consideration in interpreting these task differences is that the picture description task required the use of
301 specific linguistic elements representing the specific visual stimulus. It is also possible that when given
302 fewer constraints in the job description task, individuals had the freedom to choose a potentially more
303 familiar and more easily accessible language. Our explanation of the underlying language abnormalities
304 in PCA is consistent with findings that showed a striking discrepancy between visual and verbal
305 comprehension tasks in this population (47). Our conclusion is also synergistic with results reporting very
306 mild impairment in semantic memory in PCA, indicating that the apparent semantic impairment in these
307 conditions may be secondary to visual impairment (47).

308 Based on the observation that language is a sensitive indicator of visuospatial impairments of
309 PCA, we performed an in-depth content unit analysis of language elicited through the picture description
310 task. First, we measured the probability of verbalizing each content unit of the picture. The most
311 distinguishing feature between PCA patients and healthy individuals was the probability of mentioning
312 the “fisherman”. While 93% of healthy participants mentioned this content unit, only 8% of PCA patients
313 did so. This discrepancy could be attributed to the smaller size of this element in the picture. In addition,
314 multiple elements are superimposed in the location of this content unit. Numerically, though not
315 significantly, certain items, such as “clouds,” had a higher likelihood of being mentioned by PCA patients
316 than healthy individuals. This type of analysis provides a naturalistic way of identifying the visuospatial
317 elements that are particularly challenging for PCA patients and could help clinicians devise rehabilitative
318 strategies to alleviate these challenges. Moreover, we observed that PCA patients often missed describing
319 the overarching theme of the image (“picnic”), even when they identified certain components related to it
320 (“basket”). We believe this finding represents the effects of simultagnosia, which prevents many PCA
321 patients from grasping the integrated theme of a composite visual entity.

322 Finally, when specific quantitative language metrics were employed to differentiate PCA patients
323 from healthy individuals, our predictive model achieved a high level of performance, as evidenced by an
324 AUC of 1 and an accuracy rate of 98.15%. Automating this linguistic evaluation from an easily acquired
325 speech sample would facilitate the integration of measures like this into digital healthcare infrastructure,
326 which a wide array of healthcare providers could potentially use once trained. This development has
327 potential implications for improving early diagnosis of PCA as well as for monitoring response to
328 disease-modifying, rehabilitative, or other therapies in this underserved atypical variant of AD.

329

330

331

332

333 **Conflict of interest**

334 The authors have no conflicts of interest to report.

335

336 **Author contributions**

337 NR, DP contributed to the conception, design of the study, and data analyses. NR, DP, and BCD drafted
338 the manuscript. NR, DP, DH, MQ, BW, SM, and BCD participated in the data collection and editing of
339 the manuscript.

340

341 **Funding**

342 This research was supported by NIH grants K23 AG065450, RF1 NS131395, R01 DC014296, R21
343 DC019567, R21 AG073744, P50 AG005134, and P30 AG062421, Alzheimer's Association grant
344 23AACSF-1029880 and by the Tommy Rickles Chair in Primary Progressive Aphasia Research.

345

346 **Data availability**

347 Language data and codes will be available upon request to Dr. Bradford C Dickerson.

348

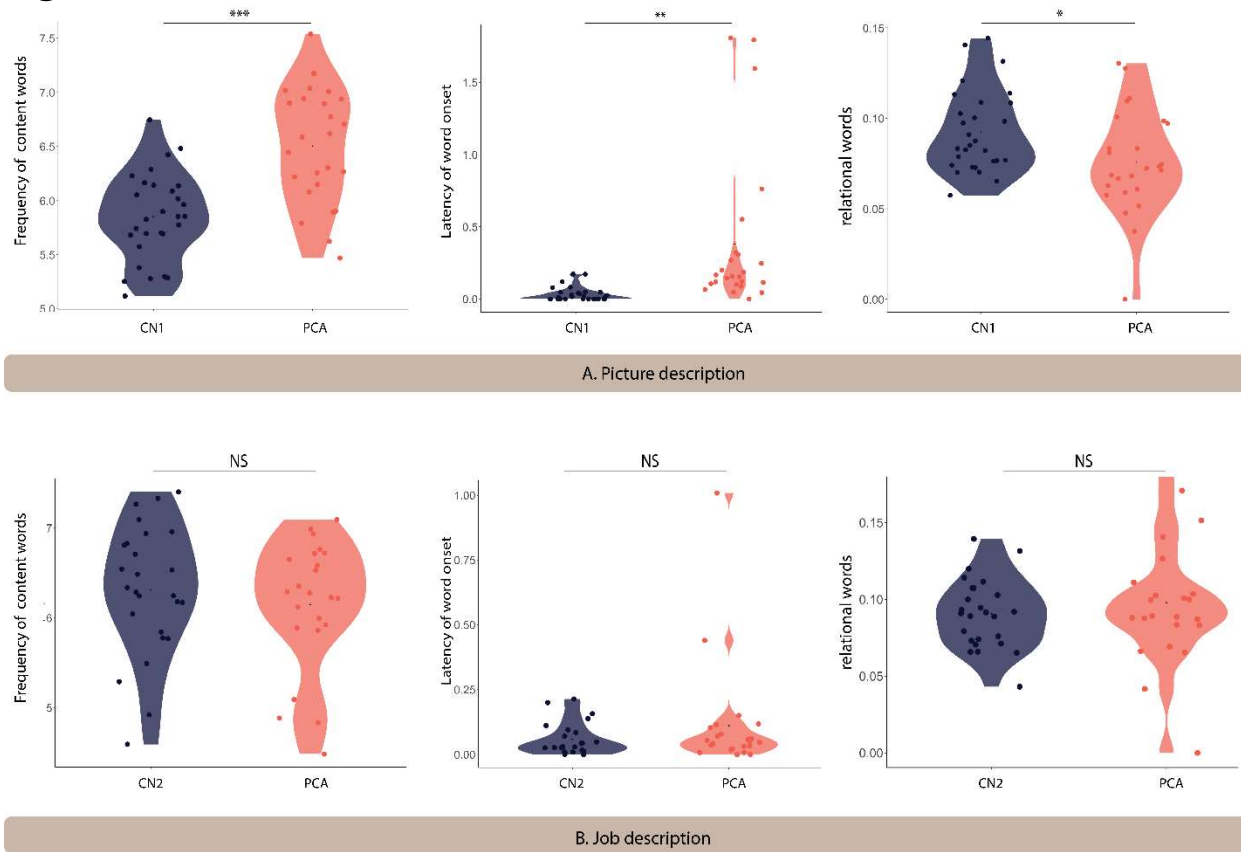
349 **References**

- 350 1. Benson DF, Davis RJ, Snyder BD. Posterior cortical atrophy. *Arch Neurol*. 1988 Jul;45(7):789–93.
- 351 2. Crutch SJ, Schott JM, Rabinovici GD, Murray M, Snowden JS, van der Flier WM, et al. Consensus
352 classification of posterior cortical atrophy. *Alzheimers Dement J Alzheimers Assoc*. 2017
353 Aug;13(8):870–84.
- 354 3. Renner JA, Burns JM, Hou CE, McKeel DW, Storandt M, Morris JC. Progressive posterior cortical
355 dysfunction: a clinicopathologic series. *Neurology*. 2004 Oct 12;63(7):1175–80.
- 356 4. Lehmann M, Crutch SJ, Ridgway GR, Ridha BH, Barnes J, Warrington EK, et al. Cortical thickness
357 and voxel-based morphometry in posterior cortical atrophy and typical Alzheimer's disease.
358 *Neurobiol Aging*. 2011 Aug;32(8):1466–76.
- 359 5. Whitwell JL, Jack CR, Kantarci K, Weigand SD, Boeve BF, Knopman DS, et al. Imaging correlates
360 of posterior cortical atrophy. *Neurobiol Aging*. 2007 Jul;28(7):1051–61.
- 361 6. Levine DN, Lee JM, Fisher CM. The visual variant of Alzheimer's disease: a clinicopathologic case
362 study. *Neurology*. 1993 Feb;43(2):305–13.
- 363 7. Alladi S, Xuereb J, Bak T, Nestor P, Knibb J, Patterson K, et al. Focal cortical presentations of
364 Alzheimer's disease. *Brain J Neurol*. 2007 Oct;130(Pt 10):2636–45.
- 365 8. McMonagle P, Deering F, Berliner Y, Kertesz A. The cognitive profile of posterior cortical atrophy.
366 *Neurology*. 2006 Feb 14;66(3):331–8.
- 367 9. Tang-Wai DF, Graff-Radford NR, Boeve BF, Dickson DW, Parisi JE, Crook R, et al. Clinical,
368 genetic, and neuropathologic characteristics of posterior cortical atrophy. *Neurology*. 2004 Oct
369 12;63(7):1168–74.
- 370 10. Putcha D, Dickerson BC, Brickhouse M, Johnson KA, Sperling RA, Papp KV. Word retrieval across
371 the biomarker-confirmed Alzheimer's disease syndromic spectrum. *Neuropsychologia*. 2020 Mar
372 16;140:107391.
- 373 11. Putcha D, McGinnis SM, Brickhouse M, Wong B, Sherman JC, Dickerson BC. Executive
374 dysfunction contributes to verbal encoding and retrieval deficits in posterior cortical atrophy. *Cortex*
375 *J Devoted Study Nerv Syst Behav*. 2018 Sep;106:36–46.
- 376 12. Crutch SJ, Lehmann M, Warren JD, Rohrer JD. The language profile of posterior cortical atrophy. *J*
377 *Neurol Neurosurg Psychiatry*. 2013 Apr;84(4):460–6.
- 378 13. Binder JR, Desai RH. The Neurobiology of Semantic Memory. *Trends Cogn Sci*. 2011
379 Nov;15(11):527–36.
- 380 14. Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL. Natural speech reveals the semantic
381 maps that tile human cerebral cortex. *Nature*. 2016 Apr;532(7600):453–8.
- 382 15. Aliko S, Wang B, Small SL, Skipper JI. The entire brain, more or less is at work: 'Language regions'
383 are artefacts of averaging [Internet]. *bioRxiv*; 2023 [cited 2023 Sep 27]. p. 2023.09.01.555886.
384 Available from: <https://www.biorxiv.org/content/10.1101/2023.09.01.555886v1>
- 385 16. Popham SF, Huth AG, Bilenko NY, Deniz F, Gao JS, Nunez-Elizalde AO, et al. Visual and linguistic
386 semantic representations are aligned at the border of human visual cortex. *Nat Neurosci*. 2021
387 Nov;24(11):1628–36.
- 388 17. Kertesz A, Kertesz A, Raven JC, PsychCorp (Firm). WAB-R: Western Aphasia Battery-Revised. San
389 Antonio, TX: PsychCorp; 2007.

- 390 18. Cui Y, Liu Y, Yang C, Cui C, Jing D, Zhang X, et al. Brain structural and functional anomalies
391 associated with simultanagnosia in patients with posterior cortical atrophy. *Brain Imaging Behav.*
392 2022;16(3):1148–62.
- 393 19. Singh TD, Josephs KA, Machulda MM, Drubach DA, Apostolova LG, Lowe VJ, et al. Clinical, FDG
394 and amyloid PET imaging in posterior cortical atrophy. *J Neurol.* 2015 Jun;262(6):1483–92.
- 395 20. Wong B, Lucente DE, MacLean J, Padmanabhan J, Quimby M, Brandt KD, et al. Diagnostic
396 evaluation and monitoring of patients with posterior cortical atrophy. *Neurodegener Dis Manag.*
397 9(4):217–39.
- 398 21. Dickerson BC, McGinnis SM, Xia C, Price BH, Atri A, Murray ME, et al. Approach to atypical
399 Alzheimer’s disease and case studies of the major subtypes. *CNS Spectr.* 2017 Dec;22(6):439–49.
- 400 22. Cho H, Mundada NS, Apostolova LG, Carrillo MC, Shankar R, Amuiri AN, et al. Amyloid and tau-
401 PET in early-onset AD: Baseline data from the Longitudinal Early-onset Alzheimer’s Disease Study
402 (LEADS). *Alzheimers Dement.* 2023 Sep 10;alz.13453.
- 403 23. Rabinovici GD, Furst AJ, Alkalay A, Racine CA, O’Neil JP, Janabi M, et al. Increased metabolic
404 vulnerability in early-onset Alzheimer’s disease is not related to amyloid burden. *Brain.* 2010 Feb
405 1;133(2):512–28.
- 406 24. Villeneuve S, Rabinovici GD, Cohn-Sheehy BI, Madison C, Ayakta N, Ghosh PM, et al. Existing
407 Pittsburgh Compound-B positron emission tomography thresholds are too high: statistical and
408 pathological evaluation. *Brain.* 2015 Jul 1;138(7):2020–33.
- 409 25. Fleisher AS, Pontecorvo MJ, Devous MD Sr, Lu M, Arora AK, Trucchio SP, et al. Positron
410 Emission Tomography Imaging With [18F]flortaucipir and Postmortem Assessment of Alzheimer
411 Disease Neuropathologic Changes. *JAMA Neurol.* 2020 Jul 1;77(7):829–39.
- 412 26. Rabinovici GD, Rosen HJ, Alkalay A, Kornak J, Furst AJ, Agarwal N, et al. Amyloid vs FDG-PET
413 in the differential diagnosis of AD and FTLD. *Neurology.* 2011 Dec 6;77(23):2034–42.
- 414 27. Sonni I, Lesman Segev OH, Baker SL, Iaccarino L, Korman D, Rabinovici GD, et al. Evaluation of a
415 visual interpretation method for tau-PET with 18F-flortaucipir. *Alzheimers Dement Diagn Assess Dis*
416 *Monit.* 2020 Jan;12(1):e12133.
- 417 28. Tomaszewski Farias S, Mungas D, Harvey DJ, Simmons A, Reed BR, Decarli C. The measurement
418 of everyday cognition: development and validation of a short form of the Everyday Cognition scales.
419 *Alzheimers Dement J Alzheimers Assoc.* 2011 Nov;7(6):593–601.
- 420 29. Cloud Speech-to-Text API v1 - Package cloud.google.com/go/speech/apiv1 (v1.19.0) | Go client
421 library | Google Cloud [Internet]. [cited 2023 Oct 4]. Available from:
422 <https://cloud.google.com/go/docs/reference/cloud.google.com/go/speech/latest/apiv1>
- 423 30. Rezaii N, Wolff P, Price BH. Natural language processing in psychiatry: the promises and perils of a
424 transformative approach. *Br J Psychiatry.* 2022 May;220(5):251–3.
- 425 31. Qi P, Zhang Y, Zhang Y, Bolton J, Manning CD. Stanza: A Python Natural Language Processing
426 Toolkit for Many Human Languages. In: Proceedings of the 58th Annual Meeting of the Association
427 for Computational Linguistics: System Demonstrations [Internet]. Online: Association for
428 Computational Linguistics; 2020 [cited 2021 Nov 24]. p. 101–8. Available from:
429 <https://aclanthology.org/2020.acl-demos.14>
- 430 32. Rezaii N, Mahowald K, Ryskin R, Dickerson B, Gibson E. A syntax-lexicon trade-off in language
431 production. *Proc Natl Acad Sci U S A.* 2022 Jun 21;119(25):e2120203119.
- 432 33. Josephy-Hernandez S, Rezaii N, Jones A, Loyer E, Hochberg D, Quimby M, et al. Automated
433 analysis of written language in the three variants of primary progressive aphasia. *Brain Commun.*
434 2023 Aug 1;5(4):fcad202.
- 435 34. Rezaii N, Ren B, Quimby M, Hochberg D, Dickerson BC. Less is more in language production: an
436 information-theoretic analysis of agrammatism in primary progressive aphasia. *Brain Commun.* 2023
437 Apr 25;5:fcad136.
- 438 35. Godfrey JJ, Holliman EC, McDaniel J. SWITCHBOARD: telephone speech corpus for research and
439 development. In: Proceedings of the 1992 IEEE international conference on Acoustics, speech and

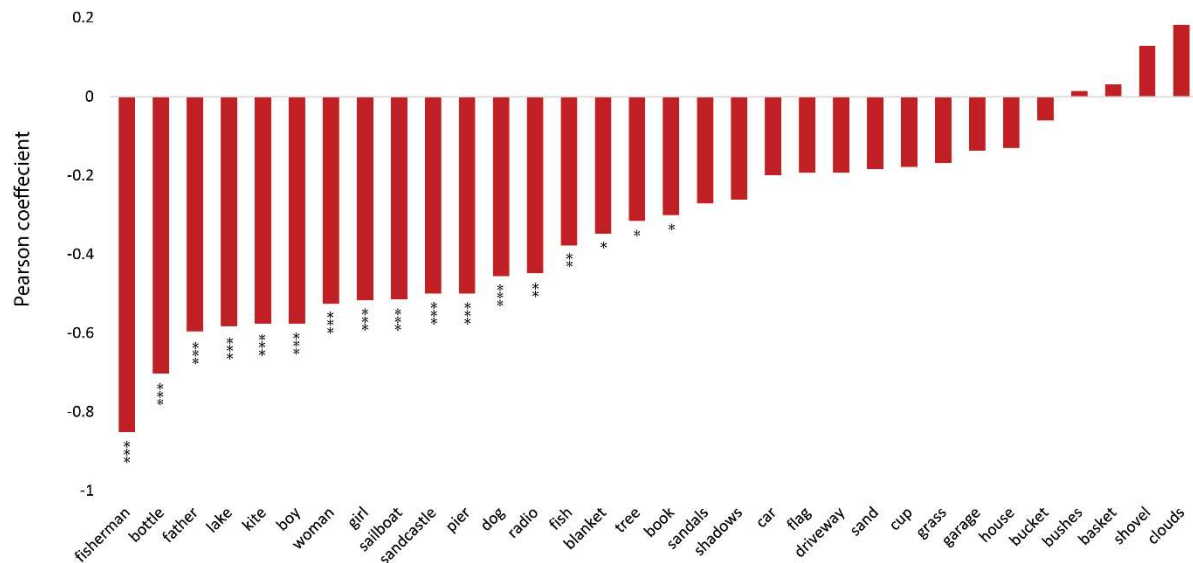
- 440 signal processing - Volume 1. San Francisco, CA, USA: IEEE Computer Society; 1992. p. 517–20.
441 (ICASSP'92).
- 442 36. Cordella C, Quimby M, Touroutoglou A, Brickhouse M, Dickerson BC, Green JR. Quantification of
443 motor speech impairment and its anatomic basis in primary progressive aphasia. *Neurology*. 2019
444 Apr 23;92(17):e1992–2004.
- 445 37. Nicholas LE, Brookshire RH. A system for quantifying the informativeness and efficiency of the
446 connected speech of adults with aphasia. *J Speech Hear Res*. 1993 Apr;36(2):338–50.
- 447 38. Ralph MAL, Jefferies E, Patterson K, Rogers TT. The neural and computational bases of semantic
448 cognition. *Nat Rev Neurosci*. 2017 Jan;18(1):42–55.
- 449 39. Damasio AR. Time-locked multiregional retroactivation: a systems-level proposal for the neural
450 substrates of recall and recognition. *Cognition*. 1989 Nov;33(1–2):25–62.
- 451 40. Mesulam MM. Behavioral Neuroanatomy. In: *Principles of Behavioral and Cognitive Neurology*.
452 Second Edition, Second Edition. Oxford, New York: Oxford University Press; 2000.
- 453 41. Devereux BJ, Clarke A, Marouchos A, Tyler LK. Representational Similarity Analysis Reveals
454 Commonalities and Differences in the Semantic Processing of Words and Objects. *J Neurosci*. 2013
455 Nov 27;33(48):18906–16.
- 456 42. Migliaccio R, Agosta F, Rascovsky K, Karydas A, Bonasera S, Rabinovici GD, et al. Clinical
457 syndromes associated with posterior atrophy: early age at onset AD spectrum. *Neurology*. 2009 Nov
458 10;73(19):1571–8.
- 459 43. Ossenkoppele R, Cohn-Sheehy BI, La Joie R, Vogel JW, Möller C, Lehmann M, et al. Atrophy
460 patterns in early clinical stages across distinct phenotypes of Alzheimer's disease. *Hum Brain Mapp*.
461 2015 Aug 11;36(11):4421–37.
- 462 44. Warren JD, Fletcher PD, Golden HL. The paradox of syndromic diversity in Alzheimer disease. *Nat*
463 *Rev Neurol*. 2012 Aug;8(8):451–64.
- 464 45. Leyton CE, Hodges JR, Piguet O, Ballard KJ. Common and divergent neural correlates of anomia in
465 amnesic and logopenic presentations of Alzheimer's disease. *Cortex J Devoted Study Nerv Syst*
466 *Behav*. 2017 Jan;86:45–54.
- 467 46. Vonk JMJ, Rizvi B, Lao PJ, Budge M, Manly JJ, Mayeux R, et al. Letter and Category Fluency
468 Performance Correlates with Distinct Patterns of Cortical Thickness in Older Adults. *Cereb Cortex N*
469 *Y N* 1991. 2019 Jun 1;29(6):2694–700.
- 470 47. Rogers TT, Ivanoiu A, Patterson K, Hodges JR. Semantic memory in Alzheimer's disease and the
471 frontotemporal dementias: A longitudinal study of 236 patients. *Neuropsychology*. 2006;20(3):319–
472 35.
473
474

475 **Figures**



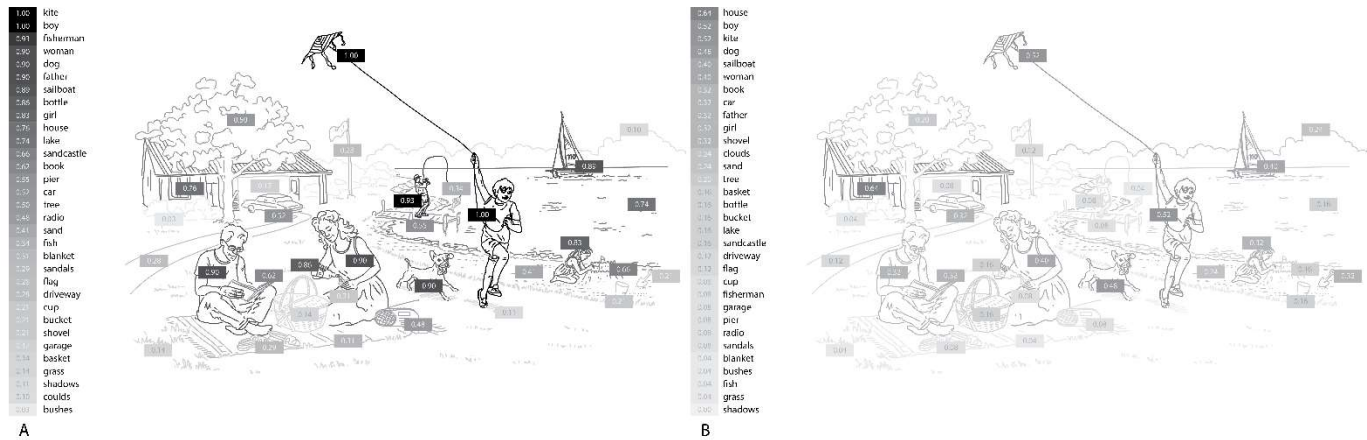
476
477
478
479
480
481

Figure 1. Language differs between PCA and healthy controls on the picture description task but not the job description task. Violin plots comparing the language features extracted from the picnic scene (A) and job (B) description tasks across healthy controls and PCA patients. *** denotes $p < 0.001$, ** indicates $0.001 < p < 0.01$, and * shows $0.01 < p < 0.05$. NS indicates not significantly different.



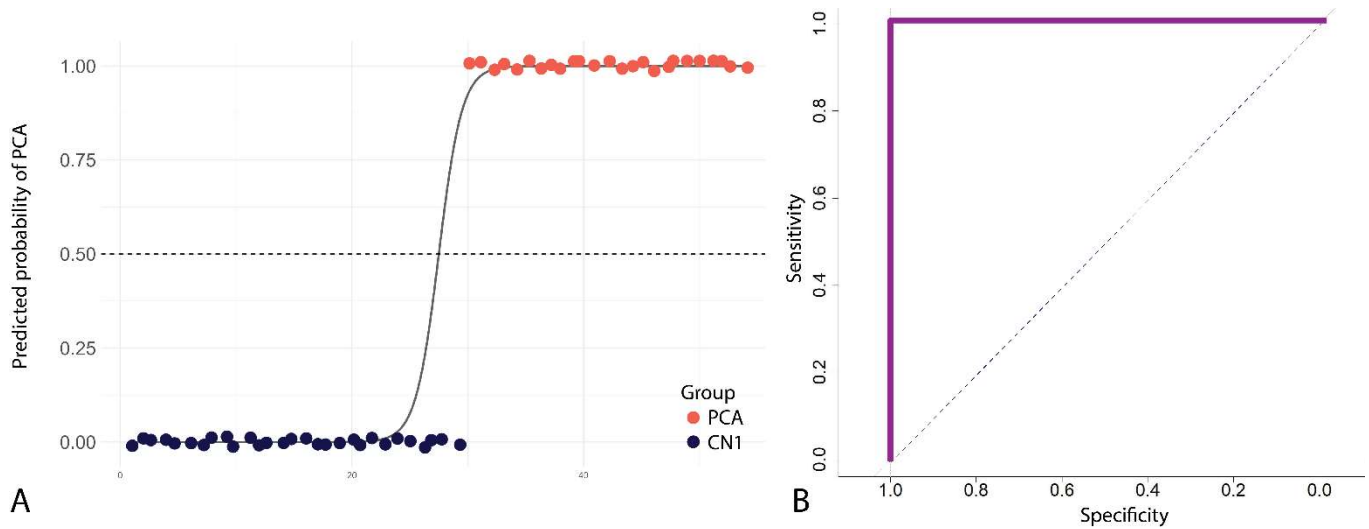
482
483 **Figure 2. PCA participants and healthy individuals described different content units on the picture**
484 **description task.** The Pearson correlation coefficients between the likelihood of reporting each content
485 unit in the picnic scene and the designated group. Negative values indicate that PCA patients have a lower
486 chance of mentioning the content unit compared to healthy individuals. *** denotes $p < 0.001$, **
487 indicates $0.001 < p < 0.01$, and * shows $0.01 < p < 0.05$. Bars without an asterisk are not significantly
488 different.
489

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .



491 **Figure 3. The likelihood of mentioning each content unit of the WAB Picnic Scene by PCA patients**
 492 **and healthy individuals.** The shading intensity of each item corresponds to its verbalization probability
 493 by participants, with darker elements indicating a higher likelihood of being mentioned by PCA patients
 494 and healthy individuals.

495
 496



497 A

498 **Figure 4. Diagnostic performance of a model that distinguishes PCA patients from healthy**
499 **individuals using linguistic features from the picture description task. (A)** The scatter plot shows the
500 predicted probabilities of classifying an individual as having PCA using selected linguistic features. Each
501 point represents an individual participant, with colors indicating the group. The sigmoid curve illustrates
502 the general trend of predicted probabilities. The dashed black line at the predicted probability of 0.5
503 serves as a decision threshold to classify between PCA patients and healthy individuals. Points are jittered
504 vertically for better visualization. **(B)** ROC curve illustrating the performance of the logistic regression
505 model in discriminating between PCA patients and healthy individuals, with an AUC value of 1
506 indicating the model's overall perfect accuracy.

507 **Table 1. Clinical characteristics of the A β + Posterior Cortical Atrophy (PCA) and Cognitively**
508 **Normal (CN) groups.** Means and Standard Deviations are reported for continuous variables. MoCA =
509 Montreal Cognitive Assessment. CDR = Clinical Dementia Rating Scale. SOB = Sum of Box scores.
510

Demographics	PCA (N = 25)	CN1 (N = 29)	CN2 (N = 26)
Age (years)	68.4 \pm 7.8	65.3 \pm 8.4	61.8 \pm 7.8
Sex (M/F)	13/12	13/16	9/17
Education (years)	17.2 \pm 2.1	15.7 \pm 1.0	14.8 \pm 1.8
Handedness (R/L/Ambidextrous)	23/0/2	20/8/1	20/3/3/
MoCA	14.6 \pm 7.8	--	
CDR	CDR 0 (N=1) CDR 0.5 (N=11) CDR 1 (N=10) CDR 2 (N=3)	--	
CDR-SOB	4.5 \pm 3.1	--	

511