Using Artificial Intelligence to detect PPA and discriminate among PPA subtypes from real-world speech

MASSACHUSETTS GENERAL HOSPITAL
FRONTOTEMPORAL DISORDERS UNIT

Gent Celaj, BA[1], Neguine Rezaii, M.D[1], Daisy Hochberg, MS, CCC-SLP[1], Megan Quimby MS, CCC-SLP[1], Nneka Watson, Bradford C. Dickerson, M.D*[1], Bonnie Wong Ph.D*[1, 2]

[1]Frontotemporal Disorders Unit, Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA,
(2) Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA
* Indicates Senior Author

HARVARD MEDICAL SCHOOL

BACKGROUND

- Advancements in artificial intelligence (AI) and natural language processing (NLP) have made it a valuable tool for diagnostic and clinical applications.
- Rezaii et al. (2024) found that AI-driven NLP tools are effective in identifying linguistic features that discriminate Primary Progressive Aphasia (PPA) subtypes from a picture description task.
- The current study aimed to replicate and extend these findings using an ecologically-valid language sample often derived from the clinical question, "What did you do for work?" (WDYDFW?).
- We hypothesized that key linguistic features identified by Rezaii and others in previous studies would differentiate PPA subtypes.

MATERIALS AND METHODS

- Speech samples from the WDYDFW? prompt were collected from 74 participants (58 PPA participants and 16 healthy controls).
- Speech samples were transcribed and manually reviewed for accuracy.
- Language markers previously identified by Rezaii et al. as discriminating PPA subtypes were extracted from the samples using Quantitext, an NLP-based toolbox that was developed to extract language markers. Ten linguistic features were extracted from each language sample, including:
  - average sentence length (average # of words per sentence)
  - word frequency (commonality in the English language of words used)
  - content words (words that provide meaning, i.e., nouns, verbs, adjectives, adverbs)
- All feature extractions was performed blind to diagnosis.
- One-way ANOVAs were conducted with linguistic features as the dependent variables and syndrome as the between-subject factor. Post-hoc pairwise comparisons were performed to assess group differences.

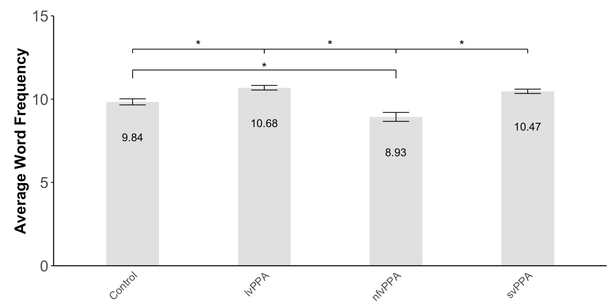
RESULTS

Table 1. Participant demographic and clinical characteristics

Demographic	nfvPPA (n = 21)	lvPPA (n = 20)	svPPA (n = 17)	Controls (n = 16)
Age (years)	70.96 (8.04)	69.37 (5.95)	67.46 (8.19)	65.56 (4.95)
Education (years)	16.07 (2.62)	16.67 (3.49)	15.96 (2.14)	14.88 (1.63)
Race (% White)	1.00	1.00	0.82	NA
Sex (M:F)	9:12	12:8	7:9	6:10
PASS Sum of Boxes (SoB)	5.28 (2.78)	4.14 (2.38)	4.85 (1.89)	NA

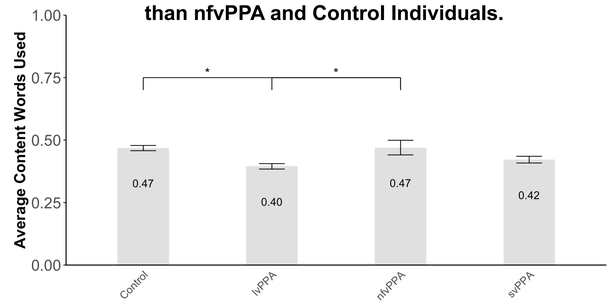
Values represent means with standard deviations in parentheses. nfvPPA = non-fluent variant Primary Progressive Aphasia; lvPPA = logopenic variant Primary Progressive Aphasia; svPPA = semantic variant Primary Progressive Aphasia; MoCA = Montreal Cognitive Assessment; NA = not available. PASS = Progressive Aphasia Severity Scale, a composite score that quantifies language impairment in PPA across multiple domains of speech and language function[3]; higher score=higher degree of impairment. Missing data by group: Years of Education – missing for 1 nfvPPA and for 2 svPPA participants; PASS SoB– missing for 3 nfvPPA and 2 lvPPA participants.

Figure 1. nfvPPA Individuals Produced Lower-Frequency Words Compared to Controls, lvPPA, and svPPA. Controls Used Lower-Frequency Words than lvPPA.



One-way ANOVA analyses revealed that nfvPPA participants used lower-frequency words in response to the WDYDFW? prompt compared to controls (F(3,70) =17.395, p = .01,), lvPPA (p < .001), and svPPA (p < .001). Controls used lower frequency words than lvPPA (F(3,70)=17.395, p = .02) Error bars represent one standard error of the mean.

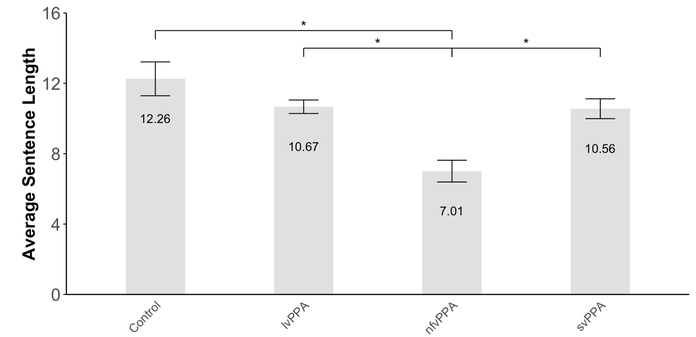
Figure 2. lvPPA Participants Produced Fewer Content Words than nfvPPA and Control Individuals.



One-way ANOVA analyses revealed that lvPPA produced a lower average proportion of content words used per sentence than nfvPPA (F(3,70) = 3.77, p < 0.025) and controls (p < .05)se to the WDYDFW? prompt. Error bars represent one standard error

RESULTS, CONT'D

Figure 3. nfvPPA Participants Produced Significantly Shorter Sentences than Other PPA Subtypes and Controls.



One-way ANOVA revealed a significant group effect (F(3,70) = 12.583, p < .001). Post-hoc analyses revealed that nfvPPA participants produced significantly shorter sentences than controls (p = .01), lvPPA (p < .001), and svPPA (p < .001). Error bars represent one standard error of the mean.

CONCLUSIONS

- NLP has the potential to discriminate among PPA subtypes using unstructured language samples commonly elicited in routine clinical office settings.
- This study underscores the potential of AI tools to classify atypical cognitive-behavioral syndromes.
- Analyses are ongoing with a larger dataset to assess the impact of symptom severity (CDR) on the presentation of linguistic features and to validate findings against structured language samples.
- Future analyses will explore whether these linguistic markers can extend to other atypical dementias, supporting broader clinical utility of NLP-based analysis.

REFERENCES & ACKNOWLEDGEMENTS

1. Rezaii, N., Hosseini, M., Sajjadi, S. A., Patterson, K., & Nili, H. (2022). *A syntax–lexicon trade-off in language production. Proceedings of the National Academy of Sciences*, 119(26), e2116560119. https://www.pnas.org/doi/10.1073/pnas.2120203119#supplementary-materials
2. Rezaii, N., Asgari, M., Sair, H. I., & Hillis, A. E. (2024). *Artificial intelligence classifies primary progressive aphasia from connected speech. Brain*, 147(9), 3070–3083. https://academic.oup.com/brain/article/147/9/3070/7697840
3. Sapolsky, D., Domoto-Reilly, K., & Dickerson, B. C. (2014). Use of the Progressive Aphasia Severity Scale (PASS) in monitoring speech and language status in PPA. *Aphasiology*, *28*(8–9), 993–1003. https://pmc.ncbi.nlm.nih.gov/articles/PMC4235969/

This research was supported by NIH grant R21 242832.

We gratefully acknowledge our participants and their families, whose partnership made this research possible.

Corresponding author: gcelaj@mgh.harvard.edu