ABSTRACT

Background

Reduced syntactic complexity is unique to the nonfluent-agrammatic variant of PPA (naPPA) compared to the semantic variant (svPPA) or the logopenic variant (lvPPA). However, there are no widely agreed, objective methods for quantifying syntactic complexity. Current methods either consider various syntactic features in isolation, or they use indirect measures as a proxy for syntactic complexity. Here, we propose a computational method for quantifying syntactic complexity in natural speech, which directly considers multiple syntactic features and combines them into a composite score. We test the sensitivity of the composite score in monitoring disease progression in naPPA in comparison to other PPA phenotypes. We examined cerebrospinal fluid (CSF) for additional biological validation. We examined two alternative indirect metrics that have been associated with syntactic complexity – namely, syntactic frequency and sentence length – and compared them with our novel score.

Methods

Speech samples of picture descriptions were collected from people with naPPA ($n = 35$, 50% males, age 70.0 ± 8), svPPA ($n = 37$, 49% males, age 64.2 ± 7) and lvPPA ($n = 33$, 49% males, age 67.6 ± 9) and from healthy controls (HC; $n = 36$, 31% males, age 68.6 ± 8). Nine syntactic features were automatically identified and tallied from the transcripts. Using principal component analysis (PCA), we calculated a syntactic complexity score that reflects the shared variability across the extracted syntactic features. After verifying group differences, we tested change over time using mixed effect models on a subset of the participants with follow-up recordings ($N = 49$). For biological validation, we examined the association of syntactic complexity with CSF concentration of neurofilament light chain (NfL).

Results

naPPA scored lower than HC ($\beta = 1.19$, CI = [0.5, 1.9]) and the other PPA groups (lvPPA: $\beta = 1.22$, CI = [0.5, 1.9]; svPPA: $\beta = 1.00$, CI = [0.3, 1.7]). Only in naPPA did syntactic complexity decrease over time ($\beta = -0.045$, CI = [−0.08, −0.01]) and lower scores associate with increased CSF NfL concentration ($\beta = -0.44$, CI = [−0.87, −0.003]). Sentence length also showed a longitudinal decline, but to a smaller extent ($\beta = -0.03$, CI

= [−0.06, −0.0009]). Syntactic frequency showed neither a longitudinal decline nor an association with NfL.

Discussion

Our novel syntactic score, derived automatically from recorded naturalistic speech, may directly capture agrammatism and can be used as a clinical outcome assessment tool to capture worsening symptoms in naPPA. Sentence length can be a useful proxy for syntactic complexity.