Title

Connected speech markers of amyloid burden in primary progressive aphasia

Permalink

https://escholarship.org/uc/item/7sh7q5ts

Authors

Slegers, Antoine
Chafouleas, Geneviève
Montembeault, Maxime
et al.

Publication Date

2021-12-01

DOI

10.1016/j.cortex.2021.09.010

Peer reviewed

HHS Public Access

Author manuscript

Cortex. Author manuscript; available in PMC 2022 December 01.

Connected speech markers of amyloid burden in primary progressive aphasia

Antoine Slegers[a,b], **Geneviève Chafouleas**[c], **Maxime Montembeault**[d], **Christophe Bedetti**[a,b], **Ariane E. Welch**[d], **Gil D. Rabinovici**[d], **Philippe Langlais**[c], **Maria L. Gorno-Tempini**[d], **Simona M. Brambati**[a,b,e,*]

[a]Department of Psychology, Université de Montréal, Canada

[b]Centre de Recherche de L'Institut Universitaire de Gériatrie de Montréal (CRIUGM), Montréal, Canada

[c]Department of Computer Science and Operational Research, Université de Montréal, Montréal, Canada

[d]Memory and Aging Center, University of California, San Francisco, San Francisco, CA, USA

[e]Centre de recherche du Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Île-de-Montréal, Montréal, Québec, Canada

Abstract

Introduction: Positron emission tomography (PET) amyloid imaging has become an important part of the diagnostic workup for patients with primary progressive aphasia (PPA) and uncertain underlying pathology. Here, we employ a semi-automated analysis of connected speech (CS) with a twofold objective. First, to determine if quantitative CS features can help select primary progressive aphasia (PPA) patients with a higher probability of a positive PET amyloid imaging result. Second, to examine the relevant group differences from a clinical perspective.

Methods: 117 CS samples from a well-characterised cohort of PPA patients who underwent PET amyloid imaging were collected. Expert consensus established PET amyloid status for each patient, and 40% of the sample was amyloid positive.

Results: Leave-one-out cross-validation yields 77% classification accuracy (sensitivity: 74%, specificity: 79%).

*Corresponding author. Centre de Recherche de L'Institut Universitaire de Gériatrie de Montréal (CRIUGM), 4565 chemin Queen-Mary, Montréal H3W 1W5, Canada. simona.maria.brambati@umontreal.ca (S.M. Brambati).

Discussion: Our results confirm the potential of CS analysis as a screening tool. Discriminant CS features from lexical, syntactic, pragmatic, and semantic domains are discussed.

1. Introduction

Primary progressive aphasias (PPA) are a group of neurodegenerative clinical syndromes characterized by the insidious and gradual onset of predominantly language-related impairments, of which three clinical variants are recognized: the non-fluent (nfvPPA), semantic (svPPA), and logopenic (lvPPA) variants (Gorno-Tempini et al., 2011). Among all patients who meet the criteria for PPA, up to 55% present with amyloidosis (Villarejo-Galende et al., 2017). AD pathology is observed in the majority of lvPPA cases (Bergeron et al., 2018; Montembeault, Brambati, Gorno-Tempini, & Migliaccio, 2018), a clinical syndrome characterized by profound word retrieval difficulties in spontaneous speech and sentence repetition deficits (Gorno-Tempini et al., 2011; McKhann et al., 2011). The prevalence of AD pathology in lvPPA cases is variable across different expert centers, ranging from 57% to 100% (Bergeron et al., 2018). The detection of amyloidosis in PPA samples represents an important and current issue for treatment and enrollment in clinical trials (Rabinovici et al., 2008; Rogalski et al., 2016).

Amyloid imaging positivity is established in vivo with positron emission tomography (PET) by the standardized uptake value ratios (SUVR) of a radioligand (e.g., Pittsburgh Compound B, Florbetapir, Florbetaben, or Flutemetamol). Appropriate use criteria state that amyloid imaging is indicated for atypical presentations of AD such as lvPPA (Dubois et al., 2010; Johnson et al., 2013) and in vivo amyloid PET measures have shown high correlation with post-mortem histopathological reports (Fink et al., 2020; Santose–Santos et al., 2018; Tan et al., 2017). However, amyloid PET scanning is 1) invasive, 2) remains prohibitively expensive for most care settings and is 3) incompatible with telemedicine. There is thus a pressing need for cost-effective methods to select patients for PET amyloid imaging and novel disease-modifying treatments.

Current research focuses on the non-invasive detection of biomarkers in individuals most likely to benefit from a full clinical and biomarker assessment (Villemagne, Doré, Burnham, Masters, & Rowe, 2018). The potential of automated approaches to connected speech (CS) analysis in the search for early and reliable markers of neurodegeneration has been underlined (Battista, Salvatore, Berlingeri, Cerasa, & Castiglioni, 2020; Beltrami et al., 2018; Clarke, Foltz, & Garrard, 2020; Laske et al., 2015; Myszczynska et al., 2020).

Previous studies on CS in PPA have shown that the three clinical variants of PPA present specific CS profiles expressing their characteristic cognitive and speech impairments (Ash et al., 2013). In patients with lvPPA, CS is marked by false starts, hesitations, filled pauses and repairs, as well as a different rate of verbs, nouns and pronouns, syntactic errors and embeddings compared to svPPA or to nfvPPA groups (Ash et al., 2013; Boschi et al.,

2017; Wilson et al., 2010). The fact that lvPPA patients produce different CS features from other PPA variants and that they are disproportionately more likely to be amyloid-positive raises the possibility that CS could be used to screen for amyloid positivity. However, it is still unknown whether such a classification is possible and, if so, which CS feature combination enables accurate screening of PPA patients for amyloidosis without first relying on a syndromic variant classification, a complex diagnostic procedure that requires access to clinical expertise, batteries of cognitive/language tests and complementary instrumental measures that may not be available in all clinical and research settings.

The general hypothesis of the study is that CS features can discriminate PPA with or without amyloidosis and that the most discriminant CS features would be related to the core problem affecting spontaneous speech in these patients, namely word retrieval difficulties, i.e., (Gorno-Tempini et al., 2011; McKhann et al., 2011). To address this issue, we investigated 117 transcripts of picture descriptions by a cohort of PPA patients who also underwent carbon 11-labeled Pittsburgh Compound-B PET (PiB). From the transcripts, we extracted a large set of features covering lexical, syntactic, semantic, and pragmatic aspects of CS. Using the currently growing approach in the field (Clarke et al., 2020), we combined CS features available in advanced natural language processing tools with other selected CS features based on recent clinical evidence on patients with PPA and/or AD pathology (Boschi et al., 2017; Fraser et al., 2014; Kamath, Sutherland, & Chaney, 2020; Slegers, Filiou, Montembeault, & Brambati, 2018). First, we based a classification experiment on this set of features to identify which patients have PET-confirmed amyloid deposits and plaques. Second, we identified the set of discriminant features that differentiate PiB+ from PiB− transcripts at the group level.

2. Methods

2.1. PPA participant database

Data used in the preparation of this article were obtained from the UCSF Memory and Aging Center database. We retrospectively identified all participants who received a diagnosis of PPA, according to current diagnostic criteria (Gorno-Tempini et al., 2011). To be included in the study, participants must have a CS sample from the picture description task from the Western Aphasia Battery (Kertesz, 1982) and carbon 11-labeled Pittsburgh Compound-B PET imaging (PiB). The general exclusion criteria for all participants were as follows: native language other than English, developmental learning disabilities, history of a psychiatric disorder, history of traumatic brain injury, and uncorrected hearing or vision problems.

All participants provided written informed consent, which was approved by the Internal Review Board of the University of California, San Francisco.

2.2. Amyloid imaging: positron emission tomography with Pittsburgh Compound-B (PiB)

Carbon 11-labeled Pittsburgh Compound-B PET (PiB) was performed at Lawrence Berkeley National Laboratory as previously described (Villeneuve et al., 2015). Native space SUVRs were created by normalizing mean images (at 50–70 min post-injection) by mean activity

in cerebellum gray matter. Visual reads of native space SUVR images were performed by experienced investigators blinded to clinical data using published criteria (Rabinovici et al., 2011). Visual inspection based on these criteria has been validated previously as reproducible and reliable estimate of increased tracer uptake when compared with quantitative analysis. Every patient was thus assigned an amyloid positive (PiB+) or amyloid negative (PiB−) status.

2.3. Cognitive and sociodemographic data

As part of their assessment, all patients underwent a multi-disciplinary evaluation including a neurological examination, neuropsychological assessment, caregiver interview, screening laboratory tests, and high-resolution structural T1-weighted magnetic resonance imaging (MRI). Neuropsychological assessment included a previously described battery of standardized tests assessing memory, visuospatial abilities, and executive functions, global cognition and functional status (Gorno-Tempini et al., 2004).

2.4. Connected speech sample

A total of 117 CS transcripts were included in the study, including 47 PiB+ and 70 PiB− PPA patients. Table 1 shows the characterization of the sample by amyloid status (PiB).

CS samples were obtained from a total of 113 PPA patients. We included two transcripts for four patients who produced picture descriptions on two separate visits. The present study aims at predicting the amyloid status (PiB). We purposefully bypass clinical diagnosis in our classification experiment. However, for sample characterization, we report that the PPA group included 20 svPPA with predominantly left anterior lobe atrophy (svPPA-l; n = 20), 12 svPPA with predominantly right anterior lobe atrophy (svPPA-r; n = 12), 38 nfvPPA, 43 lvPPA.

CS was elicited through the "picnic scene" picture description task from the Western Aphasia Battery (Kertesz, 1982). Patients were instructed the following: "take a look at this picture, tell me what you see, and try to talk in sentences". The audio was imported into Audacity (http://audacity.sourceforge.net) and manually transcribed in English orthography. Unintelligible words were removed in preprocessing. Previous reports have shown highly reliable transcriptions using similar guidelines (Wilson et al., 2010).

2.5. CS feature extraction

We contribute to the growing trend in CS research (Clarke et al., 2020) with a hybrid approach similar to Fraser et al. (2014); Garrard, Rentoumi, Gesierich, Miller, and Gorno-Tempini (2014). We first extract a large array of features focusing exclusively on variables that can be computed automatically with natural language processing tools and add features from experimental (Ash et al., 2013; Fraser et al., 2014; Wilson et al., 2010), review (Auclair-Ouellet, 2015; Boschi et al., 2017; Slegers et al., 2018) and meta-analytic (Kamath et al., 2020) studies comparing groups of PPA subtypes to one another and/or amnestic AD to controls. We then perform a priori feature selection. This framework is common in studies deploying AI on language samples of AD patients and ensures that the model includes

features describing clinically relevant CS characteristics (cf. de la Fuente Garcia, Ritchie, and Luz (2020) for a systematic review).

Text pre-processing was performed using spaCy (https://spacy.io/) version 2.3.2 and the "en_core_web_lg" model with standard pre-processing steps of tokenization, Named Entity Recognition (NER), Part of Speech tagging (POS) and dependency parsing (DEP). Variables computed with spaCy covered many lexical and syntactic features typical for the analysis of CS in neurodegenerative diseases. We supplemented psycholinguistic, semantic, and pragmatic features that were found to be crucial for the study of patients with possible amyloid pathology (or more generally PPA) and that are not already implemented in standard NLP libraries. In total, 261 features were computed: lexical = 53; syntactic = 157; semantic = 37; pragmatic = 14. Table A1 provides the list of extracted features with their definition and operationalisation. Analysis code is available (https://osf.io/uf3zb/).

2.5.1. Lexical features—Part-of-speech tags (POS) were extracted. We examined accuracy of the tags within a predefined (random seed set at 42) random 10% subset of transcripts. Upon close inspection of the spacy POS-tags, we identified two anomalies that required correction. First, 70% of gerunds (i.e., fishing, sailing, drinking, swimming, flying, reading) were wrongfully tagged as "nouns" and were manually changed to "verbs". Second, 2.4% of the hesitation markers "uhm" were wrongfully tagged and was manually assigned an "interjection" tag.

Several different measures of lexical diversity (i.e., the ratio of different words relative to the total number of words used) have been reported in the study of CS in neurodegenerative diseases. Here, we extracted three popular metrics, namely Type-token ratio (TTR), Honoré (R) and Brunet (W) as in the study of Bucks, Singh, Cuerden, and Wilcock (2000). Brunet's W is an inverted scale: a higher Brunet's W indicates lower lexical diversity.

The absolute number of inflected verbs and their proportion relative to sample length were computed. Multiple psycholinguistic databases were queried for ratings of imageability and familiarity (Scott, Keitel, Becirspahic, Yao, & Sereno, 2019), concreteness (Brysbaert, Warriner, & Kuperman, 2014), age of acquisition (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012), frequency (Brysbaert & New, 2009), valence, arousal, and dominance (Warriner, Kuperman, & Brysbaert, 2013). These psycholinguistic properties concern the conceptual and emotional content of individual words as they were subjectively rated by a group of participants in the published norms. Averages of these psycholinguistic variables were computed for 1) all words and for 2) nouns, 3) verbs, 4) adverbs, and 5) adjectives separately.

2.5.2. Syntactic features—Universal dependency[1] relationships (DEP) were extracted and proportions were established relative to both the number of words and the number of sentence roots in the transcript. For example, the word *family* in the utterance "*the family is having a picnic*" is assigned by spaCy the part-of-speech tag "NOUN" and the universal dependency "*nsubj*" (nominal subject), reflecting its syntactic function in the utterance.

---

[1] https://universaldependencies.org/.

The number of children, i.e., the direct dependents of a word, which are connected to it by a single arc to the left or to the right in the dependency tree, was averaged. Mean sentence length and "noun chunks" were extracted using spaCy's sentence and noun_chunk boundaries, respectively.

2.5.3.   Semantic features—Information content units (ICU) were computed based on the list proposed by Jensen, Chenery, and Copland (2006). The ICU list includes 36 elements, which denote subjects, objects, places, and actions in the picture. ICUs are extracted as 36 separate Boolean features i.e., as they were mentioned (True) or not (False) in the patient's description. The total number of ICUs mentioned was also tallied.

2.5.4.   Pragmatic features—Features for pragmatic elements of speech were inspired by the findings of Garrard et al. (2014); Rentoumi, Raoufian, Ahmed, de Jager, and Garrard (2014) and were counted as follows, where words in brackets denote one instance of the pragmatic event: filled pauses ("uhm"), word-finding difficulties ("know", "remember", "unable"), deictic words ("this", "that", "here"), and uncertainty ("think", "look", "like", "kind", "see", "maybe", "can", "something"). We acknowledge that these groupings and the labels "word-finding difficulty" and "uncertainty" are simplifications of diverse clinical phenomena that we postulate to be reflected in transcriptions of CS. These simplifications - or groupings - were a compromise to include relevant words highlighted in previous literature while not inflating the number of individual features by inputting single words as features. We implemented a measure of semantic idea density by computing the average cosine distance between all pairwise combinations of words within a sliding window (Ivensky, 2019). The word embeddings to average cosine distances were extracted from the same spacy "*en_core_web_lg*" model that supported grammatical dependency and POS tagging. Sliding windows of 3, 10, 20, and 50 words with a step size of half the window length were implemented, outputting four idea density scores per participant.

2.6.   Classification and significant group differences

2.6.1.   Classification with leave-one-out cross-validation—In total, 261 features were extracted (Table A1, appendix). To reduce dimensionality of the classification model, feature selection is performed within each of the 117 folds of the cross-validation scheme. For our first objective of classification, feature selection operates as follows: 1) perform Welch's t-tests across groups; 2) correct the associated *p values* for multiple comparisons by the Benjamini-Hochberg method for False Discovery Rate (FDR) at .05; 3) subset only significant features at $p < .05$ after correction for FDR; 4) starting with the largest Welch's coefficient, enter the features in the model in a stepwise manner at the condition that the added feature is not correlated >.75 to a previously entered feature with a higher *t* statistic. The classifier model is built with the selected features and then used to predict the unseen, "left-out" sample. This procedure is repeated within each fold of the leave-one-out cross-validation. As no dedicated test set is established for final model evaluation, we report performance metrics (accuracy, sensitivity, specificity) as averages of the leave-one-out cross-validation on the whole dataset (117 folds).

Data analysis was performed using Python with scikit-learn (version .21.2) implementations of two classifiers: Logistic Regression with "liblinear" solver and Support Vector Machine (SVM) with linear kernel. The classifiers and kernels were chosen for three reasons. First, using two classifiers demonstrates robustness and stability of classification accuracy (Fraser, Lundholm Fors, Eckerström, Öhman, & Kokkinakis, 2019b), i.e., ensuring that results are not specific to a single classifier. Second, linear kernels yield readily interpretable results (Fraser, Linz, Li, Fors, Rudzicz, & Königet, 2019a). Third, linear kernels tend to perform well when the number of features is relatively high compared to the number of data points (Hsu, Chang, & Lin, 2003), which is arguably the case in this experiment.

2.6.2. Significant and discriminant group differences—For our second objective of discussing group-level differences, we considered the whole sample (117 transcripts) instead of selecting features within each fold (116 transcripts in each of the 117 folds) as in Section 2.6.1. We identify discriminant group differences with a similar procedure: a series of Welch's tests corrected for multiple comparisons with the Benjamini-Hochberg method for False Discovery Rate (FDR) at .05 and stepwise removal of multicollinearity, i.e., removal of features when correlated >.75 with any feature with a higher t statistic.

3. Results

3.1. Classification of amyloid burden (PiB) from CS features

We sought to benchmark our approach against other biomarkers (cf. Discussion), which show relatively balanced sensitivity and specificity figures (sensitivity: 78%; specificity: 80% in Giannini et al., 2017; sensitivity: 78%; specificity: 77% in Ashton et al., 2019). Initially, our unweighted (100%) classification scheme yielded an unbalanced ratio (sens: 70%; spec: 77%). We thus a posteriori weighted the PiB + class in two increments of 50% (150%, 200%) to skew the sensitivity to specificity ratio toward benchmark figures. We report the results from the 150% weight for the PiB + because it yielded satisfying results (acc: 77%; sens: 74%; spec: 79%) and was stable on both classifiers. SVM returned an average classification accuracy of 77% with a sensitivity of 74% and specificity 79%. As Table 2 shows, Logistic Regression returned similar, but slightly weaker results.

3.2. Group differences: discriminant CS features for PiB + PPA patients

Fig. 1 shows distributions of the discriminant features that differentiate between PiB+ and PiB− groups on the whole dataset. In line with our objectives and methods, we only report and discuss discriminant features that survived the filter for multicollinearity.[2] The twelve discriminant features form an array that spans multiple linguistic domains: eight lexical features, one syntactic feature, one semantic feature, and two pragmatic features. In the lexical domain, four features were lower in the PiB + group: average imageability of all words, average imageability of nouns, number of nouns per sentence root, and average concreteness of verbs. Conversely, adjective dominance and frequency of verbs and adjectives, as well as Brunet's W index were higher in the PiB + group. In the

---

[2]Supplementary Table A4 contains the list of all significant features after correction for multiple comparison, but before filtering-out of multicollinearity.

syntactic domain, PiB + patients had a lower ratio of the average number of children to the total number of sentence roots. In the semantic domain, PiB + patients were less likely to precisely name the drink being poured by the woman in the picture. In the pragmatic domain, PiB + patients produced more words denoting uncertainty (e.g., "know", "remember", "looks like") relative to their total output in number of words (proportion of uncertainties). The PiB + group also exhibited lower semantic idea density. To ensure that the difference in years of education between the groups does not skew results, post-hoc regression analyses were carried out and showed that all variables retained a significant association to PiB status beyond the effect of education. Fig. 1 and Table 3 summarize these findings.

4. Discussion

The present study confirms our hypotheses by providing two main results. First, we showed that a short, non-invasive, and inexpensive sample of connected speech (CS) can identify PiB + patients with 77% accuracy, sensitivity of 74% and specificity of 79%, confirming the potential of CS in detecting the presence of amyloid in the brain of patients with PPA. Although methodological and sampling differences preclude a direct comparison of results with previous research, our analysis compares favorably to other approaches to identification of underlying pathophysiology such as clinical diagnosis and cognitive profiling (sensitivity: 78%; specificity: 80%) (Giannini et al., 2017) and plasma biomarkers (sensitivity: 78%; specificity: 77%) (Ashton et al., 2019).

Second, we demonstrated that a relatively small array of discriminant lexical, semantic, and pragmatic CS features, mostly attributable to word retrieval difficulties, play a central role in determining the CS profile characteristic of PIB + patients. The set of discriminant features indicates that PiB + patients present lower idea density, struggle to produce concrete nouns (imageability, frequency and concreteness), and that the nouns they manage to access are more abstract and are interspersed among many other words (nouns proportion), which tend to be less varied on average (Brunet W index). In addition, PIB + PPA patients present more uncertainties (e.g., think, maybe, looks like, etc.), which likely reflect CS attempts at circumventing word retrieval problems. Thus, this aspect of the laborious speech production of PiB + PPA patients can be captured by analyzing lexical features of CS transcripts as in Garrard et al. (2014); Rentoumi et al. (2014). The groups did not differ in the number of word-finding pauses, filled pauses, or other disruptions to fluency typically associated with word retrieval difficulties (Ash et al., 2013; Wilson et al., 2010). This latter set of features may not be sensitive markers since all PPA patients experience - albeit for different pathophysiological and neuropsychological reasons – a significant degree of pragmatic difficulties and disruptions to fluency. Beyond word retrieval difficulties, these features could reflect semantic processing struggles, as in svPPA, or hesitation in grammatical construction, as in nfPPA. The pragmatic features included in the analysis were crafted from observations garnered in related CS studies of svPPA (Garrard et al., 2014) and multifactorial AD (Rentoumi et al., 2014). These features are more subjective than the remainder of the feature set.

Future studies should include more automatic and fewer task-dependent features describing the characteristics (filled and unfilled) and durations of the pauses. In addition, the inclusion of features related to distortions and phonological paraphasia, which are known to occur in lvPPA, should also be considered to better capture word retrieval difficulties in PIB + PPA (Wilson et al., 2010). The inclusion of more sensitive variables could help improve classification accuracy in future studies, considering that the present experiment achieves 77%, from 60% chance accuracy.

So far, the set of discussed features can be traced to word retrieval difficulties previously described in PIB + PPA patients but can also be observable in patients with semantic deterioration (PiB− svPPA patients). Interestingly, the analysis also revealed the presence of a discriminant syntactic feature that is not shared by svPPA patients, namely a higher ratio of the number of average left "children" to the total number of sentence roots. Short sentences as a marker of PiB + speech could appear surprising considering they do not usually present with agrammatism. However, consistent with the absence of agrammatism, PIB + patients do not produce fewer verbs. In this framework, the lower ratio of the average number of children to the total number of sentence roots suggests that PiB + patients start many sentences with root verbs but that these have few dependents. This simplification of syntax in the absence of agrammatism is likely to reflect verbal working memory deficits that are well-documented in PIB + PPA but also described in AD patients. The pattern of CS features is generally compatible with the neuroanatomical correlates of lexical retrieval and verbal working memory deficits, namely atrophy in the posterior superior and middle temporal gyri and inferior parietal lobule, and damage to the posterior inferior longitudinal fasciculus, anterior inferior longitudinal fasciculus, uncinate fasciculus, and superior longitudinal fasciculus (Montembeault et al., 2018).

In addition, PIB + PPA patients also displayed higher average dominance of adjectives that is driven by the presence of emotional words such as "happy", "good", "strong", "active", and "nice". PiB + patients tended to include more vivid adjectives related to characters' feelings and general atmosphere of the picnic scene. This finding is compatible with the better processing of emotional concepts by left anterior insular regions in lvPPA compared to nfPPA patients (Piguet, Leyton, Gleeson, Hoon, & Hodges, 2015). Finally, PiB + patients were significantly less likely to identify the drink being poured by the woman in the picture. Indeed, 60% of the PiB− group specified what the drink could be, whereas only 26% of the PiB + group did. Patients of the latter group often substituted the information with uncertainty, e.g., "*not sure what she is drinking*" (not specifying the drink) instead of "*she may be pouring coffee or tea*" (specifying the drink). A drawback of this marker is its specificity to the picnic picture description task, and in this context, we cannot exclude that its significance is artefactual.

5. Conclusion

For patients presenting with language symptoms suggestive of PPA, a full clinical examination is indicated before scanning per current recommendations on clinical amyloid imaging (Laforce et al., 2016; Witte et al., 2015). When extensive testing is not feasible

or delayed, an advantageous approach could be the collection of a short sample of CS. We showed that CS analysis has good sensitivity and specificity to AD pathology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Declaration of competing interest

The funders were not involved in the study design, collection, analysis or interpretation of data, nor were they involved in writing the paper or the decision to submit this article for publication.

To our knowledge, no part of the study analyses was pre-registered in a time-stamped, institutional registry prior to the research being conducted. Here, we report how we determined our sample size (maximum data available), all data exclusions (N/A), all inclusion/exclusion criteria (cf. Methods section), whether inclusion/exclusion criteria were established prior to data analysis (yes), all manipulations, and all measures in the study. Public archiving of anonymized data is not contemplated by the study's IRB approval. Specific requests can be submitted through the UCSF MAC Resource Request form: http://memory.ucsf.edu/resources/data. Following a UCSF-regulated procedure, access will be granted to designated individuals in line with ethical guidelines on the reuse of sensitive data. This would require submission of a Material Transfer Agreement, available at: https://icd.ucsf.edu/material-transfer-and-data-agreements. Commercial use will not be approved. Dr. Maria Luisa Gorno-Tempini is responsible for granting access to the raw data.
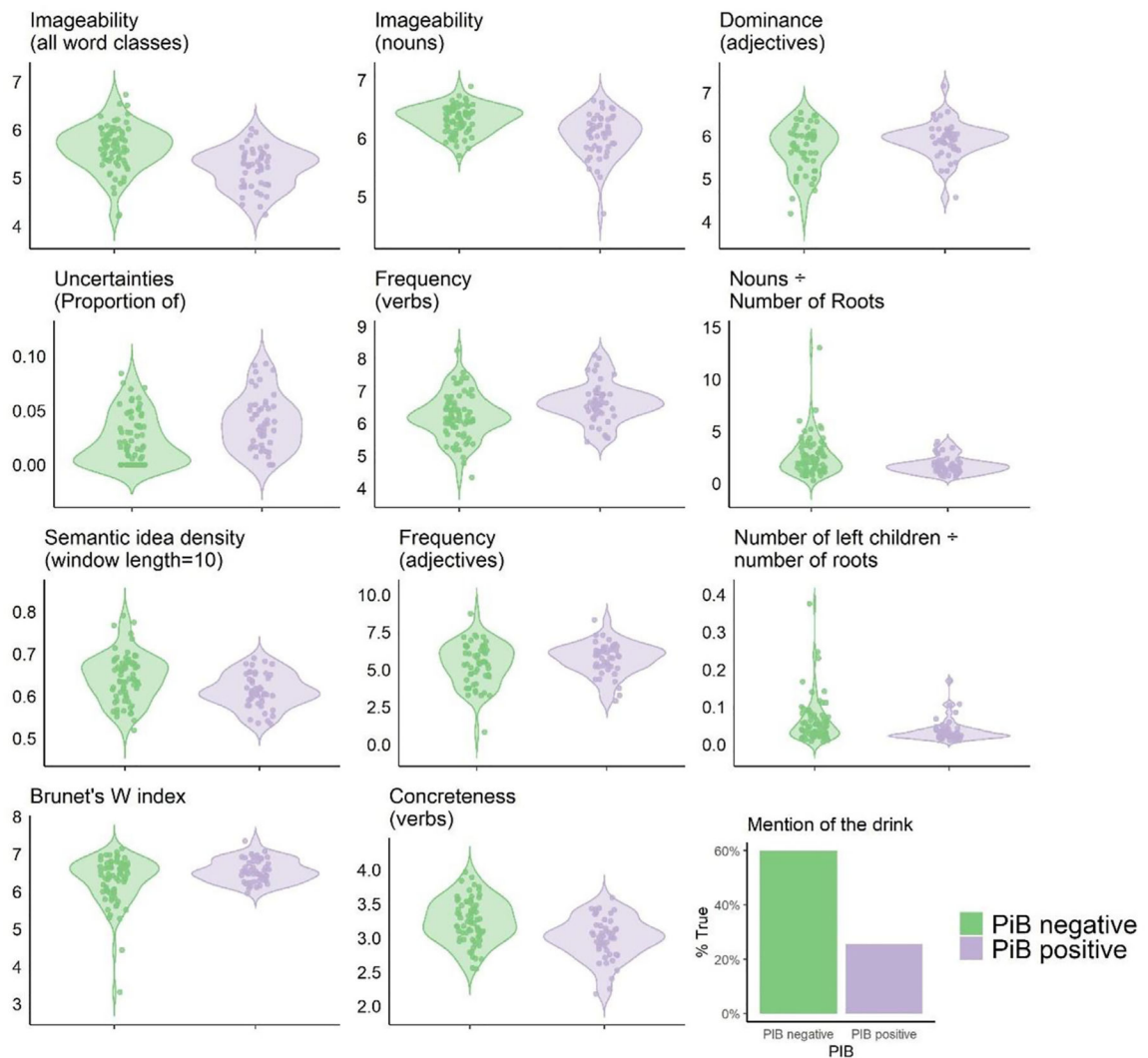
REFERENCES

Ash S, Evans E, O'shea J, Powers J, Boller A, Weinberg D, et al. (2013). Differentiating primary progressive aphasias in a brief sample of connected speech. Neurology, 81(4), 329–336. [PubMed: 23794681]

Ashton NJ, Nevado-Holgado AJ, Barber IS, Lynham S, Gupta V, Chatterjee P, et al. (2019). A plasma protein classifier for predicting amyloid burden for preclinical Alzheimer's disease. Science Advances, 5(2), Article eaau7220.

Auclair-Ouellet N (2015). Inflectional morphology in primary progressive aphasia and Alzheimer's disease: A systematic review. Journal of Neurolinguistics, 34, 41–64.

Battista P, Salvatore C, Berlingeri M, Cerasa A, & Castiglioni I (2020). Artificial intelligence and neuropsychological measures: The case of Alzheimer's disease. Neuroscience & Biobehavioral Reviews, 114, 211–228. 10.1016/j.neubiorev.2020.04.026 [PubMed: 32437744]

Beltrami D, Gagliardi G, Rossini Favretti R, Ghidoni E, Tamburini F, & Calzà L (2018). Speech analysis by natural language processing techniques: A possible tool for very early detection of cognitive decline? Frontiers in Aging Neuroscience, 10, 369. [PubMed: 30483116]

Bergeron D, Gorno-Tempini ML, Rabinovici GD, Santos-Santos MA, Seeley W, Miller BL, et al. (2018). Prevalence of amyloid-β pathology in distinct variants of primary progressive aphasia. Annals of Neurology, 84(5), 729–740. [PubMed: 30255971]

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

Brysbaert M, & New B (2009). Moving beyond Ku era and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. [journal article]. Behavior Research Methods, 41(4), 977–990. 10.3758/brm.41.4.977 [PubMed: 19897807]

Brysbaert M, Warriner AB, & Kuperman V (2014). Concreteness ratings for 40 thousand generally known English word lemmas. Behavior Research Methods, 46(3), 904–911. [PubMed: 24142837]

Bucks RS, Singh S, Cuerden JM, & Wilcock GK (2000). Analysis of spontaneous, conversational speech in dementia of Alzheimer type: Evaluation of an objective technique for analysing lexical performance. Aphasiology, 14(1), 71–91.

Clarke N, Foltz P, & Garrard P (2020 Aug). How to do things with (thousands of) words: Computational approaches to discourse analysis in Alzheimer's disease. Cortex, 129, 446–463. 10.1016/j.cortex.2020.05.001. Epub 2020 May 19. [PubMed: 32622173]

de la Fuente Garcia S, Ritchie CW, & Luz S (2020). Artificial intelligence, speech, and language processing approaches to monitoring Alzheimer's disease: A systematic review. Journal of Alzheimer's Disease, 78(4), 1547–1574. 10.3233/JAD-200888.

Dubois B, Feldman HH, Jacova C, Cummings JL, DeKosky ST, Barberger-Gateau P, et al. (2010). Revising the definition of Alzheimer's disease: A new lexicon. The Lancet Neurology, 9(11), 1118–1127. [PubMed: 20934914]

Fink HA, Hemmy LS, Linskens EJ, Silverman PC, MacDonald R, McCarten JR, et al. (2020). Diagnosis and treatment of clinical Alzheimer's-type dementia: A systematic review.

Fraser KC, Meltzer JA, Graham NL, Leonard C, Hirst G, Black SE, et al. (2014). Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex, 55, 43–60. [PubMed: 23332818]

Fraser KC, Linz N, Li B, Fors KL, Rudzicz F, König A, et al. (2019a). Multilingual prediction of Alzheimer's disease through domain adaptation and concept-based language modelling. In Proceedings of the Conference NAACL HLT 2019: 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Vol. 1, pp. 3659–3670). Association for Computational Linguistics (Long and Short Papers).

Fraser KC, Lundholm Fors K, Eckerström M, Öhman F, & Kokkinakis D (2019b). Predicting MCI status from multimodal language data using cascaded classifiers. Frontiers in Aging Neuroscience, 11, 205. 10.3389/fnagi.2019.00205. [PubMed: 31427959]

Garrard P, Rentoumi V, Gesierich B, Miller B, & Gorno-Tempini ML (2014). Machine learning approaches to diagnosis and laterality effects in semantic dementia discourse. Cortex, 55, 122–129. [PubMed: 23876449]

Giannini LA, Irwin DJ, McMillan CT, Ash S, Rascovsky K, Wolk DA, et al. (2017). Clinical marker for Alzheimer disease pathology in logopenic primary progressive aphasia. Neurology, 88(24), 2276–2284. [PubMed: 28515265]

Gorno-Tempini ML, Dronkers NF, Rankin KP, Ogar JM, Phengrasamy L, Rosen HJ, et al. (2004). Cognition and anatomy in three variants of primary progressive aphasia. Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, 55(3), 335–346.

Gorno-Tempini M, Hillis A, Weintraub S, Kertesz A, Mendez M, Cappa SE, et al. (2011). Classification of primary progressive aphasia and its variants. Neurology, 76(11), 1006–1014. [PubMed: 21325651]

Hsu C-W, Chang C-C, & Lin C-J (2003). A practical guide to support vector classification. Taipei.

Ivensky I (2019). Automatic detection of semantic dementia from scene description. Montréal: M.Sc., Université de Montréal

Jensen AM, Chenery HJ, & Copland DA (2006). A comparison of picture description abilities in individuals with vascular subcortical lesions and Huntington's disease. Journal of Communication Disorders, 39(1), 62–77. [PubMed: 16154582]

Johnson KA, Minoshima S, Bohnen NI, Donohoe KJ, Foster NL, Herscovitch P, et al. (2013). Appropriate use criteria for amyloid PET: A report of the amyloid imaging task force, the society of nuclear medicine and molecular imaging, and the Alzheimer's association. Journal of Nuclear Medicine, 54(3), 476–490. [PubMed: 23359661]

Kamath V, Sutherland ER, & Chaney G-A (2020). A meta-analysis of neuropsychological functioning in the logopenic variant of primary progressive aphasia: Comparison with the semantic and non-fluent variants. Journal of the International Neuropsychological Society: JINS, 26(3), 322–330. [PubMed: 31658919]

Kertesz A (1982). Western aphasia battery test manual. Psychological Corp.

Kuperman V, Stadthagen-Gonzalez H, & Brysbaert M (2012). Age-of-acquisition ratings for 30,000 English words. Behavior Research Methods, 44(4), 978–990. [PubMed: 22581493]

Laforce R, Rosa-Neto P, Soucy J-P, Rabinovici GD, Dubois B, & Gauthier S (2016). Canadian consensus guidelines on use of amyloid imaging in Canada: Update and future directions from the specialized task force on amyloid imaging in Canada. The Canadian Journal of Neurological Sciences, 43(4), 503–512. [PubMed: 26916179]

Laske C, Sohrabi HR, Frost SM, López-de-Ipiña K, Garrard P, Buscema M, et al. (2015). Innovative diagnostic tools for early detection of Alzheimer's disease. Alzheimer's & Dementia, 11(5), 561–578.

McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia, 7(3), 263–269.

Montembeault M, Brambati SM, Gorno-Tempini ML, & Migliaccio R (2018). Clinical, anatomical, and pathological features in the three variants of primary progressive aphasia: A review. Frontiers in Neurology, 9, 692. [PubMed: 30186225]

Myszczynska MA, Ojamies PN, Lacoste AMB, Neil D, Saffari A, Mead R, … Ferraiuolo L (2020). Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. Nature Reviews Neurology. 10.1038/s41582-020-0377-8

Rabinovici GD, Jagust WJ, Furst AJ, Ogar JM, Racine CA, Mormino EC, et al. (2008). Aβ amyloid and glucose metabolism in three variants of primary progressive aphasia. Annals of Neurology, 64(4), 388–401. [PubMed: 18991338]

Piguet O, Leyton CE, Gleeson LD, Hoon C, & Hodges JR (2015). Memory and emotion processing performance contributes to the diagnosis of non-semantic primary progressive aphasia syndromes. Journal of Alzheimer's Disease, 44(2), 541–547. 10.3233/JAD-141854.

Rabinovici G, Rosen H, Alkalay A, Kornak J, Furst A, Agarwal N, et al. (2011). Amyloid vs FDG-PET in the differential diagnosis of AD and FTLD. Neurology, 77(23), 2034–2042. [PubMed: 22131541]

Rentoumi V, Raoufian L, Ahmed S, de Jager CA, & Garrard P (2014). Features and machine learning classification of connected speech samples from patients with autopsy proven Alzheimer's disease with and without additional vascular pathology. Journal of Alzheimer's Disease, 42(s3), S3–S17.

Rogalski E, Sridhar J, Rader B, Martersteck A, Chen K, Cobia D, et al. (2016). Aphasic variant of Alzheimer disease: Clinical, anatomic, and genetic features. Neurology, 87(13), 1337–1343. [PubMed: 27566743]

Santos-Santos MA, Rabinovici GD, Iaccarino L, Ayakta N, Tammewar G, Lobach I, et al. (2018). Rates of amyloid imaging positivity in patients with primary progressive aphasia. JAMA Neurology, 75(3), 342–352. [PubMed: 29309493]

Scott GG, Keitel A, Becirspahic M, Yao B, & Sereno SC (2019). The Glasgow Norms: Ratings of 5,500 words on nine scales. Behavior Research Methods, 51(3), 1258–1270. [PubMed: 30206797]

Slegers A, Filiou R, Montembeault M, & Brambati S (2018). Connected speech features from picture description in Alzheimer's disease: A systematic review. Journal of Alzheimer's Disease: JAD, 65(2), 519. [PubMed: 30103314]

Tan RH, Kril JJ, Yang Y, Tom N, Hodges JR, Villemagne VL, et al. (2017). Assessment of amyloid β in pathologically confirmed frontotemporal dementia syndromes. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 9, 10–20.

Villarejo-Galende A, Llamas-Velasco S, Gómez-Grande A, Puertas-Martin V, Contador I, Sarandeses P, et al. (2017). Amyloid pet in primary progressive aphasia: Case series and systematic review of the literature. Journal of Neurology, 264(1), 121–130. [PubMed: 27815682]

Villemagne VL, Doré V, Burnham SC, Masters CL, & Rowe CC (2018). Imaging tau and amyloid-β proteinopathies in Alzheimer disease and other conditions. Nature Reviews Neurology, 14(4), 225–236. [PubMed: 29449700]

Villeneuve S, Rabinovici GD, Cohn-Sheehy BI, Madison C, Ayakta N, Ghosh PM, et al. (2015). Existing Pittsburgh Compound-B positron emission tomography thresholds are too high: Statistical and pathological evaluation. Brain, 138(7), 2020–2033. [PubMed: 25953778]

Warriner AB, Kuperman V, & Brysbaert M (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. Behavior Research Methods, 45(4), 1191–1207. [PubMed: 23404613]

Wilson SM, Henry ML, Besbris M, Ogar JM, Dronkers NF, Jarrold W, et al. (2010). Connected speech production in three variants of primary progressive aphasia. Brain, 133(7), 2069–2088. [PubMed: 20542982]

Witte MM, Foster NL, Fleisher AS, Williams MM, Quaid K, Wasserman M, et al. (2015). Clinical use of amyloid-positron emission tomography neuroimaging: Practical and bioethical considerations. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 1(3), 358–367.

Fig. 1 –.

Significant group differences for discriminant features. All group differences are significant at $p < .05$ after Benjamini-Hochberg correction for False discovery rate of .05. Most Y-axis labels were omitted to lighten the plot. Imageability, Dominance, Frequency, Concreteness are based on Likert scales from the psycholinguistic databases listed in Section 2.5.1. Proportion of Uncertainties is the ratio of number of uncertainties in the transcript divided by the total number of words. Nouns/number of roots is the ratio of number of nouns divided by the number of sentence roots in the transcript. Number of left children/number of roots is the ratio of number of left children divided by number of sentence roots in the transcripts. Brunet's W index results from the equation in Table A1 and yields an inverted scale: a higher Brunet's W index corresponds to lower lexical diversity. Mention of the drink is plotted as the percentage of transcripts in which the drink was mentioned.

Table 1 –

Average sociodemographic information and severity of impairment by group. Welch's t-test with significance level, uncorrected. MMSE: Mini-Mental State Evaluation; CDR: Clinical Dementia Rating scale.

Sociodemographic information and severity	PiB negative N = 70	PiB positive N = 47	Welch's t-test	*p* value
Age	65.83	63.71	−1.38	.17
Handedness (% right-handed)	77	77	.07	.97
Gender (% female)	53	60	.69	.49
Education (years)	15.72	16.91	2.05	.04
MMSE	24.03	22.28	−1.38	.17
CDR	.59	0.6	.14	.89
CDR Box Score	2.98	3.33	.73	.46

Table 2 –

Averaged classification outcomes for leave-oneout cross-validation.

Classifier	Accuracy	Sensitivity	Specificity
Logistic Regression	.74	.72	.74
SVM	.77	.74	.79

Table 3 –

Significant group differences for discriminant features. Corrected *p* values with Benjamini-Hochberg correction and False discovery rate of .05.

Connected speech feature	PiB– mean	PiB+ mean	Welch's t-test	*p* value	corr. *p* value
Imageability (all word classes)	5.62	5.19	−5.05	.000002	.0004
Imageability (nouns)	6.32	6.04	−4.59	.00002	.001
ICU: drink	.60	.26	−3.95	.0001	.005
Dominance (adjectives)	3.91	5.40	3.66	.0004	.01
Uncertainties (Proportion of)	.02	.04	3.52	.0007	.01
Frequency (verbs)	6.13	6.65	3.43	.0009	.01
Nouns/Number of roots	2.61	1.73	−3.43	.0009	.01
Semantic idea density (window length = 10)	.64	.61	−3.41	.0009	.01
Frequency (adjectives)	4.13	5.34	3.03	.003	.03
Number of left children/number of roots	.06	.04	−3.02	.003	.03
Brunet's W index	6.30	6.55	2.91	.004	.04
Concreteness (verbs)	3.20	2.99	−2.81	.006	.045