Automated profiling of spontaneous language production in primary progressive aphasia

Zimmerer, V.C., Hardy, C. J. D., Eastman, J., Dutta, S., Varnet, L., Bond, R.L., Russell, L., Rohrer, J.D., Warren, J.D., Varley, R.A.

Abstract

Objective: Spontaneous language produced by aphasic individuals is challenging to quantify. We used a computerized tool to profile and compare spontaneous language output in canonical primary progressive aphasia (PPA) syndromes in relation to behavioral variant frontotemporal dementia (bvFTD) and healthy controls.

Methods: We examined "last holiday" narrative speech samples from 29 speakers with semantic variant PPA (svPPA), 25 speakers with logopenic variant PPA (lvPPA), 34 speakers with non-fluent variant PPA (nfvPPA), 14 speakers with bvFTD and 20 older normal controls (NCs). We applied the Frequency in Language Analysis Tool which uses frequency and collocation strength measures to determine how common and familiar word and word combinations were in the speech samples. It also computes word counts, content word ratios and a measure of connected language.

Results: Taking age into account, all dementia subtypes differed from NCs. Speakers with svPPA produced more familiar content words. Speakers with lvPPA produced less connected language, fewer and more familiar content words and more familiar word combinations. Speakers with nfvPPA produced fewer words over all, less connected language, fewer and more familiar content words and more familiar combinations. Speakers with bvFTD produced fewer words and less connected language. Variables also distinguished all dementia groups from another. A machine learning classifier using these variables achieved 59.4% accuracy in matching samples to their diagnosis, and a 90% accuracy when classifying samples under NC or dementia.

Conclusions: Automated quantification of natural language in progressive aphasia, but also non-language led dementia, is feasible and extracts syndromic profiles that complement those derived from standard neurolinguistic tests, warranting further evaluation as a candidate biomarker.

1. Introduction

Diagnosis of the 'language-led dementias' or primary progressive aphasias (PPA) commonly relies on a range of formal neuropsychological tests.[1–3] In addition to test scores, features of spontaneous speech are determined through qualitative observation, but with little quantification beyond error counting.

However, there is considerable clinical interest in quantification of broader aspects of spontaneous speech production: this might not only give insight into the functional difficulties experienced by patients and their communication partners, but support identification of early decline in language function and allow highly sensitive tracking of behavior change.[4–8]

Here we analyzed language samples from the three major PPA variant syndromes: semantic variant (svPPA), logopenic variant (lvPPA) and non-fluent variant (nfvPPA) using the Frequency in Language Analysis Tool (FLAT), a novel automated script for quantification of language. FLAT captureslanguage usage frequency not only of single words, but also word combinations. More frequent forms are easier to process, and speakers with svPPA and nfvPPA use more frequent words in their output.[4]. More common combinations are processed differently,[9] requiring less combinatorial effort than less frequent or novel formulations. Previous work with our methods showed that speakers with Alzheimer's disease[5] and stroke aphasia[10,11] use more common combinations than NCs in spontaneous language production. FLAT also determines content word ratio and provides a measure of degree of connectedness of language. We compared automated language measures in PPA syndromes with measures derived from patients with behavioral variant frontotemporal dementia (bvFTD) as a disease control group with a distinct neurolinguistic phenotype,[12] and with older normal control individuals (NC).

2. Methods

2.1 Hypotheses

We hypothesized that in pairwise group comparisons each group would differ from each other on at least one variable. While we had expectations about specific variables and directions, such as that svPPA would produce more common words than NCs and that speakers with nfvPPA would produce less connected language, we chose bi-directional hypotheses given the novelty of the research.

2.2 Participants and samples

All participants were recruited and tested in London. The study was approved by the University College London institutional ethics committee (reference no. Q6/Q051/52) and all participants gave informed consent in accordance with the Declaration of Helsinki. Speech samples and MMSE scores were recorded on first assessment. Time points varied relative to initial diagnosis.

We selected recorded spontaneous speech samples from the UCL PPA database, based on the following criteria: (1) diagnosis of one of the four canonical dementia types, or healthy older control sample, and

(2) sample of adequate audio quality for transcription. Each sample was based on the participant's narrative description of their last holiday. Speech samples were orthographically transcribed from the audio recording. In cases of portions of unclear audio or speech, an experienced clinician (R.A.V.) was consulted and consensus was reached. In cases of phonological errors, the target word was transcribed if it was recognizable on the basis of context and phonological form. If not, it was excluded. Place and person names were also excluded.

The final sample set consisted of 20 NCs (mean age = 62.8, SD = 7.3; 10 female), 29 participants with svPPA (mean age = 64.0, SD = 7.8; 12 female), 25 participants with lvPPA (mean age = 63.3, SD = 13.6; 13 female), 34 participants with nfvPPA (mean age = 69.9, SD = 8.4; 20 female), and 14 people with bvFTD (mean age = 64.4, SD = 64.4; 12 female). All patients met consensus criteria for the relevant diagnosis.[1,13] Age differed significantly across groups, $\chi^2(4) = 13.68$, $p = .008$ and was therefore included as a covariate in further analyses. The effect was driven by the nfvPPA group being significantly older than the NC and svPPA groups ($p = .004$ in both comparisons).

2.3 Sample analysis procedure

Samples were annotated for analysis using the Language Analysis Tool (FLAT).[5,10,11] The FLAT is a computer program which parses any text, e.g. transcribed speech, retrieves each word, bigram (two-word combination) and trigram (three-word combination) and looks up their usage frequency (reported in instances per million words) in the spoken subsection of the British National Corpus (BNC; 2007) a 10 million word collection of normative samples from a range of communication contexts, geographical regions and demographic groups. Based on these values, it also computes t-scores as a measure of collocation strength, i.e. how often words in a combination appear together as opposed to in other contexts. Collocation strength is not a direct function of usage frequency: A high-frequency sentence like *I don't know* has high collocation strength because the individual words occur together very often, but so does a low-frequency compound such as *plate tectonics*, since *tectonics* appears in almost no other context. Frequency and collocation strength are indicators of familiarity of words or word combinations, and these values increase in the output of people with stroke aphasia.[10,11] We present these values as averages for words and bigrams types (as opposed to tokens), meaning that each instance of a word or word combination is only entered once to compute the individual's average, ignoring repetitions of the respective unit.

The FLAT also characterizes words as content words which refer to things (e.g., *table*), actions (e.g., *climb*) and properties (e.g., *fast*), or function words which include pronouns (e.g., s*he*), articles (e.g., *the*)

and question particles (e.g. *when*). Measures of word frequency are restricted to content words. A content word ratio is computed, which characterizes the extent to which content words are replaced by function words such as pronouns. A "combination ratio" is also computed by dividing the number of trigrams by the number of words, with higher values reflecting more connected speech. In this way, the analysis captures both lexical change (e.g. shift to high-frequency vocabulary) and grammatical change (omission of function words, reliance on more familiar utterances). In line with previous investigations, our analysis excluded immediate word and phrase repetitions unless considered intentional e.g., very, very big), and also excluded ungrammatical bigrams. Previous research found that for word combinations, collocation strength for bigrams provided the most information.[5,10] We therefore included bigram collocation strength only.

To compare groups on each variable, we computed analyses of covariance with participant group as the independent variable, age as covariate and the language measure as dependent variable. We treated each of the variables word count, combination ratio, content word ratio, content word frequency and bigram t-score (collocation strength) as separate hypotheses and adjusted significance thresholds for multiple comparisons between groups. We examined the relationship between language measures and age, as well as MMSE scores within each group using Pearson's r. We further tested how FLAT variables could be used for the classification samples by training a machine learning classifier. For this analysis we used a support vector machine (SVM).

3. Results

3.1 Group comparisons

Table 1 displays group averages and main effects for between group comparisons with age entered as covariate. All independent variables yielded significant main effects in between-group comparisons. The biggest effect size was for combination ratio, followed by content word frequency, word count, content word ratio and collocation strength. Figure 1 visualizes the language profiles of the dementia subtypes in relation to NCs.

For pairwise comparisons we used Bonferroni adjustments of significance threshold based on ten comparisons for each variable. The adjusted threshold was $p < .005$. For mean differences, confidence intervals and p values, see Appendix A. Compared to NCs, speakers with svPPA produced more familiar content words. Speakers with lvPPA produced language that was less connected, contained fewer and more familiar content words, and more familiar word combinations. Speakers with nfvPPA produced fewer words, less connected language, more familiar content words and more familiar combinations.

Speakers with bvFTD produced fewer words and less connected output. However, each dementia subgroup differed from each other subgroup on at least one variable (Table 2; Appendix A).

Table 1. Group averages and main effects as calculated by ANCOVAs with age as covariate. Age had no significant effect.

Variable	**Normal controls**	**Semantic variant PPA**	**Logopenic variant PPA**	**Non-fluent variant PPA**	**Behavioural variant FTD**	**Main effect (Age as covariate)**
Word count	260.7 (SD = 169)	248.7 (SD = 141.3)	156.1 (SD = 98.0)	87.2 (SD = 80.7)	166.2 (SD = 134.6)	$F(4,116) = 9.813$, $p < .001$, $\eta^2 = .25$
Combination ratio	.77 (SD = .05)	.69 (SD = .07)	.60 (SD = .09)	.43 (SD = .18)	.60 (SD = .09)	$F(4,116) = 36.717$, $p < .001$, $\eta^2 = .56$
Content word ratio	.37 (SD = .04)	.33 (SD = .04)	.28 (SD = .05)	.35 (SD = .09)	.33 (SD = .07)	$F(4,116) = 5.878$, $p < .001$, $\eta^2 = .17$
Content word frequency (per million)	524 (SD = 185)	969 (SD = 337)	1278 (SD = 461)	966 (SD = 489)	820 (SD = 382)	$F(4,116) = 10.154$, $p < .001$, $\eta^2 = .26$
Bigram collocation strength (t-scores)	24.51 (SD = 4.07)	29.60 (SD = 4.00)	35.32 (SD = 11.90)	32.29 (SD = 12.27)	30.37 (SD = 10.92)	$F(4,116) = 3.56$, $p = .009$, $\eta^2 = .11$

Figure 1. Radar plot visualization of language profiles of different dementia groups. Data were residualized over participant age in order to account for age differences, and then normalized using control means and standard deviations. The outer line in each plot represents the control mean; each line towards the center represents a distance of one standard deviation from the control mean.

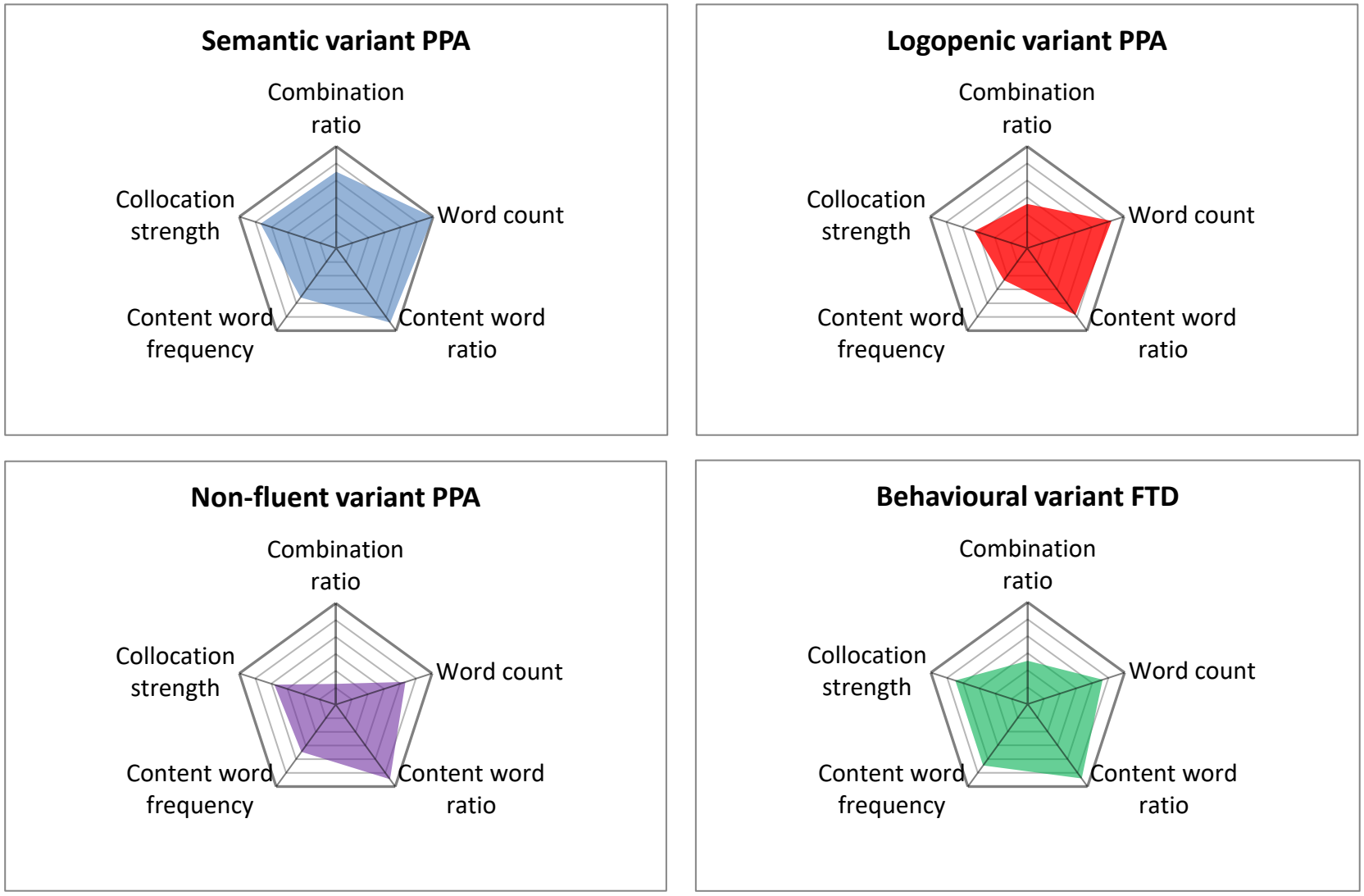


Table 2. List of features which distinguished spontaneous language production between groups. Features were only included if difference between groups was significant at $p < .005$ (see Appendix A for mean differences and inferential statistics). Cells describe difference of group in the left column relative to the group in the top row, e.g., top left cell: speakers with semantic variant PPA produce more familiar content words than neurotypical speakers. PPA = primary progressive aphasia; FTD = fronto-temporal dementia.

	Neurotypical speakers	**Semantic variant PPA**	**Logopenic variant PPA**	**Non-fluent variant PPA**	**Behavioral variant FTD**
Semantic variant PPA	Higher content word frequency	-	More combinations; Higher content word ratio	Higher word count; More combinations	Higher word count
Logopenic variant PPA	Fewer combinations; Lower content word ratio; Higher content word frequency; stronger collocations	Fewer combinations; Lower content word ratio	-	More combinations; Lower content word ratio	Higher content word frequency
Non-fluent variant PPA	Lower word count; Fewer combinations; Higher content word frequency; Stronger collocations	Lower word count; Fewer combinations	Fewer combinations; Higher content word frequency	-	Fewer combinations
Behavioral variant FTD	Lower word count; Fewer combinations	Lower word count	Lower content word ratio	More combinations	-

3.2 Relationship with age and MMSE

We next investigated the relationship between language variables as generated by the FLAT and age and general cognitive capacities as measured using the MMSE within each dementia group. MMSE scores were not obtained from NCs, and were not available from all participants with dementia. We had MMSE scores from 24 people with svPPA, 18 people with lvPPA, 23 people with nfvPPA and 12 people with bvFTD. We report all effects significant at $p < .05$. Age was not significantly correlated with language variables in any group. There were some significant correlations with MMSE scores. In people with svPPA, participants with lower MMSE scores produced more frequent content words, $r(24) = -.593$, $p = .003$. We found the same relationship in the lvPPA group, $r(18) = -.488$, $p = .04$. In the bvFTD group, people with lower MMSE scores had a lower combination ratio, $r(12) = .673$, $p = .016$, a lower content word ratio, $r(12) = .725$, $p = .008$ and higher bigram collocation strength, $r(12) = -.587$, $p = .045$. MMSE scores did not correlate with language variables in nfvPPA.

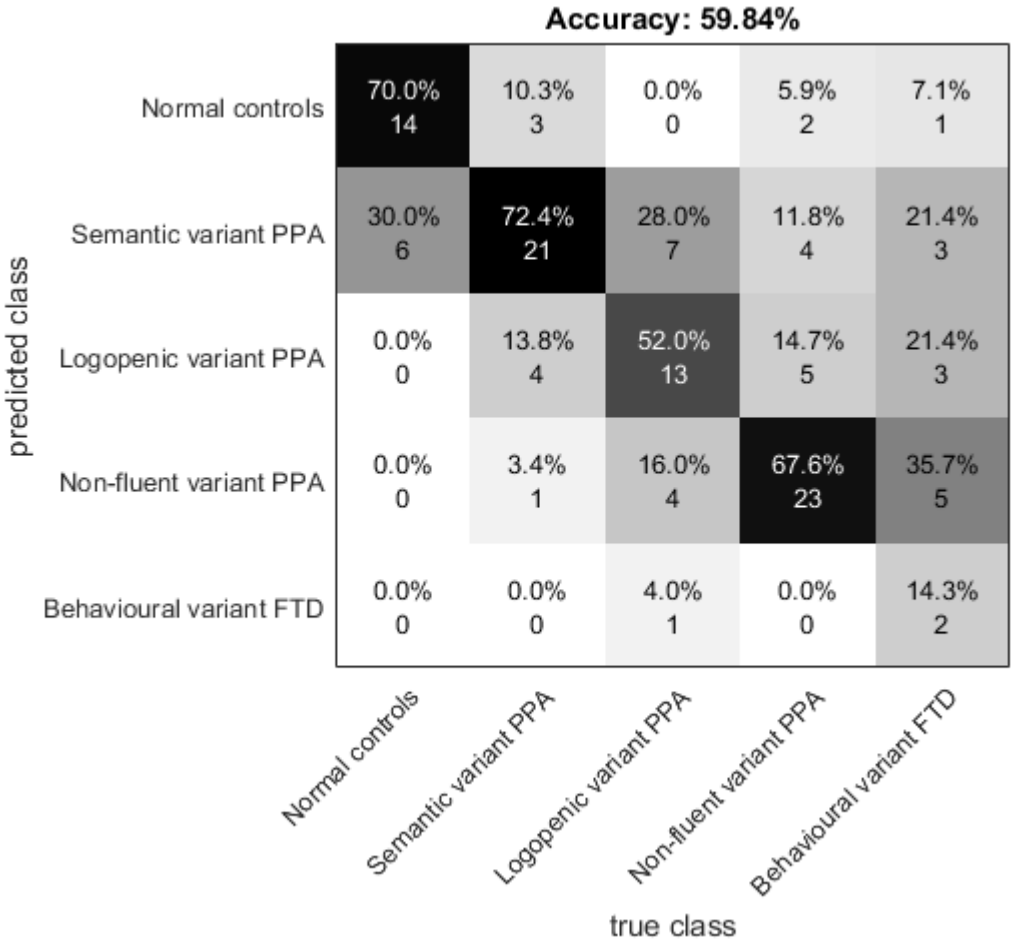
3.3 Machine learning classifier of dementia type

In a final, post-hoc analysis we applied machine learning methods to determine in how FLAT data can be used to categorize participants into the five groups based solely on their last holiday samples. We used a SVM, a complementary linear classifier approach to compare the five groups based on the results of the FLAT analysis. The SVM classifier was trained on a subset of the dataset (80%, randomly selected) to categorize each individual as belonging to one of the five groups based on a linear combination of the five variables (word count, combination ratio, content word ratio, content word frequency, collocation strength). The prediction accuracy of the obtained classifier was then evaluated on the remaining 20% of the data set (test set), in terms of the percentage of correct prediction, and the confusion matrix. This procedure was then repeated five times so that each participant was in the test set once. The accuracy of the five tests was averaged to determine overall accuracy of the model.

Against a chance level of 20%, the SVM classifier achieved a prediction accuracy of 59.8%. The confusion matrix for the model prediction is shown in Figure 2. The model was most successful in identifying NCs (70% correct classification), speakers with svPPA (72.4% correct) and speakers with nfvPPA (67.6% correct). The model's performance was less accurate for lvPPA (52% correct), and strikingly inaccurate for bvFTD (14.3% correct), with 78.5% of bvFTD samples being classified as one of the PPA types. Of all NCs, 30% (six participants) were classified as svPPA. Six participants with dementia (10.3% of speakers with svPPA, 5.9% of speakers with nfvPPA, 7.1% of speakers with bvFTD) were classified as NCs. The confusion matrix is therefore in line with the group comparisons which showed that speakers with svPPA

were the group most similar to NCs. It also illustrates the large overlap between language features in bvFTD, as measured by FLAT, and the features of PPAs. If regarded as a detector of dementia in general (two categories: NCs vs. pooled dementia groups), accuracy increases to 90%, with a true positive rate of 94% and a false positive rate of 30%.

Figure 2. Confusion matrix based on the SVM classifier. Columns display the diagnosis on the basis of diagnosis. Rows show the predicted class on the basis of the five FLAT variables. Percentages represent the proportion of members of the true class with a given diagnosis. Beneath is a raw count of the same members. Prediction accuracy was measured on data on which the model had not been trained.



4. Discussion

We ran an automated analysis of spontaneous language production in healthy speakers, canonical PPAs and bvFTD. Group comparisons showed significant differences between all dementia groups and healthy speakers as well as between each dementia subtype (Table 1). In some groups, language variables

correlated with MMSE scores. A machine learning classifier using the output from the analysis categorized samples with a success rate three times better than chance.

Current guidelines mostly limit their description of spontaneous output to reduction, grammatical simplification or errors. There is strong evidence for another type of simplification, namely increased use of common words and/or phrases. Previous research showed that speakers with svPPA and, to a lesser degree, speakers with nfvPPA, tend to produce more common words.[4] Our data support these conclusions and show that this pattern also exists in lvPPA. We also saw differences in the ratio of content words. Speakers with svPPA had a lower content word ratio, consistent with reports of higher pronoun use as a sign of semantic difficulties (e.g. "it" instead of a concrete noun).[4] Speakers with lvPPA too displayed a lower ratio, likely a result of lexical impairment.

Moreover, we show that speakers with nfPPA and lvPPA use more familiar word combinations, independently of individual word frequency. According to usage-based language theories such as Construction Grammar,[14] familiar combinations are not produced by retrieving each individual word and combining words using abstract grammatical rules, but by accessing holistic representations of the entire phrase or sentence. In many cases, word combinations may be represented as a single "word". Holistic processing poses fewer demands on cognitive systems both at the lexical level, as fewer units need to be retrieved, and at the grammatical level, as fewer combinatorial operations need to be performed. Lexical and grammatical impairment should therefore result in an overuse of familiar forms (sometimes called "formulaic language"[15]), and our data suggest that this is the case for nfvPPA, primarily a grammatical impairment, and lvPPA, where lexical processing is primarily impaired. Similar analyses have shown that this is also the case for Alzheimer's disease and focal aphasia.[5,10] Overuse of formulaic language can have a substantial impact on communication, and therefore quality of life. A speaker who is more restricted to familiar sentences and phrases will find it difficult to adapt to unfamiliar situations and speak about new thoughts and needs.[16] This is one of the ways in which language measures not only indicate cognitive change, but reflect how well the individual can socialize by holding conversations or consuming media.

Formulaic language can also explain why differences in content/function word ratio were relatively small even between non-fluent speakers and NCs. Many strongly collocated combinations contain function word clusters, such as *I don't* in the expression *I don't know*. They make up for a substantial proportion of aphasic production.[10] So while one striking feature of non-fluent aphasia is omission of function words, resulting in a "telegraphic" style, formula overuse can result in non-fluent speakers producing a

similar or ratio of content words than neurotypical speakers over the entire sample, as is the case in the current study.

Our data also support earlier studies which found that bvFTD comes with significant changes in linguistic behavior.[12] In our bvFTD samples, language change affected less the quality of words and word combinations, and rather their quantity, as speakers spoke less and produced shorter utterances. This reduction was roughly on a par with speakers with lvPPA, but while in the latter group this likely reflects impairment of linguistic representations, the linguistic profile in bvFTD may be rather explained by impairment to executive function and related changes to mood and social behavior. However, differences between bvFTD and NC groups in content word ratio and usage frequency were significant before correction for multiple comparisons (Appendix A), and these findings are consistent with results from studies of word naming and comprehension in bvFTD that also revealed lexical disruption.[12] Similarly, other differences which were close to significance thresholds may prove to be important for modelling language in dementia.

In PPAs, language change is the most important indicator of disease onset and progression, at least in early and mid-stages. A longitudinal study may shed more light in this case. MMSE scores are another measure of progression, and language variables generally correlated with these. Two groups however stood out: in speakers with nfvPPA, none of the variables were significantly associated with MMSE scores, and in the bvFTD group, multiple variables were associated, including some which did not distinguish that group from NCs. MMSE scores are more indicative of cognitive impairment in bvFTD than in PPAs,[17] which suggests that the differences observed in bvFTD are indeed related to executive dysfunction, while in speakers with PPA MMSE performance is more dependent on the language components of the test and the individual's ability to understand instructions.

Our application of a machine learning classifier was a first exploration into using FLAT as a classification tool. While the overall accuracy is encouraging, more work needs to be carried out to understand its errors and to build better models. Some errors may reflect the nature of the dementia subtypes. In particular, svPPA and nfvPPA has been associated with distinct language features, while lvPPA has been more difficult to classify, its most distinguishing feature being difficulties in sentence repetition.[3] The model had the most difficulties with bvFTD, and it is possible that this subtype, while causing significant changes in language use, either has no clear language production profile at all, or has one that cannot be captured using the variables which were part of our design. However, errors can also be caused by small sample sizes. Given that the bvFTD group was the smallest of the groups, low accuracy can be the result

of the relatively small number of data points that were available to training the algorithm. Finally, a large number of NCs have been classified under svPPA. While this may reflect limitations of the model, it is imaginable that some speakers in this subset were at risk, or in early stages of a neurological condition. Future studies can explore these explanations by implementing cognitive testing for NCs, or by examining them longitudinally.

At this stage however it is safer to assume in case of errors, the language analysis is at fault, and consider further steps for improvement. Sample sizes can be increased, other language elicitation tasks can be chosen and more variables included. There is to date no attempt of a full integrative language model for dementias which includes properties of the acoustic signal (such as prosody and speech pauses),[7,18,19] phonological, lexical and grammatical properties. A range of biographical variables can affect language use, and only very large databases will be able effectively explore the effects of education, socio-economic status and gender in conjunction with age. Ultimately, the trajectory of change within an individual may be crucial. The most effective use of language as a marker of cognitive function may involve multiple measurements over an individual in conjunction a multifactorial model of language change in ageing.

Our study demonstrates that for additional sensitivity of the model, language features in production should ideally not be coded as binary variables (e.g. “present” or “absent”), but as continuous (degree of deviation from typical language). The advantages of computerized tools are the speed of analysis and the lack of bias as no raters are involved. Recordings of a few hundred words, as used in this study, can be made quickly and at bedside. As transcription technology advances, computerized methods like that described here will become increasingly practical and widely available for clinical applications. This in turn would open up the potential use of such methods as biomarkers, to detect more subtle and ecologically relevant dysfunction of the language system and to track this over time.

Authorship statements and acknowledgments

This project was conceived by VCZ, RAV and JDW. Data were collected by CJDH, RLB, and LR. Samples were transcribed by JE and SD. The analysis was carried out by VCZ, JE and SD, with help from CJDH. The classifier analysis was carried out by LV. VCZ wrote the first draft of the manuscript. CJDH, LV, RAV and JDW contributed to subsequent drafts.

This project was supported by Alzheimer’s Society grants to RAV, VCZ and JDW. CJDH is supported by an Action on Hearing Loss – Dunhill Medical Trust Pauline Ashley Postdoctoral Fellowship. JDR is an MRC

Clinician Scientist. This work was also supported by the NIHR UCLH Biomedical Research Centre and the Wellcome Trust (Senior Fellowship in Clinical Science to JDW). RLB is supported by an MRC studentship.

References

1. Gorno-Tempini ML, Hillis AE, Weintraub S, et al. Classification of primary progressive aphasia and its variants. Neurology [online serial]. 2011;76:1006–1014. Accessed at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3059138&tool=pmcentrez&rendertype=abstract. Accessed January 24, 2014.

2. Wilson SM, Henry ML, Besbris M, et al. Connected speech production in three variants of temporary progressive aphasia. Brain. 2010;133:2069–2088.

3. Marshall CR, Hardy CJD, Volkmer A, et al. Primary progressive aphasia: a clinical approach. J Neurol [online serial]. Springer Berlin Heidelberg; 2018;265:1474–1490. Accessed at: http://link.springer.com/10.1007/s00415-018-8762-6. Accessed August 21, 2018.

4. Fraser KC, Meltzer JA, Graham NL, et al. Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex [online serial]. 2014;55:43–60. Accessed at: http://www.sciencedirect.com/science/article/pii/S0010945212003413. Accessed December 12, 2013.

5. Zimmerer VC, Wibrow M, Varley RA. Formulaic language in people with probable Alzheimer's Disease: a frequency-based approach. J Alzheimer's Dis. 2016;53:1145–1160.

6. Fraser KC, Meltzer JA, Rudzicz F. Linguistic Features Identify Alzheimer's Disease in Narrative Speech. J Alzheimers Dis [online serial]. IOS Press; 2015;49:407–422. Accessed at: http://content.iospress.com/articles/journal-of-alzheimers-disease/jad150520. Accessed December 3, 2015.

7. Nevler N, Ash S, Jester C, Irwin DJ, Liberman M, Grossman M. Automatic measurement of prosody in behavioral variant FTD. Neurology [online serial]. 2017;89:650–656. Accessed at: http://www.ncbi.nlm.nih.gov/pubmed/28724588. Accessed August 21, 2018.

8. Ash S, Evans E, O'Shea J, et al. Differentiating primary progressive aphasias in a brief sample of connected speech. Neurology [online serial]. 2013;81:329–336. Accessed at: http://www.ncbi.nlm.nih.gov/pubmed/23794681. Accessed November 9, 2018.

9. Siyanova-Chanturia A, Conklin K, Caffarra S, Kaan E, van Heuven WJB. Representation and processing of multi-word expressions in the brain. Brain Lang [online serial]. 2017;175:111–122.

Accessed at: http://www.ncbi.nlm.nih.gov/pubmed/29078151. Accessed December 8, 2017.

10. Zimmerer VC, Newman L, Thomson R, Coleman MJ, Varley RA. Automated analysis of language production in aphasia and right hemisphere damage: Frequency and collocation strength. Aphasiology. 2018;32:1267–1283.

11. Bruns C, Varley RA, Zimmerer VC, Carragher M, Brekelmans G, Beeke S. "I don't know": a usage-based approach to familiar collocations in non-fluent aphasia. Aphasiology. Epub 2018.

12. Hardy CJD, Buckley AH, Downey LE, et al. The language profile of behavioral variant frontotemporal dementia. J Alzheimer's Dis. 2015;50.

13. Rascovsky K, Hodges J, Knopman D, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia [online]. BRAIN OXFORD UNIV PRESS; 2011. Accessed at: http://discovery.ucl.ac.uk/1325627/. Accessed January 15, 2014.

14. Goldberg AE. Constructions: a new theoretical approach to language. TRENDS Cogn Sci. 2003;7:219–224.

15. Wray A. What do we (think we) know about formulaic language? An evaluation of the current state of play. Annu Rev Appl Linguist. 2012;32:231–254.

16. Wray A. Formulaic language as a barrier to effective communication with people with Alzheimer's Disease. Can Mod Lang Rev. 2011;67:429–458.

17. Osher JE, Wicklund AH, Rademaker A, Johnson N, Weintraub S. The Mini-Mental State Examination in Behavioral Variant Frontotemporal Dementia and Primary Progressive Aphasia. Am J Alzheimer's Dis Other Dementiasr [online serial]. Sage PublicationsSage CA: Los Angeles, CA; 2008;22:468–473. Accessed at: http://journals.sagepub.com/doi/10.1177/1533317507307173. Accessed September 28, 2018.

18. López-de-Ipiña K, Alonso JB, Solé-Casals J, et al. On Automatic Diagnosis of Alzheimer's Disease Based on Spontaneous Speech Analysis and Emotional Temperature. Cognit Comput [online serial]. Springer US; 2015;7:44–55. Accessed at: http://link.springer.com/10.1007/s12559-013-9229-9. Accessed September 28, 2018.

19. Angelopoulou G, Kasselimis D, Makrydakis G, et al. Silent pauses in aphasia. Neuropsychologia [online serial]. 2018;114:41–49. Accessed at: http://www.ncbi.nlm.nih.gov/pubmed/29634961. Accessed October 1, 2018.

Appendix A. Pairwise comparisons of properties of spontaneous language output, with age entered as covariate. Suggested Bonferroni adjustment of significance threshold due to ten comparisons for each variable/hypothesis: $p < .005$. MD = mean difference, calculated as: mean of group in the left column subtracted by the mean of the group in the top row. PPA = primary progressive aphasia; FTD = fronto-temporal dementia.

Word count	Semantic variant PPA	Logopenic variant PPA	Non-fluent variant PPA	Behavioral variant FTD
Neurotypical controls	MD = 12 95% CI [-60, 83] p = .749	MD = 106 95% CI [32, 180] p = .006	MD = 177 95% CI [105, 248] p < .001	MD = 145 95% CI [59, 231] p = .001
Semantic variant PPA		MD = 94.124 95% CI [27, 162] p = .007	MD = 165 95% CI [101, 229] p < .001	MD = 133 95% CI [53, 214] p = 001
Logopenic variant PPA			MD = 71 95% CI [4, 138] p = .039	MD = 40 95% CI [-43, 122] p = .346
Non-fluent variant PPA				MD = -31 95% CI [-111, 48] p = .436

Combination ratio	Semantic variant PPA	Logopenic variant PPA	Non-fluent variant PPA	Behavioral variant FTD
Neurotypical controls	MD = .08 95% CI [.01, .14] p = .021	MD = .18 95% CI [.11, .24] p < .001	MD = .35 95% CI [.29, .41] p < .001	MD = .18 95% CI [.1, .25] p < .001
Semantic variant PPA		MD = .1 95% CI [.04, .16] p = .001	MD = .27 95% CI [.22, .33] p < .001	MD = .1 95% CI [.03, .17] p = 006
Logopenic variant PPA			MD = .17 95% CI [.11, .23]	MD = -.001 95% CI [-.07, .07]

			p < .001	p = .985
Non-fluent variant PPA				MD = -.17 95% CI [-.24, -.1] p < .001

Content word ratio	Semantic variant PPA	Logopenic variant PPA	Non-fluent variant PPA	Behavioral variant FTD
Neurotypical controls	MD = .38 95% CI [.002, .075] p = .039	MD = .09 95% CI [.05, .13] p < .001	MD = .03 95% CI [-.002, .07] p = .065	MD = .05 95% CI [.003, .09] p = .035
Semantic variant PPA		MD = .05 95% CI [.02, .09] p = .004	MD = -.004 95% CI [-.04, .03] p = .8	MD = .009 95% CI [-.03, .5] p = .67
Logopenic variant PPA			MD = -.06 95% CI [-.09, -.02] p = .002	MD = -.04 95% CI [-.08, 0] p = .048
Non-fluent variant PPA				MD = .01 95% CI [-.03, .05] p = .525

Content word frequency	Semantic variant PPA	Logopenic variant PPA	Non-fluent variant PPA	Behavioral variant FTD
Neurotypical controls	MD = -445 95% CI [-677, -214] p < .001	MD = -754 95% CI [-993, -515] p < .001	MD = -442 95% CI [-673, -211] p < .001	MD = -296 95% CI [-573, -179] p = .037
Semantic variant PPA		MD = -309 95% CI [-526, -914] p = .006	MD = 3 95% CI [-203, 210] p = .975	MD = 150 95% CI [-110, 409] p = .255

Logopenic variant PPA			MD = 312 95% CI [96, 538] p = .005	MD = 458 95% CI [193, 724] p = .001
Non-fluent variant PPA				MD = 146 95% CI [-110, 403] p = .26

Bigram collocation strength	Semantic variant PPA	Logopenic variant PPA	Non-fluent variant PPA	Behavioral variant FTD
Neurotypical controls	MD = -5.3 95% CI [-11.3, .8] p = .085	MD = -10.9 95% CI [-17.1, -4.7] p = .001	MD = -9 95% CI [-15, -3] p = .004	MD = -6.1 95% CI [-13.6, 1.1] p = .097
Semantic variant PPA		MD = -5.6 95% CI [-11.3, 0] p = .052	MD = -3.8 95% CI [-9.1, 1.6] p = .169	MD = -.8 95% CI [-7.6, 5.9] p = .807
Logopenic variant PPA			MD = 1.9 95% CI [-3.8, 7.5] p = .514	MD = 4.8 95% CI [-2.1, 11.7] p = .174
Non-fluent variant PPA				MD = 2.9 95% CI [-3.8, 9.6] p = .388