Published in final edited form as:

J Neurolinguistics. 2010 March 1; 23(2): 127–144. doi:10.1016/j.jneuroling.2009.12.001.

A computerized technique to assess language use patterns in patients with frontotemporal dementia

Serguei VS Pakhomov, Ph. D.[†], Glenn E. Smith, PhD.^{††}, Susan Marino, PhD[†], Angela Birnbaum, PhD[†], Neill Graff-Radford, M.D.^{†††}, Richard Caselli, M.D.^{*}, Bradley Boeve, MD^{††}, and David S. Knopman, M.D.^{††}

[†] Center for Clinical and Cognitive Neuropharmacology, University of Minnesota, Twin Cities, Minnesota

^{††} Mayo Alzheimer's Disease Research Center, Rochester, Minnesota

^{†††} Department of Neurology, Mayo Clinic, Jacksonville, Florida

^{*} Department of Neurology, Mayo Clinic, Scottsdale, Arizona

Abstract

Frontotemporal lobar degeneration (FTLD) is a neurodegenerative disorder that affects language. We applied a computerized information-theoretic technique to assess the type and severity of language-related FTLD symptoms. Audio-recorded samples of 48 FTLD patients from three participating medical centers were elicited using the Cookie Theft picture stimulus. The audio was transcribed and analyzed by calculating two measures: a perplexity index and an out-of-vocabulary (OOV) rate. The perplexity index represents the degree of deviation in word patterns used by FTLD patients compared to patterns of healthy adults. The OOV rate represents the proportion of words used by FTLD patients that were not used by the healthy speakers to describe the stimulus. In this clinically well-characterized cohort, the perplexity index and the OOV rate were sensitive to spontaneous language manifestations of semantic dementia and the distinction between semantic dementia and progressive logopenic aphasia variants of FTLD. Our study not only supports a novel technique for the characterization of language-related symptoms of FTLD in clinical trial settings, it also validates the basis for the clinical diagnosis of semantic dementia as a distinct syndrome.

Keywords

frontotemporal lobar degeneration; semantic dementia; perplexity; entropy; statistical language modeling

1. Introduction

Frontotemporal lobar degeneration (FTLD) is a neurodegenerative disorder that severely affects cognitive function and, in many cases, manifests itself through impaired language

Corresponding author: Serguei VS Pakhomov, 7-125F Weaver-Densford Hall, 308 Harvard St. S.E., Minneapolis, MN, Tel: +1 612-624-1198, Fax: +1 612-625-9931, pakh0002@umn.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

use (Kertesz, McMonagle, Blair, Davidson, & Munoz, 2005). Currently, FTLT comprises 4 syndromes: behavioral variant frontotemporal dementia (bvFTD), progressive non-fluent aphasia (PNFA), progressive logopenic aphasia (PLA) and semantic dementia (SD). These syndromes are typically diagnosed using standard clinical criteria, neuropsychological testing and neuroimaging; however, the definition of syndromes and phenotypes remains a key theme in research on dementia in general and FTLT in particular (Rascovsky, et al., 2007). Although neuroimaging is a powerful way to determine structural changes associated with FTLT, careful clinical evaluation remains critical to FTLT diagnosis, particularly in the early stages of disease progression. FTLT currently has no known cure, but research efforts are underway to design and test therapeutic interventions. In order to assess the efficacy of therapies and to characterize the disease progression, consistent and objective instruments are required for measuring changes in cognition manifest in language.

1.1 Speech and language characteristics in FTLT

Over half of all patients with FTLT exhibit language-related symptoms on initial presentation (Hodges, et al., 2004). A number of speech and language characteristics were shown to be differentially sensitive to the effects of FTLT variants. The progressive non-fluent aphasia variant has been characterized in terms of dysfluent, effortful, and agrammatical speech (Ash, et al., 2008; Bird, Lambon Ralph, Patterson, & Hodges, 2000; Gorno-Tempini, et al., 2004; Grossman, 2002; Pelle, Cooke, Moore, Vesely, & Grossman, 2007; Weintraub, Rubin, & Mesulam, 1990). The semantic dementia variant involves multi-modal non-verbal, as well as verbal, naming and recognition deficits with relatively preserved grammar (Hodges, Patterson, Oxbury, & Funnell, 1992; Neary, et al., 1998). However, despite these differences between the non-fluent and fluent aphasic variants of FTLT, there is considerable overlap between their language specific manifestations (Thompson, Ballard, Tait, Weintraub, & Mesulam, 1997). Apart from the overlap between fluent and non-fluent types of primary progressive aphasia, the distinction between the fluent subtype of aphasia and semantic dementia is also being debated. Some researchers treat the not otherwise specified primary progressive aphasia (PPA - NOS) as distinct from either semantic dementia or progressive non-fluent aphasia variants of FTLT (Josephs, et al., 2006). However, the distinction between these two classifications may be a matter of emphasis rather than differences in the underlying pathophysiology of the phenomenon (Adlam, et al., 2006).

Although the behavioral, progressive non-fluent aphasia and semantic dementia syndromes are likely to represent FTLT pathologically (Knopman, et al., 2008), the grouping of the progressive logopenic aphasia syndrome with FTLT vs. Alzheimer's disease is debatable. Similarly to progressive non-fluent aphasia, spontaneous speech production in progressive logopenic aphasia has also been characterized by slower speaking rate, hesitations and pauses attributable to word-finding difficulties (Gorno-Tempini, et al., 2008). Some of the cases of primary progressive aphasia distinct from both semantic dementia and progressive non-fluent aphasia also exhibited these altered prosodic characteristics of speech with relatively preserved grammar, and could possibly be classified as progressive logopenic aphasia (Josephs, et al., 2006).

In summary, the characterization of FTLT variants remains challenging and necessitates further investigation of novel techniques for the assessment of the linguistic aspects of the disorder.

1.2 Quantitative analysis of speech and language in semantic dementia

A number of diverse speech and language features have been identified and used to characterize fluent primary progressive aphasia and semantic dementia in general, and the

semantic dementia variant of FTLN in particular. Gordon (Gordon, 2006) used a Quantitative Production Analysis protocol (Berndt, Wayland, Rochon, Saffran, & Schwartz, 2000; Saffran, Berndt, & Schwartz, 1989) to compare fluent and non-fluent aphasic speech productions elicited with a picture description task. The measures used in the Quantitative Production Analysis protocol were found to be sensitive to the severity of both fluent and non-fluent aphasia, but could not reliably discriminate between these two subtypes. In a subsequent study, Gordon (Gordon, 2008) tested additional measures of correct information units (Nicholas & Brookshire, 1993; Yorkston & Beukelman, 1980) and type-to-token ratio. Although these measures correlated with those obtained with the Quantitative Production Analysis protocol and were sensitive to aphasia severity, they also failed to distinguish between fluent and non-fluent groups.

Our study addresses the need for quantitative and objective instruments sensitive to language manifestations of dementia by making use of the fact that patients with semantic dementia are more likely to experience word-finding difficulties (Amici, Gorno-Tempini, Ogar, Dronkers, & Miller, 2006; Bird, et al., 2000; Hodges, et al., 1992; Neary, et al., 1998; Snowden, 1999; Westbury & Bub, 1997). Thus their speech, while fluent, tends to contain unexpected, albeit mostly understandable, words and word sequences (e.g., “she is doing too dropping too much water” to describe a woman standing by a kitchen sink that’s overflowing with water). Our methodology for capturing and quantifying such unusual words and sequences of words relies on the notion of language model perplexity originally developed for conducting research on automatic speech recognition and natural language processing. The technique consists of constructing a statistical language model (detailed in the Methods) based on language samples from one population (e.g., picture descriptions by healthy adults) and using this model to predict word sequences in language samples from another population (e.g., picture descriptions by patients with FTLN). A model that is efficient in predicting such word sequences is said to have lower *perplexity* (Bahl, Baker, Jelinek, & Mercer, 1977). Thus, theoretically, the unexpected word sequences (measured by perplexity) and unexpected words (measured by the out-of-vocabulary rate) found in the speech of patients with semantic dementia are likely to result in higher values, which may be used to index the degree of impairment to semantic networks in patients with FTLN, as well as other forms of dementia (Roark, Hosom, Mitchell, & Kaye, 2007).

Our study investigated the use of information-theoretic measures (perplexity index and out-of-vocabulary rate) to measure the degree of deviation in utterances produced by patients with FTLN on a picture description task from those of healthy adults. We expected to find significant differences in the perplexity score and the out-of-vocabulary rate among at least some of the FTLN variants. The perplexity score was expected to be low for the behavioral variant, as their picture descriptions sounded closest to those produced by healthy adults. We also expected the out-of-vocabulary rate to be high for the semantic dementia variant, as patients with this variant were anticipated to have word finding difficulties. Thus these patients would be more likely to substitute words used by healthy adults on this picture description task with either neologisms or other vocabulary that was not used by healthy adults performing the same task.

2. Methods

The overall study design is illustrated in Figure 1. The study took place in two phases. In Phase I, we constructed a statistical language model that was subsequently used in Phase II to assess the language contained in picture descriptions provided by the study participants.

2.1 Participants

All aspects of these studies have been approved by the Institutional Review Boards at the Mayo Clinic as well as the University of Minnesota. A total of 80 subjects participated in this study. The patient group consisted of 48 people diagnosed with one of the 4 syndromes (behavioral variant frontotemporal dementia (n=19), progressive non-fluent aphasia (n=12), progressive logopenic aphasia (n=6) and semantic dementia (n=11)). These patients were recruited for the study at 3 academic medical centers – Mayo Clinic (Rochester, MN, Scottsdale, AZ, Jacksonville, FL). There were two control groups consisting of younger and older adults. The younger control group consisted of 23 volunteers recruited at the University of (ANONYMIZED). The older control group consisted of 9 nursing home residents recruited at three nursing home facilities in the Minneapolis/St. Paul metropolitan area. The nursing home residents were selected from a random sample based on a manual review of their medical charts to exclude anyone with a diagnosis of dementia. The controls were used during Phase I for statistical language model development, which was applied in Phase II to assess language differences among the four groups of FTLT patients and compare them to the two control groups.

2.2 Diagnostic criteria

Diagnostic criteria for FTLT variants have been previously reported (Knopman, et al., 2008) and are briefly summarized below. The exclusion/inclusion criteria for this study were based on the Neary criteria (Neary, et al., 1998) and are also described in detail in a previous study (Knopman, et al., 2007). The initial diagnosis was made by neurologists skilled in the diagnosis of FTLT using these criteria. The neuropsychological tests described in this study were not used in the initial diagnosis and were intended as part of a longitudinal battery investigating the suitability of standard neuropsychological tests in clinical trials. In addition, to support the diagnosis of FTLT, all patients were required to have imaging studies demonstrating focal cerebral atrophy consistent with a degenerative etiology. In brief, we defined the following 4 syndromes:

Behavioral variant frontotemporal dementia (bvFTD) was diagnosed with a change in personality and behavior sufficient to interfere with work or interpersonal relationships. These symptoms constituted the principal deficits and the initial presentation and with at least 5 core symptoms in the domains of aberrant personal conduct and impaired interpersonal relationships.

Progressive non-fluent aphasia (PNFA) was diagnosed with expressive speech characterized by at least 3 of the following: reduced numbers of words per utterance, speech hesitancy or labored speech, word finding difficulty, or agrammatism, where these symptoms constitute the principal deficits and the initial presentation.

Progressive logopenic aphasia (PLA) was diagnosed with anomia but intact word meaning and object recognition, where these symptoms constitute the principal deficits and the initial presentation. Progressive logopenic aphasia was treated as a category separate from progressive non-fluent aphasia and semantic dementia.

Semantic dementia (SD) was diagnosed with loss of comprehension of word meaning, object identity or face identity, where these symptoms constitute the principal deficits and the initial presentation.

2.3 Clinical Assessments

We used standard manually administered and scored Clinical Dementia Rating (CDR) scales (Morris, 1993) consisting of six dimensions (Memory, Orientation, Judgment, Community

affairs, Home and hobbies, and Personal care) augmented to assess the FTLT syndromes. The augmentation consisted of two additional dimensions: Behavioral, Comportment and Personality scale, and the Language-specific scale. Generally, the scores on the CDR scales range between 0 and 3 and represent normal functioning (0), minimal impairment (0.5), mild impairment (1), moderate impairment (2), or severe impairment (3). Further details on the use of FTLT specific CDR scales are available elsewhere (Knopman, et al., 2008); however, since the language specific dimension is particularly relevant to the current study, we describe it here in more detail for convenience. The score of 0 on the Language-specific CDR scale indicates normal speech and comprehension, 0.5 – minimal but noticeable word-finding problems, minimal dysfluency and normal comprehension, 1 – mild word finding problems that do not significantly degrade speech or mild comprehension difficulties, 2 – moderate word finding problems that interfere significantly with communication and moderate dysfluency and comprehension difficulty, 3 – severe deficits in word finding, expressive speech and comprehension making conversation virtually non-existent. The CDR scales were dichotomized in order to separate participants with no or mild impairment (CDR < 2) from participants with moderate to severe impairment (CDR ≥ 2). In addition to the eight individual dimensions, we calculated their sum (CDRTOTAL variable). The CDRTOTAL variable was dichotomized using 8 as the cutoff representing the sum of maximum values for no or mild dementia across all eight dimensions.

2.4 Cognitive Measures

As part of another longitudinal study, all 48 FTLT patients underwent a standard neuropsychological test battery which included the Boston Diagnostic Aphasia Examination Cookie-Theft Picture Description Task (Goodglass & Kaplan, 1983). The Cookie Theft picture stimulus was also used to collect speech samples from the control subjects. In addition to the Cookie Theft stimulus, all of the 48 FTLT patients were administered a standard neuropsychological test battery consisting of the following tests: California Verbal Learning Test (CVLT) Free and Delayed Recall (Delis, Kramer, Kaplan, & Ober, 2000), Simplified Trail Making (Part A only) (Knopman, et al., 2008), Two-number Number Cancellation (Mohs, et al., 1997), Digits Backward Test from Wechsler Memory Scale-Revised (Wechsler, 1987), Stroop Test (Stroop, 1935), Digit-Symbol Substitution Test (Wechsler, 1981), Verbal Fluency Test for Letters and Categories (Benton, Hamsher, & Sivan, 1983), Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1978), and the Wechsler Adult Intelligence Scale - Revised (WAIS-R) Verbal Similarities Test (Wechsler, 1981). The selection of the tests was dictated by their performance in the FTLT population (Kramer, et al., 2003) as well as pragmatic and logistical considerations. The test battery was targeted to be limited to under 1 hour and to contain a mix of tests requiring verbal and non-verbal responses that are not too easy or too difficult for the patients (Knopman, et al., 2008). All tests were scored by board-certified behavioral neuropsychologists.

2.5 Speech Transcription

The speech obtained from each subject on the picture description task was digitized and subsequently manually transcribed by a staff member trained to perform verbatim transcription. An example of a transcribed segment is shown below in (1):

- (1) ... E_go E_ah ahead there's a mother T_NOISE FILLEDPAUSE_ah there's a boy T_BREATH and there's g- FILLEDPAUSE_ah j- jub a little girl ...

where “E_” indicates the speech that belongs to the examiner and “T_” indicates non-speech events. We transcribed all speech and non-speech acoustic events including loud breathing, throat clearing and laughter, speech dysfluencies consisting of filled pauses (um's and ah's) and false starts (e.g. “g- j-” in “g- j- jub”) as well as backchannels (e.g. “yeah” and “uh-

huh”). However, these speech and non-speech events were subsequently removed from the data prior to analysis. Phonological distortions due to possible dysarthria were transcribed phonetically to the best of the transcriptionist’s ability. Difficult cases with speech overlap and excessive noise were resolved through consultation with one of the study investigators (SP). On average, the transcription time for each subject’s picture description was approximately 15 minutes.

2.6 Statistical Language Model

To represent the language use patterns in healthy adults, we trained a statistical language model based on the data from 15 younger controls. The 8 remaining younger controls as well as the 9 elderly controls were used to establish the perplexity and out-of-vocabulary rate measurements that were compared with those of the FTLN subjects.

This statistical model captures the probabilities of 1 and 2 word sequences occurring in verbal descriptions of the picture stimulus. Below is an excerpt from the model trained for this study using the Hidden Markov Toolkit (v3.4) (Young, et al., 2006).

(1)

```
/data/
/1-grams:
...
-2.1088 jar
-2.9839 just
-3.2849 keeping
-2.9839 kid
-3.2849 kid's
-2.2435 kids
...
/2-grams:
-0.9171 kid looks
-0.9171 kid stealing
-0.6161 kid's falling
-1.6575 kids appear
-0.5306 kids are
-0.9472 kids have
-1.6575 kids stealing
```

The first column contains log probabilities (base 10) of 1 and 2 word sequences found in the picture descriptions used for training of the model. For example, the probability of the sequence “kids are” is $10^{-0.5306} = 0.29$, whereas the probability of the sequence “kids have” is $10^{-0.9472} = 0.11$. Thus, this model simply reflects the fact that we are more likely to see the word “kids” followed by the word “are” than by the word “have” as estimated from the speech of healthy adults. This statistical model, to which we will refer as the “BDAE Model”, was then used to assess the speech samples recorded from FTLN patients.

Roark and colleagues (Roark, Mitchell, & Hollingshead, 2007) have previously used an information-theoretic measure of cross-entropy between a statistical part-of-speech model and speech obtained from patients with mild cognitive impairment. In general, cross-entropy constitutes an upper bound on the entropy of a stochastic process. When applied to human

language, entropy measures how much information is encoded by the grammar of the language and has been experimentally shown to be correlated with the amount of effort involved in processing sentences (Keller, 2004). Perplexity is a more readily interpretable derivative of cross-entropy; however, the two measures represent the same property of statistical language models – their ability to predict words in new utterances. For example, the perplexity of 173.1 on a set of picture descriptions by patients with FTLN may be interpreted as the language model having to make on average 173 independent choices to predict each word in the text of the descriptions. Thus the notion of perplexity may be regarded as a way to indirectly capture deviations in local (span of 2–3 words) syntactic and semantic dependencies from the “norm” represented by the language model. A more in-depth exposition of both perplexity and cross-entropy can be found in the computational linguistics literature (e.g., (Brown, Della Pietra, Mercer, Della Pietra, & Lai, 1992), (Manning & Shutze, 1999)).

In addition to the perplexity index, we also investigated a measure of the out-of-vocabulary rate for each picture description. The out-of-vocabulary rate represents the percentage of unexpected words that were spoken by the FTLN patients that were not found in the language model trained on healthy participants’ speech. For example, if the subject’s picture description consisted of 100 words not including filled pauses, false starts and unintelligible speech, and 10 of these words were not found in the statistical language model, the out-of-vocabulary rate was calculated to be 10%. Thus, the out-of-vocabulary rate complements the perplexity index by providing additional information on the degree of deviation in the language patterns of FTLN patients from the “norm.”

2.7 Narrative Representations of Semantic Dementia

Bird and colleagues created a set of 6 artificial narratives to simulate the content of Cookie Theft picture descriptions expected to be generated by healthy adults and people with progressively worsening stages of semantic dementia (Bird, et al., 2000). They refer to these narrative representations of semantic dementia as “models”, not to be confused with the statistical language model used in the current study. For clarity, we will refer to Bird’s models as “Narrative Models” in contrast to the “BDAE Model” used in our study.

Bird’s subjects comprise a group completely independent from the subjects recruited for our study. The composite narrative by healthy adults (Narrative Model 1) was based on the content of 20 control subjects’ narratives from Bird et al.’s study. Language manifestations of semantic memory deficits were then simulated by removing low-frequency words from the “healthy” Narrative Model 1 in bands defined by progressively increasing thresholds. The deleted words were replaced with appropriate substitutions frequently heard in the speech of people with progressive fluent aphasia (e.g. “sort of”, “I forget what you call it”, “things on your feet”). Narrative Model 2 excluded words that occurred less than 10 times per million; Narrative Model 3 excluded words occurring less than 32 times per million, Narrative Model 4 – less than 100 times per million; Narrative Model 5 – less than 317 times per million; and Narrative Model 6 – less than 1000 times per million. Thus, Narrative Model 2 represents only a slight impairment, whereas the Narrative Model 6 represents a very severe impairment. The full text of the Narrative Models can be found in the appendix to Bird’s publication (Bird, et al., 2000).

Bird et al. (Bird, et al., 2000) found a striking similarity between these artificial narrative models based on word frequency restrictions and the actual Cookie Theft picture descriptions by 3 patients with semantic dementia in a longitudinal study. This similarity was further validated by a follow-up cross-sectional study of 21 narratives from 8 patients with different semantic dementia severity as determined by standard neuropsychological tests. In our study, we used these 6 Narrative Models created by Bird et al., to provide an

independent test of the hypothesis that the perplexity index is sensitive to language manifestations of semantic memory deterioration. If this hypothesis is correct, we should observe the lowest perplexity index on Bird's Narrative Model 1 (healthy control) and the perplexity index should become progressively higher on subsequent Narrative Models 2 through 6.

2.8 Statistical Analysis

We did not assume that our data were normally distributed; therefore, we used the non-parametric Kruskal-Wallis counterpart to the one-way ANOVA to test for the differences between the subgroups. For those tests indicating significance, we examined pairwise comparisons between the groups using the Mann-Whitney test with p-values adjusted for multiple comparisons using the Holm method. Effect size measures were calculated using the non-parametric equivalent of the eta-square method by taking the ratio of the χ^2 -squared value from the Kruskal-Wallis test to N-1. Correlations between perplexity, out-of-vocabulary rate, clinical and cognitive variables were computed using the Spearman rank correlation method. Regression modeling was performed with standard simple linear regression. Results were considered significant if the p-value was less than 0.05. All statistical computations were carried out using R (version 2.9.1) statistical software package.

3. Results

3.1 Participant characteristics

The mean age of the 48 FTLN patients at the time of the testing was 64.7 (stdev = 8.7). Twenty-three of the FTLN patient (48%) were women, 25 (52%) were men. The mean education was 15.0 (stdev = 2.4) years. Nineteen (39%) had a clinical diagnosis of behavioral variant frontotemporal dementia; twelve (25%) had a diagnosis of progressive non-fluent aphasia; six (13%) had a diagnosis of progressive logopenic aphasia; and eleven (23%) were diagnosed with semantic dementia. The mean scores of the neuropsychological tests stratified by FTLN variants are presented in Table 1. No significant differences according to age were found among any of the four FTLN variants.

The mean age in the younger controls group was 32.5 (stdev = 11.3). The mean age of the older controls group was 72.66 (stdev = 7.30). The mean age of the younger control group was significantly different from all variants in the FTLN group as well as the older control group. The mean age of the older control group was not significantly different from the mean age of the semantic dementia (p-value = 0.98), progressive logopenic aphasia (p-value = 0.44) or progressive non-fluent aphasia (p-value = 0.65) variants. A significant difference in age was found between the behavioral variant and the older controls group (p-value = 0.04) with the subjects in the behavioral variant group being slightly younger than the older controls.

3.2 Statistical Language Model Perplexity

Table 2 shows correlations between the perplexity scores of the BDAE model and the test scores obtained with the neuropsychological test battery. These results indicate that the perplexity of the BDAE model negatively correlated with category fluency but did not correlate with letter fluency. Statistically significant correlations were also found between the BDAE perplexity index and the CVLT Free and Delayed Recall tasks, Boston Naming, and WAIS-R Verbal Similarities test scores.

BDAE Model perplexity index correlated with Memory ($r = 0.35$, p-value < 0.05), Orientation ($r = 0.37$, p-value < 0.05), Language ($r = 0.52$, p-value < 0.01) and CDRTOTAL ($r = 0.34$, p-value < 0.05) CDR dimensions, as illustrated in Table 3. None of the other

dimensions showed significant correlations. The comparison between unimpaired and moderately/severely impaired individuals, also summarized in Table 3, showed that the mean BDAE perplexity scores tended to be lower for the group with CDR scores less than 2 (no or mild impairment). The group with CDR scores of 2 or greater (moderately or severely impaired) had only 3 subjects for Memory and one for Orientation, whereas it had 17 subjects for Language and 6 for CDRTOTAL. This asymmetry indicates a relatively greater proportion of impairment manifest in Language than in other domains such as Memory, Orientation, Judgment, Community Affairs, Home and Hobbies, Personal care, and Behavior.

The means and standard deviations of the BDAE Model perplexity scores for the four diagnostic variants of FTLN are summarized in Figure 2. These results show that the mean perplexity is highest for the semantic dementia variant (111.0) and lowest for the bvFTD group (57.5). The differences between the means among the FTLN variants were statistically significant with Kruskal-Wallis test ($\chi^2 = 20.11$, $df = 5$, $p\text{-value} = 0.001$). Subsequent post-hoc analysis conducted with pair-wise Mann-Whitney tests adjusted for multiple comparisons confirmed a statistically significant difference between a) behavioral and semantic dementia variants ($W = 180$; adjusted $p\text{-value} = 0.009$), b) the semantic dementia variant and younger controls ($W = 88$; adjusted $p\text{-value} = 0.0004$) and older controls ($W = 90$, adjusted $p\text{-value} = 0.016$). The non-parametric eta-square was 0.31 indicating a fairly large effect size. None of the other comparisons revealed significant differences including young vs. old controls.

3.3 Out-of-vocabulary rate

Figure 3 shows the mean out-of-vocabulary rates for the four FTLN variants. The out-of-vocabulary rate is lowest for the progressive logopenic aphasia variant (9.5%) and highest for the semantic dementia variant (17.6). The differences between the out-of-vocabulary rate means among the FTLN variants were statistically significant on Kruskal-Wallis test ($\chi^2 = 13.74$, $df = 5$, $p\text{-value} = 0.017$). Subsequent post-hoc analysis conducted with pair-wise Mann-Whitney tests adjusted for multiple comparisons confirmed a statistically significant difference between a) progressive logopenic aphasia and semantic dementia variants ($W = 66$; adjusted $p\text{-value} = 0.029$), b) semantic dementia and older controls ($W = 91$, adjusted $p\text{-value} = 0.013$). The difference between semantic dementia and the behavioral variant was not significant after adjustment for multiple comparisons ($W = 39$; adjusted = 0.07). None of the other comparisons revealed significant differences including young vs. old controls.

3.4 Perplexity of BDAE Model on narrative representations of semantic dementia

The perplexity indices computed using the BDAE statistical model and the six Narrative Models created by Bird and colleagues to represent different levels of severity of semantic dementia were distributed as illustrated in Figure 4. The perplexity indices increased positively with the degree of semantic dementia simulated with Bird's Narrative Models. A polynomial regression model indicated a strong relationship between the severity of semantic impairment reflected in the Narrative Models and the perplexity scores produced by the BDAE Model ($R^2 = 0.98$; $df = 3$; $p\text{-value} = 0.003$).

4. Discussion

Our study demonstrates a novel use of a standard information-theoretic measure of language model perplexity for the characterization of FTLN syndromes. This study suggests that the perplexity index is sensitive to the differences in speech patterns of patients with semantic dementia and behavioral variant of FTLN. In addition, the out-of-vocabulary rate is sensitive to differences in the speech of patients with semantic dementia and progressive logopenic

aphasia. The perplexity index discriminated mild from moderate-to-severe language impairment across all FTLD variants.

4.1 Perplexity index as a measure of semantic memory impairment in FTLD

The language model based on healthy adults' picture descriptions was the most "perplexed" having on average 111 choices in predicting the next word in the narrative picture descriptions by FTLD participants with the diagnosis of semantic dementia. This perplexity value is almost double that of the means for the behavioral variant. The next highest perplexity (on average 106 choices per word) was obtained with the progressive logopenic aphasia group. The lowest perplexity of 57.5 was obtained from the patients with the behavioral variant. The impairment associated with the behavioral variant affects executive functioning more than language, thus resulting in narratives that are relatively fluent, grammatically and semantically intact with some deficits at the higher discourse level (Ash, et al., 2006). Our results are consistent with these observations as the perplexity mean for the behavioral variant group is only slightly higher than that for the healthy participants group.

Despite the age difference between the young control group and the FTLD patients, the perplexity scores for the young controls were not significantly different from the behavioral variant subgroup but were different from the semantic dementia group. The mean age of the semantic dementia group was not significantly different from the mean age in any other FTLD group including the behavioral variant. This indicates that the perplexity index is not age-sensitive (for the age groups included in this study). Comparisons of mean perplexity scores between the older controls and the FTLD patients confirm this finding. A significant difference was evident between the older controls and the semantic dementia variant. The absence of an age-related effect in perplexity and out-of-vocabulary rate is also supported by previous studies of language production on picture description tasks in healthy aging (Glosser & Deser, 1992; Marini, Boewe, Caltagirone, & Carlomagno, 2005). These studies showed relative stability of microlinguistic abilities (e.g., word use, syntax, phonology at an individual utterance level) across the young adult (25–39 years old) and young elderly (60–74 years old) groups with significant and sharp declines present in more advanced age (> 74 years old). In our study, the mean age of the younger healthy participants group was 32.5 and the mean age of the FTLD group was 65.2. Thus, both the younger healthy and the older FTLD participants were well within the age range shown to have stable microlinguistic abilities. Prior work on language and aging did identify significant age-related differences in language processing but these differences were limited to higher levels of linguistic analysis including anaphoric reference, propositional content and discourse structure (Marini, et al., 2005; North, Ulatowska, Macaluso-Haynes, & Bell, 1986; Ulatowska, Hayashi, Cannito, & Fleming, 1986). Since language patterns involved in the computation of the perplexity index are contained to 1–2 consecutive words, the perplexity technique can be said to capture local or microlinguistic rather than macrolinguistic features that are not significantly affected by age within the microlinguistically stable range. In keeping with these prior findings on language in aging, we also did not find a significant difference either in perplexity scores or in the out-of-vocabulary rates between the younger and the older control groups. The insensitivity of our approach to age differences within the age range covered in this article suggests that the perplexity index may generalize to other acute and progressive disorders affecting language that are more prevalent in younger individuals.

The distribution of the mean perplexity scores across the FTLD variants is consistent with the phenomenology of the disease. Our study suggests that the perplexity of a language model trained on the speech of healthy adults is sensitive to semantic deficits in FTLD that manifest themselves through syntactically intact but statistically "unexpected/perplexing" sequences of words. These findings are also in keeping with previous studies in which

patients with semantic dementia were found to be significantly more impaired on a picture naming test as compared to the progressive non-fluent aphasia and behavioral variants (Libon, et al., 2009; Nestor, et al., 2003). Patients with progressive non-fluent aphasia also produced more errors on the Boston Naming test than healthy controls; however, these errors were predominantly phonological in nature suggesting intact semantic store in this group (Nestor, et al., 2003).

Previous work on progressive logopenic aphasia demonstrated that the speech produced by patients with this syndrome is characterized by slowed speaking rate, anomia and presence of phonological paraphasias while having preserved grammaticality (Amici, et al., 2006; Gorno-Tempini, et al., 2008; Josephs, et al., 2008). These symptoms of progressive logopenic aphasia are not easily distinguishable from the symptoms of semantic dementia (Westbury & Bub, 1997) that may also manifest through anomia (Hodges, et al., 1992) and with relatively preserved grammaticality (Amici, et al., 2006; Mesulam, et al., 2009). The problem in distinguishing between these syndromes has been highlighted by Bird who demonstrated that even though patients with early stages of semantic dementia exhibit word-finding difficulties on picture naming and category fluency tests, these deficits are not obvious on a picture description task (Bird, et al., 2000). The latter effect was attributed by Bird to the patients' compensating for their inability to refer to objects in the Cookie Theft stimulus with other generally acceptable vocabulary. Our methodology may help in distinguishing between these two variants as the out-of-vocabulary rate "compares" the vocabulary used by healthy adults to the vocabulary used by the FTLN patients. Patients with semantic dementia may be using generally acceptable vocabulary to describe the picture but this vocabulary differs from what would be typically expected from a healthy person on this task. Both the out-of-vocabulary rate and the perplexity index help capture this discrepancy.

The results of our study also corroborated Bird's findings with respect to the 6 narrative models simulating semantic memory impairment. The perplexity index computed on these artificial narratives increased in direct proportion to the increasing degree of semantic memory impairment simulated with Bird's Narrative Models. The perplexity indices computed on Bird's Narrative Models 1 (57.11) and 2 (55.24), representing speech of healthy controls and people with minimal semantic memory impairment, were very similar to the perplexity index calculated on the speech of healthy and behavioral variant frontotemporal dementia participants in our study (48.7 and 57.5, respectively). The perplexity index calculated on the narratives of the semantic dementia group in our study was 111, which is similar to the perplexity index calculated on Bird's Narrative Model 5. This Narrative Model was constructed to represent more severe semantic memory impairment. This is consistent with our data showing that 7 out of 11 (64%) semantic dementia patients in our study had a language-related clinical dementia rating score greater or equal to 2 (moderate-to-severe impairment). Only one semantic dementia patient had a language-related clinical dementia rating of 0.5 (mild impairment). These results provide further evidence in favor of the hypothesis that the perplexity index is sensitive to manifestations of semantic dementia in spontaneous speech and may be used as an indicator of the severity of semantic memory impairment.

The subjects with progressive non-fluent aphasia variant had a nominally higher perplexity than the subjects with either the progressive logopenic aphasia or behavioral variants, or the healthy subjects. Both progressive non-fluent aphasia and semantic dementia are distinct subtypes of the general diagnosis of primary progressive aphasia; however, the characteristic features of progressive non-fluent aphasia that distinguish it from the semantic dementia variant include phonological problems (e.g., phonemic paraphasias) and agrammatism, whereas semantic processing remains relatively intact (Grossman & Ash, 2004). Both

phonemic paraphasias and agrammatism are likely to negatively affect the perplexity scores as phonemic paraphasias results in out-of-vocabulary words (or non-words), whereas agrammatism results in word sequences that one does not expect to find in normal conversational speech.

An unexpected finding was that the perplexity index showed a difference between the semantic dementia and the behavioral variant groups but not between semantic dementia and the logopenic aphasia groups. This was unexpected because the out-of-vocabulary rate measure was correlated with the perplexity measure ($r = 0.52$, $p\text{-value} < 0.001$) and did show a significant difference between the logopenic aphasia group and the semantic dementia group. This divergence in measurements on the logopenic aphasia group was likely due to the presence of a single subject in this group with a perplexity score of 329.2 which is more than 2 standard deviations over the mean (mean = 105.9 (stdev = 110.7)). Removing this subject from the PLA group reduces the mean to 61.3 (stdev = 19.6). However, the difference in means between the reduced PLA group and the SD group is still not significant (after adjustment for multiple comparisons) but it does follow the same pattern as the out-of-vocabulary rate measure and indicates that a larger sample size may reveal significant differences.

4.2 Comparison between the perplexity index and neuropsychological test results

The pattern of neuropsychological test results was consistent with what would be expected based on the diagnostic formulations of FTLN variants and the severity of impairment. For example, patients with progressive logopenic aphasia and semantic dementia performed worse than the other variants on naming, similarities and category fluency tasks. The progressive non-fluent aphasia patients had worse scores on letter fluency compared to other groups, whereas the behavioral variant patients performed better on free recall, category fluency and verbal similarities tests. These patterns are also consistent with previous work in FTLN populations (Amici, et al., 2006; Rohrer, et al., 2009). A comparison of the neuropsychological test results to the perplexity index showed that the category fluency test had a statistically significant negative correlation with the perplexity index whereas the letter fluency test did not. These results are in keeping with prior work showing decreased performance on category fluency tests by patients with semantic dementia (Clark, Charuvastra, Miller, Shapira, & Mendez, 2005; Monsch, et al., 1992). These findings are also consistent with studies showing an accelerated deterioration of semantic features of concepts in Alzheimer's disease, whereas structural information such as syntax and grammar remain relatively intact (Kempler, Curtiss, & Jackson, 1987) albeit with lower complexity (Garrard, Maloney, Hodges, & Patterson, 2005; Harper, 2000; Roark, Mitchell, et al., 2007; Williams, Holmes, Kemper, & Marquis, 2003). Thus, the fact that language model perplexity correlates with category fluency measures associated with semantic impairment provides additional support for the main findings of our study. Specifically, the deterioration of semantic features of concepts in semantic dementia leads to using words that are not semantically coherent with other words in the same utterance resulting in unexpected word sequences.

We also found that the BDAE model perplexity scores were correlated with CVLT Free and Delayed Recall, Boston Naming test and WAIS-R Verbal Similarities tests. CVLT Free and Delayed Recall tests have been previously shown to elicit memory problems in Alzheimer's patients (Bayley, et al., 2000). Lexical retrieval and semantic deficits elicited with the Boston Naming test and WAIS-R Verbal Similarities test have also been shown to be sensitive to the effects of Alzheimer's disease (Hart, Kwentus, Taylor, & Hamer, 1988; Laine, Vuorinen, & Rinne, 1997). Alternatively, these findings are consistent with the severity analyses supporting the notion that perplexity scores will increase as general neuropsychological integrity decreases with disease progression in all forms of FTLN.

The fact that we found 17 out of 48 subjects to have a clinical dementia rating of 2 or greater on the language-specific CDR scale, whereas there were at most 7 subjects with this level of severity on other dimensions, indicates that language is more severely affected than other functional domains in our sample of patients with FTLN. This finding is important as it suggests that language assessment may be a primary outcome measure in studies of new therapies for FTLN.

There is increasing recognition that the different subtypes of progressive aphasia including progressive non-fluent aphasia, semantic dementia and progressive logopenic aphasia have different anatomic and biochemical bases (Mesulam, 2003; Rohrer, et al., 2009; Westbury & Bub, 1997). Proper identification of the expressive speech disorder plays an important role in differential diagnosis as well as the assessment of daily functioning (Mesulam, et al., 2009). Although there are no effective treatments for the different subtypes at this time, the prospects are quite favorable for the emergence of specific treatments for the tauopathies that are associated with progressive non-fluent aphasia and the TDP-43 proteinopathy associated with semantic dementia (Josephs, et al., 2008). Although the measures of language functioning cannot replace the current clinical assessment for dementia, they do offer a standardized and objective way of characterizing expressive speech and could serve as a means of classifying and monitoring the functioning of subjects in a clinical trial, either by supporting or calling into question a clinical diagnosis.

5. Limitations

A number of limitations must be discussed in order to facilitate the interpretation of the study results. First, stimuli that elicit greater amounts of speech than the Cookie Theft stimulus may achieve better test-retest reliability than our current approach. However, the Cookie Theft is a standard stimulus used in the clinical diagnosis of aphasia. In our study, the mean duration of a picture description by FTLN patients was 99.6 seconds and the mean number of words was 108. Although we may not be able to detect differences of less than 10% with a single stimulus of this size (Brookshire & Nicholas, 1994), the differences in perplexity and out-of-vocabulary rate between the semantic dementia variant and the behavioral variant as well as controls are much greater than 10%. Thus we believe that the Cookie Theft stimulus was sufficient for the current study, while recognizing that greater power would be achieved with larger and/or multiple samples. Second, the older controls consisted of nursing home residents that did not have a diagnosis of dementia but did have other diagnoses including depression. Depression may influence one's speech production; however, patients with FTLN also tend to suffer from depression (Mesulam, 2003), thus possibly making nursing home residents (without dementia) a better control population than community dwelling elderly. Third, the current analysis is based on English-only speech samples limiting the generalizability of our findings to FTLN patients that speak other languages. The measures of perplexity index and out-of-vocabulary rate may be adapted to capture word distribution patterns in other languages; however, further validation will be required.

6. Conclusion

Measures of language model perplexity and out-of-vocabulary rate obtained from models trained on healthy adults' picture description narratives is sensitive to language impairments characteristic of frontotemporal lobar degeneration, particularly the semantic dementia variant of the disease. Our multidisciplinary approach demonstrates the utility of information technology to measure and categorize language impairments associated with frontotemporal lobar degeneration in an objective and reproducible manner. This approach may be particularly useful for a quantitative characterization of language impairment in a

clinical trial or observational study settings and may also be applicable to other neurodegenerative diseases.

Acknowledgments

The work presented in this paper was supported by the United States National Institute of Aging grants: R01-AG023195, P50-AG 16574 (Mayo Alzheimer's Disease Research Center), P30-AG19610 (Arizona ADC) and a Grant in Aid of Research from the University of Minnesota. We would also like to thank Dustin Chacon for helping with transcription of speech samples.

References

- Amici S, Gorno-Tempini ML, Ogar JM, Dronkers NF, Miller BL. An overview on Primary Progressive Aphasia and its variants. *Behav Neurol* 2006;17(2):77–87. [PubMed: 16873918]
- Ash S, Moore P, Antani S, McCawley G, Work M, Grossman M. Trying to tell a tale: discourse impairments in progressive aphasia and frontotemporal dementia. *Neurology* 2006;66(9):1405–1413. [PubMed: 16682675]
- Ash S, Moore P, Vesely L, Giunawardena D, McMillan C, Anderson C, et al. Non-fluent speech in frontotemporal lobar degeneration. *Journal of Neurolinguistics* 2008;22(4):370–383.
- Bahl, L.; Baker, J.; Jelinek, E.; Mercer, R. Perplexity--a measure of the difficulty of speech recognition tasks. Paper presented at the 94th Meeting of the Acoustical Society of America; 1977.
- Bayley PJ, Salmon DP, Bondi MW, Bui BK, Olichney J, Delis DC, et al. Comparison of the serial position effect in very mild Alzheimer's disease, mild Alzheimer's disease, and amnesia associated with electroconvulsive therapy. *J Int Neuropsychol Soc* 2000;6(3):290–298. [PubMed: 10824501]
- Benton, AL.; Hamsher, K.; Sivan, AB. Multilingual Aphasia Examination. 3. Iowa City, IA: AJA Associates; 1983.
- Berndt, RS.; Waylannd, S.; Rochon, E.; Saffran, EM.; Schwartz, MF. Quantitative production analysis: A training manual for the analysis of aphasic sentence production. Hove, UK: Psychology Press; 2000.
- Bird H, Lambon Ralph MA, Patterson K, Hodges JR. The rise and fall of frequency and imageability: noun and verb production in semantic dementia. *Brain Lang* 2000;73(1):17–49. [PubMed: 10872636]
- Brookshire RH, Nicholas LE. Speech sample size and test-retest stability of connected speech measures for adults with aphasia. *Journal of Speech and Hearing Research* 1994;37(2):399–407. [PubMed: 8028321]
- Brown P, Della Pietra V, Mercer R, Della Pietra S, Lai J. An estimate of an upper bound for the entropy of English. *Computational Linguistics* 1992;18(1):31–40.
- Clark DG, Charuvastra A, Miller BL, Shapira JS, Mendez MF. Fluent versus nonfluent primary progressive aphasia: a comparison of clinical and functional neuroimaging features. *Brain Lang* 2005;94(1):54–60. [PubMed: 15896383]
- Delis, DC.; Kramer, JH.; Kaplan, E.; Ober, BA. California Verbal Learning Test. 2. San Antonio, TX: The Psychological Corp; 2000.
- Garrard P, Maloney L, Hodges J, Patterson K. The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. *Brain* 2005;128:250–260. [PubMed: 15574466]
- Glosser G, Deser T. A comparison of changes in macrolinguistic and microlinguistic aspects of discourse production in normal aging. *J Gerontol* 1992;47(4):P266–272. [PubMed: 1624704]
- Goodglass, H.; Kaplan, E. The Assessment of Aphasia and Related Disorders. Philadelphia: Lea and Febiger; 1983.
- Gordon J. A quantitative production analysis of picture description. *Aphasiology* 2006;20(2/3/4):188–204.
- Gordon J. Measuring the lexical semantics of picture description in aphasia. *Aphasiology* 2008;22(7–8):839–852.

- Gorno-Tempini ML, Brambati SM, Ginex V, Ogar J, Dronkers NF, Marcone A, et al. The logopenic/phonological variant of primary progressive aphasia. *Neurology* 2008;71(16):1227–1234. [PubMed: 18633132]
- Gorno-Tempini ML, Dronkers NF, Rankin KP, Ogar JM, Phengrasamy L, Rosen HJ, et al. Cognition and anatomy in three variants of primary progressive aphasia. *Ann Neurol* 2004;55(3):335–346. [PubMed: 14991811]
- Grossman M. Frontotemporal dementia: a review. *J Int Neuropsychol Soc* 2002;8(4):566–583. [PubMed: 12030310]
- Grossman M, Ash S. Primary progressive aphasia: a review. *Neurocase* 2004;10(1):3–18. [PubMed: 15849155]
- Harper, L. Sentence Production in Individuals with Alzheimer's Disease. University of Calgary; Calgary: 2000.
- Hart RP, Kwentus JA, Taylor JR, Hamer RM. Productive naming and memory in depression and Alzheimer's type dementia. *Arch Clin Neuropsychol* 1988;3(4):313–322. [PubMed: 14591421]
- Hodges J, Davies R, Xuereb J, Casey B, Broe M, Bak T, et al. Clinicopathological Correlates in Frontotemporal Dementia. *Annals of Neurology* 2004;56:399–406. [PubMed: 15349867]
- Hodges J, Patterson K, Oxbury S, Funnell E. Semantic dementia. Progressive fluent aphasia with temporal lobe atrophy. *Brain* 1992;115(Pt 6):1783–1806. [PubMed: 1486461]
- Josephs KA, Duffy JR, Strand EA, Whitwell JL, Layton KF, Parisi JE, et al. Clinicopathological and imaging correlates of progressive aphasia and apraxia of speech. *Brain* 2006;129(Pt 6):1385–1398. [PubMed: 16613895]
- Josephs KA, Whitwell JL, Duffy JR, Vanvoorst WA, Strand EA, Hu WT, et al. Progressive aphasia secondary to Alzheimer disease vs FTLTD pathology. *Neurology* 2008;70(1):25–34. [PubMed: 18166704]
- Kaplan, E.; Goodglass, H.; Weintraub, S. The Boston Naming Test. Boston: Lea & Fabiger; 1978.
- Keller, F. The Entropy Rate Principle as a Predictor of Processing Effort: An Evaluation against Eye-tracking Data. Paper presented at the Conference on Empirical Methods in Natural Language Processing; Barcelona, Spain. 2004.
- Kempler D, Curtiss S, Jackson C. Syntactic Preservation in Alzheimer's Disease. *Journal of Speech and Hearing Research* 1987;30:343–350. [PubMed: 3669641]
- Kertesz A, McMonagle P, Blair M, Davidson W, Munoz DG. The evolution and pathology of frontotemporal dementia. *Brain* 2005;128(9):1996–2005. [PubMed: 16033782]
- Knopman DS, Boeve BS, Caselli RJ, Graff-Radford NR, Kramer JH, Mendez MF, et al. Longitudinal tracking of FTLTD: toward developing clinical trial methodology. *Alzheimer Dis Assoc Disord* 2007;21(4):S58–63. [PubMed: 18090426]
- Knopman DS, Kramer JH, Boeve BF, Caselli RJ, Graff-Radford NR, Mendez MF, et al. Development of methodology for conducting clinical trials in frontotemporal lobar degeneration. *Brain* 2008;131(11):2957–2968. [PubMed: 18829698]
- Kramer JH, Jurik J, Sha SJ, Rankin KP, Rosen HJ, Johnson JK, et al. Distinctive neuropsychological patterns in frontotemporal dementia, semantic dementia, and Alzheimer disease. *Cogn Behav Neurol* 2003;16(4):211–218. [PubMed: 14665820]
- Laine M, Vuorinen E, Rinne JO. Picture naming deficits in vascular dementia and Alzheimer's disease. *J Clin Exp Neuropsychol* 1997;19(1):126–140. [PubMed: 9071647]
- Libon DJ, Xie SX, Wang X, Massimo L, Moore P, Vesely L, et al. Neuropsychological decline in frontotemporal lobar degeneration: a longitudinal analysis. *Neuropsychology* 2009;23(3):337–346. [PubMed: 19413447]
- Manning, C.; Shutze, H. Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press; 1999.
- Marini A, Boewe A, Caltagirone C, Carlomagno S. Age-related differences in the production of textual descriptions. *J Psycholinguist Res* 2005;34(5):439–463. [PubMed: 16177935]
- Mesulam M. Primary progressive aphasia--a language-based dementia. *N Engl J Med* 2003;349(16):1535–1542. [PubMed: 14561797]

- Mesulam M, Rogalski E, Wieneke C, Cobia D, Rademaker A, Thompson C, et al. Neurology of anomia in the semantic variant of primary progressive aphasia. *Brain* 2009;132(9):2553–2565. [PubMed: 19506067]
- Mohs RC, Knopman D, Petersen RC, Ferris SH, Ernesto C, Grundman M, et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord* 1997;11(Suppl 2):S13–21. [PubMed: 9236948]
- Monsch AU, Bondi MW, Butters N, Salmon DP, Katzman R, Thal LJ. Comparisons of verbal fluency tasks in the detection of dementia of the Alzheimer type. *Arch Neurol* 1992;49(12):1253–1258. [PubMed: 1449404]
- Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 1993;43(11):2412–2414. [PubMed: 8232972]
- Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* 1998;51(6):1546–1554. [PubMed: 9855500]
- Nestor PJ, Graham NL, Fryer TD, Williams GB, Patterson K, Hodges JR. Progressive non-fluent aphasia is associated with hypometabolism centred on the left anterior insula. *Brain* 2003;126(Pt 11):2406–2418. [PubMed: 12902311]
- Nicholas LE, Brookshire RH. A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *J Speech Hear Res* 1993;36(2):338–350. [PubMed: 8487525]
- North AJ, Ulatowska HK, Macaluso-Haynes S, Bell H. Discourse performance in older adults. *Int J Aging Hum Dev* 1986;23(4):267–283. [PubMed: 3557641]
- Peelle J, Cooke A, Moore P, Vesely L, Grossman M. Syntactic and thematic components of sentence processing in progressive nonfluent aphasia and nonaphasic frontotemporal dementia. *Journal of Neurolinguistics* 2007;20(6):482–494. [PubMed: 18978927]
- Rascovsky K, Hodges JR, Kipps CM, Johnson JK, Seeley WW, Mendez MF, et al. Diagnostic criteria for the behavioral variant of frontotemporal dementia (bvFTD): current limitations and future directions. *Alzheimer's Disease and Associated Disorders* 2007;21(4):S14–18.
- Roark, B.; Hosom, J.; Mitchell, M.; Kaye, JA. Automatically derived spoken language markers for detecting Mild Cognitive Impairment. Paper presented at the International Conference on Technology and Aging (ICTA); 2007.
- Roark, B.; Mitchell, M.; Hollingshead, K. Syntactic complexity measures for detecting Mild Cognitive Impairment. Paper presented at the ACL 2007 Workshop on Biomedical Natural Language Processing (BioNLP); Prague, Czech Republic. 2007.
- Rohrer JD, Ridgway GR, Crutch SJ, Hailstone J, Goll JC, Clarkson MJ, et al. Progressive logopenic/phonological aphasia: Erosion of the language network. *Neuroimage* 2009;49(1):984–983. [PubMed: 19679189]
- Saffran EM, Berndt RS, Schwartz MF. The quantitative analysis of agrammatic production: procedure and data. *Brain Lang* 1989;37(3):440–479. [PubMed: 2804622]
- Snowden JS. Semantic dysfunction in frontotemporal lobar degeneration. *Dement Geriatr Cogn Disord* 1999;10(Suppl 1):33–36. [PubMed: 10436337]
- Stroop JR. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 1935;18:643–662.
- Thompson CK, Ballard KJ, Tait ME, Weintraub S, Mesulam M. Patterns of language decline in non-fluent primary progressive aphasia. *Aphasiology* 1997;11(4–5):297–321.
- Ulatowska HK, Hayashi MM, Cannito MP, Fleming SG. Disruption of reference in aging. *Brain Lang* 1986;28(1):24–41. [PubMed: 3719297]
- Wechsler, DA. Wechsler Adult Intelligence Scale-Revised. New York: Psychological Corporation; 1981.
- Wechsler, DA. Wechsler Memory Scale-Revised. New York: Psychological Corporation; 1987.
- Weintraub S, Rubin NP, Mesulam M. Primary progressive aphasia. Longitudinal course, neuropsychological profile, and language features. *Arch Neurol* 1990;47(12):1329–1335. [PubMed: 2252450]

- Westbury C, Bub D. Primary progressive aphasia: a review of 112 cases. *Brain Lang* 1997;60(3):381–406. [PubMed: 9398390]
- Williams K, Holmes F, Kemper S, Marquis J. Written Language Clues to Cognitive Changes of Aging: An Analysis of the Letters of King James VI/I. *Journal of Gerontology: Psychological Sciences* 2003;58B(1):42–44.
- Yorkston KM, Beukelman DR. An analysis of connected speech samples of aphasic and normal speakers. *Journal of Speech and Hearing Disorders* 1980;45(1):27–36. [PubMed: 7354627]
- Young, SJ.; Evermann, G.; Gales, MJF.; Hain, T.; Kershaw, D.; Moore, G., et al. The HTK Book, version 3.4. Cambridge, UK: Cambridge University Engineering Department; 2006.

Figure 1.
Study design and data flow.

Figure 2. Perplexity results obtained with the BDAE model for the four FTLN variants and younger and older controls.

Figure 3. Out-of-vocabulary rate results obtained with the BDAE model for the four FTL variants and younger and older controls.

Figure 4. Perplexity index scores computed based on the BDAE statistical language model for 6 narrative models representing different degrees of semantic memory

Table 1

FTLD variant group differences on standard cognitive assessments

N=48	bvFTD(n=19) mean (std.)	PNFA(n=12) mean (std.)	PLA(n=6) mean (std.)	SD(n=11) mean (std.)	p-value
Age	61.10 (8.70)	66.33 (6.90)	63.50 (9.56)	70.00 (7.88)	0.06
CVLT Free Recall	19.74 (6.81)	14.75 (10.57)	12.50 (7.47)	12.55 (6.77)	0.05
CVLT Delayed Recall	3.21 (2.89)	3.75 (2.80)	2.33 (2.42)	1.55 (2.38)	0.25
Trail Making Part A					
total time to complete	56.11 (36.46)	78.83 (38.16)	112.17 (12.00)	62.91 (33.17)	0.06
number of correct lines	12.74 (3.02)	9.25 (5.63)	10.67 (4.84)	11.36 (3.98)	0.52
number of errors	1.11 (1.82)	1.83 (1.53)	1.50 (0.55)	1.64 (2.94)	0.45
Number Cancellation					
total correct	27.74 (10.52)	24.42 (11.78)	19.00 (9.40)	24.55 (7.75)	0.90
times reminded	0.37 (0.83)	0.08 (0.29)	0.17 (0.41)	0.73 (1.19)	0.54
Digits Backward^{§¶}	3.84 (1.68)	2.25 (1.22)	2.33 (1.21)	4.00 (1.18)	0.02
Stroop Test					
color naming correct	45.11 (23.76)	34.08 (19.72)	27.33 (8.61)	43.18 (17.50)	0.29
color-word naming correct	30.63 (22.33)	18.42 (17.25)	8.83 (4.11)	18.64 (9.33)	0.05
color-word errors *	2.95 (5.17)	2.83 (4.41)	8.67 (12.07)	0.73 (1.10)	0.02
Digit-Symbol Substitution	48.79 (17.80)	39.83 (24.80)	28.67 (11.86)	45.73 (15.94)	0.06
Verbal Fluency (Ph)					
letter C	9.26 (6.10)	4.33 (2.42)	5.33 (3.88)	6.36 (4.06)	0.13
letter F	8.79 (5.14)	4.33 (3.77)	6.50 (4.76)	7.36 (3.96)	0.10
letter L	7.89 (5.00)	4.42 (2.78)	5.66 (3.98)	7.64 (4.68)	0.25
Verbal Fluency (Sem)					
Animals [§]	12.47 (4.67)	9.75 (6.64)	7.00 (3.74)	6.36 (4.39)	0.01
Fruits	7.89 (3.54)	6.17 (3.81)	6.16 (3.43)	4.45 (4.37)	0.08
Vegetables [§]	7.37 (3.66)	5.75 (3.91)	5.83 (2.13)	3.36 (4.63)	0.02
Boston Naming Test[§]	23.21 (6.76)	18.58 (10.70)	15.16 (9.76)	6.55 (5.41)	< 0.001
WAIS-R Verbal Similarities	16.00 (6.30)	11.55 (9.68)	10.00 (4.69)	9.09 (7.54)	0.08

* indicates significance on Mann-Whitney test at 0.05 level between PLA and SD

\$ indicates significance on Mann-Whitney test at 0.05 level between bvFTD and SD
 § indicates significance on Mann-Whitney test at 0.05 level between PNFA and SD
 ¶ indicates significance on Mann-Whitney test at 0.05 level between PNFA and bvFTD

Table 2

Correlations between perplexity scores obtained with the BDAE model and neuropsychological measures of cognitive functioning

N=48	<u>Spearman rank correlation coefficients BDAE model perplexity index</u>
CVLT Free Recall	-.47 **
CVLT Delayed Recall	-.32 *
Trail Making Part A	
total time to complete	.13
number of correct lines	-.10
number of errors	.01
Number Cancellation	
total correct	-.27
times reminded	.07
Digits Backward	-.16
Stroop Test	
color naming correct	-.10
color-word naming correct	-.12
color-word errors	-.06
Digit-Symbol Substitution	-.17
Verbal Fluency (Letters)	
letter C	-.17
letter F	-.10
letter L	-.09
Verbal Fluency (Categories)	
animals	-.52 **
fruits	-.38 **
vegetables	-.42 **
Boston Naming Test (N correct)	-.57 **
WAIS-R Verbal Similarities (N correct)	-.46 **

* indicates correlations significant at 0.05 level (two-tailed)

** indicates correlations significant at 0.01 level (two-tailed)

Table 3
Differences in mean perplexity scores obtained with the BDAE language model between mild and severe dementia cases (N=48)

N=48	N subjects	BDAE perplexity	F-score	p-value	Spearman correlation ^{##}
CDR Memory [#]	< 2	45	76.7	-	0.40 ^{**}
	≥ 2	3	149.6		
CDR Orientation [#]	< 2	47	79.3	-	0.44 ^{**}
	≥ 2	1	173.1		
CDR Judgment	< 2	42	76.6	3.204	0.080
	≥ 2	6	117.9		0.14
CDR Community [#]	< 2	44	76.4	4.287	0.44 [*]
	≥ 2	4	133.8		0.13
CDR Home/Hobbies	< 2	42	76.7	3.204	0.080
	≥ 2	6	117.9		0.17
CDR Self-care [#]	< 2	47	79.27	-	0.02
	≥ 2	1	173.6		
CDR Behavior	< 2	41	80.4	0.067	0.797
	≥ 2	7	86.2		0.01
CDR Language	< 2	31	62.3	14.383	<0.001 ^{***}
	≥ 2	17	117.0		0.53 ^{**}
CDR TOTAL	< 9	42	74.0	6.506	0.014 [*]
	≥ 8.5	6	131.8		0.33 [*]

* indicates correlations significant at 0.05 level (two-tailed)

** indicates correlations significant at 0.01 level (two-tailed)

fewer than 3 cases assessed as moderate or severe were found

BDAE perplexity scores were correlated with the ranks of CDR scores for each CDR domain using Spearman rank correlation