Computerized Analysis of Speech and Language to Identify Psycholinguistic Correlates of Frontotemporal Lobar Degeneration

Serguei V. S. Pakhomov, PhD[*], **Glenn E. Smith, PhD**[†], **Dustin Chacon**[*], **Yara Feliciano**[‡], **Neill Graff-Radford, MD**[§], **Richard Caselli, MD**[||], and **David S. Knopman, MD**[†]

[*]University of Minnesota, Twin Cities, MN

[†]Mayo Alzheimer's Disease Research Center

[‡]University of Puerto Rico Medical School

[§]Department of Neurology, Mayo Clinic Jacksonville

[||]Department of Neurology Mayo Clinic Arizona

Abstract

Objective—To evaluate the use of a semiautomated computerized system for measuring speech and language characteristics in patients with frontotemporal lobar degeneration (FTLD).

Background—FTLD is a heterogeneous disorder comprising at least 3 variants. Computerized assessment of spontaneous verbal descriptions by patients with FTLD offers a detailed and reproducible view of the underlying cognitive deficits.

Methods—Audiorecorded speech samples of 38 patients from 3 participating medical centers were elicited using the Cookie Theft stimulus. Each patient underwent a battery of neuropsychologic tests. The audio was analyzed by the computerized system to measure 15 speech and language variables. Analysis of variance was used to identify characteristics with significant differences in means between FTLD variants. Factor analysis was used to examine the implicit relations between subsets of the variables.

Results—Semiautomated measurements of pause-to-word ratio and pronoun-to-noun ratio were able to discriminate between some of the FTLD variants. Principal component analysis of all 14 variables suggested 4 subjectively defined components (length, hesitancy, empty content, grammaticality) corresponding to the phenomenology of FTLD variants.

Conclusion—Semiautomated language and speech analysis is a promising novel approach to neuropsychologic assessment that offers a valuable contribution to the toolbox of researchers in dementia and other neurodegenerative disorders.

Keywords

spontaneous speech; language; prosody; frontotemporal lobar degeneration; automated speech analysis

The need for improved definition of syndromes and phenotypes is a key theme in dementia research in general, and frontotemporal lobar degeneration (FTLD) in particular.[1] Currently, FTLD comprises 3 syndromes: behavioral variant frontotemporal dementia (bvFTD), progressive nonfluent aphasia (PNFA), and semantic dementia (SD). The inclusion of progressive logopenic aphasia (PLA) as either a variant of FTLD or Alzheimer is currently being debated.[2–8] These syndromes are diagnosed using standard clinical criteria, neuropsychologic testing, and neuroimaging. Careful clinical evaluation is critical to FTLD diagnosis, particularly in the early stages of disease progression. A systematic analysis of spontaneous speech samples is considered "the single most valuable aspect of the diagnosis"[9] for the aphasic FTLD syndromes. Research in aphasia has contributed a set of instruments in the form of picture description tasks designed to elicit and rate spontaneous speech, including the Boston Diagnostic Aphasia Examination Cookie Theft stimulus.[10] The assessments with these instruments, however, are traditionally carried out manually, which is subjective and may not have the detail and precision of measurements necessary to define either the full range of syndromes, or their nuances. Detailed measurements are particularly important to the assessment of syntax, semantics, and prosody—the 3 areas identified in a survey of clinicians' views on the clinical usefulness of aphasia test batteries.[11]

Recent advances in computerized natural language processing (NLP) and automatic speech recognition (ASR)[12,13] make it possible to develop objective and precise instruments for automated or semiautomated analysis of speech and language patterns present in the spontaneous speech elicited with standard stimuli. In addition to clinical diagnostic and treatment purposes, precise and fine-grained measurements of prosodic and linguistic characteristics of spontaneous speech are necessary to enable grouping of these characteristics across populations of patients to define the linguistic phenotypes associated with FTLD and other neurodegenerative disorders affecting language. Although FTLD is a degenerative and currently untreatable disease, availability of objectively defined characteristics with acceptable variability within and across subjects is critical to designing clinical trials to test therapeutic interventions that are under development. In this study, we use a semiautomated system for language and speech analysis for objective measurement of speech and language characteristics elicited from patients with FTLD on a standard picture description task.

SPEECH AND LANGUAGE CHARACTERISTICS IN FRONTOTEMPORAL DEMENTIA

Over half of all patients with symptoms of FTLD exhibit language-related manifestations on initial presentation. [14] A number of speech and language characteristics were shown to be differentially sensitive to the effects of FTLD variants. The PNFA variant has been characterized in terms of dysfluent, effortful, and agrammatical speech.[3,15–19] The SD variant involves multimodal non-verbal and verbal naming and recognition deficits with relatively preserved grammar.[20,21] However, despite these differences between the nonfluent and fluent aphasic variants of FTLD, there is considerable overlap between their language specific manifestations.[22]

Apart from the overlap between fluent and non-fluent types of primary progressive aphasia, the distinction between the fluent subtype of aphasia and SD is also being debated. For example, Josephs et al[23] treat not otherwise specified primary progressive aphasia (PPA-NOS) as separate from either SD or PNFA variants of FTLD. However, a recent study by Adlam et al[24] suggests that the distinction between these 2 classifications is a matter of emphasis rater that differences in the underlying pathophysiology of the phenomenon.

The PLA variant of FTLD has been introduced into the diagnosis of FTLD relatively recently. Whereas bvFTD, PNFA, and SD syndromes are likely to represent FTLD pathologically,[7] the grouping of the PLA syndrome with FTLD or Alzheimer disease is debatable. Similarly to PNFA, spontaneous speech production in PLA has also been characterized by slower speaking rate, hesitations, and pauses attributable to word-finding difficulties.[2] Some of the cases of primary progressive aphasia distinct from SD and PNFA in the Josephs et al study also exhibited these altered prosodic characteristics of speech with relatively preserved grammar and could possibly be classified as PLA.[23]

Language-related deficits in patients diagnosed with bvFTD tend to be observed at the higher discourse level rather than syntax, phonology, and semantics found with other FTLD variants. Studies of patients with bvFTD (or social/dysexecutive variant) showed impaired working memory that manifests itself through deficits in sentence comprehension,[25] thematic role processing in verbs,[18,26] and altered discourse characteristics.[27] The latter, including discourse coherence, cohesion, and "empty" speech (overuse of pronouns) are also sensitive to diffuse neural degradation characteristic of Alzheimer disease on tasks involving elicitation of spontaneous speech.[28–30]

Several diverse speech and language features have been identified and used to characterize fluent PPA and SD in general, and the SD variant of FTLD in particular. Gordon[31] used a Quantitative Production Analysis (QPA) protocol[32,33] to compare fluent and nonfluent aphasic speech productions elicited with a picture description task. The measures used in QPA protocol were found to be sensitive to the severity of both fluent and nonfluent aphasia but could not reliably discriminate between these 2 subtypes. In a subsequent study, Gordon[34] tested additional measures of correct information units[35,36] and type-to-token ratio. Although these measures correlated with those obtained with the QPA protocol and were sensitive to aphasia severity, they also failed to distinguish between fluent and nonfluent groups.

In summary, language-specific manifestations of FTLD (and progressive aphasias in general) are diverse with significant overlap across different variants and are currently assessed using 2 main types of approaches. One approach consists of subjective assessment conducted by trained neuropsychologists or speech-language pathologists that use Likert-style scales to judge the performance of the subject/patient on a set of language dimensions (eg, speaking rate, hesitancy, speech sound distortions, telegraphic speech, grammaticality). The other approach consists of manual psycholinguistic analysis of speech and language samples in terms of their phonologic, syntactic, semantic, and pragmatic features. This latter approach attempts to identify linguistic features sensitive to manifestations of the disease and then quantify the occurrence of these features in speech and language samples. The first approach is more suitable in a clinical setting as it is easier and less time consuming to conduct, whereas the second approach is likely to produce more objective and reproducible results and, therefore, is more suitable in a research setting. However, the second approach still relies on manual identification of linguistic features including utterance and clause boundaries, verb argument structure, thematic roles of verb arguments, and parts-of-speech.

Manual annotation of linguistic features in spontaneous speech samples is a very labor-intensive process and is subject to variable agreement among the annotators, particularly in content analysis involving multiple semantic categories.[37–39] Although manual linguistic analysis is an indispensable exploratory tool, validated computerized linguistic analysis has the potential to minimize the variability in detecting and measuring linguistic phenomena, thus, offers a better reproducibility and comparability of measurements. High temporal resolution of computerized analysis of the kind we propose in this article also has the potential advantage over manual methods limited by human perceptual abilities.

Neuropsychologic Tests and FTLD Assessment

In addition to the analysis of spontaneous speech and language, research in FTLD has also relied on standard neuropsychologic test batteries, including visual confrontation naming, trail-making, verbal fluency, digits backward, number cancellation, verbal memory and learning, and Stroop test to differentiate between FTLD variants among others.[3,7,14,40] However, one of the issues with the use of standardized neuropsychologic tests in FTLD research has to do with the lack of consistency in test selection and scoring across different clinical sites.[41] These and other studies indicate that standard neuropsychologic tests are certainly capable of distinguishing between FTLD variants and between FTLD and AD. However, some studies found that the group differences identified with standard neuropsychologic tests did not occur consistently across different tests. For example, Thompson et al[22] suggested that neuropsychologic tests may be too narrow to capture the richness of the behavioral and cognitive features of FTLD and even obscure the differences between patient groups. Qualitative assessments of error types were proposed to complement neuropsychologic test scores in obtaining a more complete characterization of FTLD variants. Computerized approaches to speech and language analysis based on computational linguistics and NLP technology that we explore in the current study may help quantify some of these qualitative measures. For example, some of the phonologic errors, perseverative behavior, poor discourse organization or confabulation on verbal fluency, Boston Naming and other tests may be correlated with quantifiable measures obtained through the alignment of audio with transcripts of speech samples and measures based on statistical language modeling discussed in the Methods section of this paper.

METHODS

Participants

Thirty-eight patients diagnosed with 1 of the 3 FTLD syndromes (bvFTD, PNFA, SD) and PLA were recruited from 3 academic medical centers. All aspects of this study were approved by the IRBs at each of the medical centers and the University of Minnesota. All 38 participants underwent a neuropsychologic test battery that included the Boston Diagnostic Aphasia Examination Cookie-Theft Picture Description Task.[10] This assessment was part of a larger study and minimizing subject burden was a key concern. The details of administering the neuropsychologic test battery and the spontaneous speech elicitation procedures were earlier reported.[7]

Diagnostic Criteria

Diagnostic and exclusion criteria for this study were reported earlier.[7] In brief, we defined 4 syndromes: bvFTD, PNFA, PLA, and SD. The inclusion in this study was based on the Neary criteria.[21] In addition, all patients were required to have imaging studies showing focal cerebral atrophy of at least 1 of these: the anterior temporal lobes, frontal lobes, insula, or caudate nuclei.

PNFA was diagnosed with expressive speech characterized by at least 3 of these: nonfluency (reduced numbers of words per utterance), speech hesitancy or labored speech, word finding difficulty, or agrammatism, in which these symptoms constitute the principal deficits and the initial presentation.

PLA was diagnosed with fluent aphasia with anomia but intact word meaning and object recognition in which these symptoms constitute the principal deficits and the initial presentation.

SD was diagnosed with loss of comprehension of word meaning, object identity, or face identity in which these symptoms constitute the principal deficits and the initial presentation.

Behavioral variant FTD 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 had at least 5 core symptoms in the domains of aberrant personal conduct and impaired interpersonal relationships.

Cognitive Instruments

Neuropsychologic Assessments—All participants were administered a standard neuropsychologic test battery consisting of these tests: California Verbal Learning Test Free Recall,[42] Number Cancellation,[43] Digits Backward from Wechsler Memory Scale-Revised,[44] Stroop Test,[45] Digit-Symbol Substitution,[46] Verbal Fluency for Letters and Categories,[47] Boston Naming Test.[48] All tests were scored by board-certified behavioral neurologists. The motivation for selecting these tests and detailed information on their performance in the FTLD population can be found in an earlier publication.[7]

Computerized Psycholinguistic Assessments—NLP and ASR comprise a set of computational techniques used for computerized analysis of speech and language. The application of ASR and NLP to psychometric testing is new; however, it is a natural extension of the capabilities afforded by this technology. We have developed a system for semiautomated language and speech analysis based on NLP and ASR technology (illustrated in Fig. 1). For this study, the system was configured to process audio recordings of speech elicited during the Cookie-Theft picture description task of the Boston Diagnostic Aphasia Examination. The audio input represented as digitized speech waveform was first manually transcribed verbatim and then automatically aligned with the transcribed text. The details of using ASR for automatic alignment including acoustic and language modeling are provided in the on-line Appendix A.

The resulting subsecond level alignment (illustrated in Figure 2) enables precise measurement and quantification of durational and frequency characteristics of the input at the level of utterances (2 or more coarticulated words), words, and individual phonemes. Although our current approach to treating mispronunciations does not account for many phonemic distinctive features and may not identify phoneme boundaries precisely in cases of dysarthric output, it can still be used to identify word or, in the case of unintelligible speech, utterance boundaries. On the basis of the alignments between the audio, verbatim transcriptions, and part-of-speech, we defined the following the variables:

1. Pause-to-word ratio
2. Fundamental frequency variance
3. Part-of-speech perplexity
4. Word-level perplexity
5. Pronoun-to-noun ratio
6. Word count
7. Total duration of speech in the sample
8. Mean prosodic phrase length
9. Correct Information Unit count
10. Normalized long pause count (silent pauses >400 ms in duration)

11. Normalized filled pause count

12. Normalized silent pause count (silent pauses >150 ms in duration)

13. Normalized false start count

14. Normalized repetition count

15. Normalized dysfluent even count (filled pauses, false starts, and repetitions)

The definitions of these variables and technical details related to their computation are provided in the on-line Appendix B.

Statistical Methods

Factor analysis based on the principal components analysis with Varimax rotation was used to examine the relationships between various semiautomated psycholinguistic measurements and to reduce the number of variables. One-way analysis of variance (ANOVA) was used to evaluate the differences in measurements using different FTLD variants as factors with Tukey post hoc tests for differences in means adjusted for multiple comparisons. Paired *t*-test was used to evaluate the differences between the timings of the word boundaries identified with semiautomated versus detailed speech and text alignments. All statistical calculations were carried out with SPSS 13.0 statistical software package.

Precision of Semiautomated Alignment and Part-of-speech Tagging

Measurements of speech and language characteristics used by our semiautomated approach rely on the alignment between the transcribed text of the picture description and the audio signal depicted in Figure 2. The precision of semiautomated alignment was estimated in terms of word beginning (WBBS) and word ending boundary shifts (WEBS)[49] as compared with word boundaries determined by aligning the transcripts with the audio manually. This detailed manual alignment is a painstaking process and takes significantly more time, effort, and training to carry out than the regular verbatim transcription. Thus, it was carried out on a subset of 19 randomly selected cases. In addition to the word boundary shifts, manual alignment also included part-of-speech annotation enabling evaluation of the accuracy of the automatic part-of-speech tagging used by our system. Detailed manual alignment was done subsequent to the verbatim transcription that was carried out for the semiautomated alignment, and thus was used to estimate the accuracy of the verbatim transcriptions. The verbatim transcriptions, the detailed manual alignments, and part of speech annotations were carried out by a linguist specifically trained for this task (DC).

RESULTS

Reduction of Variables to Principal Components

On the basis of the results of the exploratory factor analysis done on all 38 samples and the semiautomatic psycholinguistic variables, we identified 4 components that cumulatively accounted for 71% of the total variance in all variables. On the basis of the values of the coefficients greater than 0.6 in the component rotation matrix (Table 1), we subjectively determined that the components represent speech length (component 1), hesitancy (component 2), empty content (component 3), and grammaticality (component 4).

Discriminating Between FTLD Variants

In this section, we present the results of a one-way ANOVA with the 4 FTLD variants (bvFTD, PNFA, PLA, and SD) used as factors for several neuropsychologic, and semiautomatic psycholinguistic variables.

Neuropsychologic Assessments

Of all tests evaluated in this study, only 2 showed statistically significant differences in means: verbal category fluency (*P*<0.001) and correctness on the Boston Naming Test (*P*<0.001). The mean verbal fluency score for the bvFTD group was 12.87, for PNFA—11.90, for PLA—7.75, and for SD—5.33. The mean correctness score on the Boston Naming Test was 23.2 for bvFTD, 23.40 for PNFA, 16.75 for PLA, and 5.00 for SD. Tukey post hoc test for pairwise differences in mean category fluency scores showed significant differences between bvFTD and SD (*P*<0.01) and a difference between bvFTD and PNFA that approaches significance (*P*=0.057). On the Boston Naming Test, the differences between the bvFTD and the SD and the PNFA groups were highly significant (*P*<0.0001). The difference between PLA and SD was also significant (*P*=0.029).

Semiautomated Psycholinguistic Measures

The measures obtained from all 38 study participants showed significant differences in means on one-way ANOVA tests for pause-to-word ratio (*P*<0.01), normalized dysfluent event (*P*<0.001) count, and the ratio of pronouns to nouns (*P*=0.01) scores. The distribution of the scores for these 3 measurements is presented in Figures 3–5, respectively. Tukey post hoc test for pairwise differences in means for the pause-to-word ratio and the normalized dysfluent event variables showed significant differences between PNFA and all 3 other groups— bvFTD (*P*<0.001), PLA (*P*=0.01) and SD (*P*<0.001). For the pronoun-to-noun ratio variable, significant difference was found only between bvFTD and SD groups (*P*=0.02). The differences in means for all the remaining measures were not statistically significant. However, we did find that the proportion of nouns and verbs was higher in the PNFA group (26% for nouns and 20% for verbs) than in the SD group (20% for nouns and 18% for verbs). None of the composite variables obtained from the principal components analysis showed significant differences in means among the FTLD groups.

Precision of Semiautomated Alignment and Part-of-speech Tagging

The mean difference between fully manual and semiautomatic alignments at the word-initial boundary was 580 ms (SD=920). The mean difference at the word-final boundary was 680 ms (SD=905). Of all semiautomated word-initial boundary alignments, 76% had a difference from manual alignments of less than 500 ms, whereas this number was 74% for the word-final boundary alignments. Overall, the accuracy of the automated part-of-speech tagging, as compared with manual review, was 84% (SD=8). The automatic part-of-speech tagger carried out best on subjects with PNFA (86%, SD=11) and worst on subjects with bvFTD (84%, SD=11). A comparison of the verbatim transcriptions created for the semiautomated alignment and those created during the subsequent manual alignment revealed an 88% (SD=5.58) absolute agreement between these 2 types of transcription.

DISCUSSION

Our computerized approach to quantifying language uses characteristics obtained from spontaneous speech, removing the subjectivity inherent in manual assessments, and thus may improve the reliability and the comparability of measurements across different studies. Current computer-based neuropsychologic tests face a number of challenges because they offer only indirect analogues of the traditional "paper-and-pencil" tests. Effective speech recognition could enable the development of direct analogues and open up a wide range of new possibilities for neuropsychologic testing.[50] Current commercial speech recognition applications do not have the robustness required, for example, to transcribe fully automatically spontaneous picture descriptions spoken by patients with FTLD. However, the semiautomated approach described in this article offers a viable alternative to bridge the gap between computerized and manual testing based on spontaneous speech. Our approach relies

on a human transcriptionist (not necessarily trained in neuropsychologic testing) to carry out tasks that are currently beyond the reach of computers (eg, using global and local semantic contexts during the process of recognizing speech). At the same time, our approach relies on computers rather than humans to do what computers do best (eg, capture frequency and duration of prosodic events). This synergistic combination of human and computer capabilities offers an opportunity to examine in detail a number of speech and language characteristics including syntax, semantics, and prosody useful in diagnostic assessments of FTLD and other syndromes affecting language.

The results of our study are consistent with other studies that investigated the use of computerized speech analysis for the diagnosis of mild cognitive impairment[13] and aphasia in children.[12] We found that speech hesitancy characterized by the ratio of silent pauses to words and the ratio of disfluent events to words are sensitive indicators of PNFA. Fluency of spontaneous speech has been earlier found to be significantly decreased in PNFA patients as compared with controls and other FTLD groups.[15] The same study found that speech fluency was generally decreased in all FTLD variants as compared with healthy controls with partially overlapping neuroanatomical sources associated with fluency.

Two clinical measures (verbal category fluency and correctness on the Boston Naming Test) and 2 semiautomated psycholinguistic measures (pronoun-to-noun ratio and pause-to-word ratio) showed significant differences between the means among the 4 FTLD diagnostic groups. The findings on the clinical measures are consistent with some of the prior work showing that category fluency and confrontation naming are among the first single-word measures to be affected in patients presenting with primary progressive aphasia[16] and FTLD.[7,40] Our results suggest that the category fluency of the PNFA group is similar to that of the bvFTD group in contrast to the PLA and the SD patients. Whereas these results are in keeping with some of the earlier studies,[51,52] other studies have found slightly more impaired performance on this test in the PNFA group,[53] possibly owing to heterogeneity of the PNFA group and the overlap between diagnostic subtypes of FTLD.

We also found that the mean scores on the verbal category fluency test were similarly high for the bvFTD (12.87) and the PNFA (11.90) variants as compared with the PLA (7.75) and SD (5.33) variants. The mean scores on the Boston Naming Test were very close for bvFTD (23.20) and PNFA (23.40) but different for the PLA (16.75) and the SD (6.00) variants. These results support earlier findings that the category fluency and the Boston Naming tests may be more sensitive to semantic deficits and are likely to be useful in the diagnosis of SD and anomia. Both of these tests fail to distinguish between the behavioral and the nonfluent aphasia variants of FTLD. The results obtained with the pronoun-to-noun ratio are similar to those with the Boston Naming Test in that they identify the SD variant as having the highest ratio of pronouns (0.78) corresponding to the lowest BNT score (6.00). However, the results obtained with computerized measurements of the pause-to-word ratio indicate that the PNFA variant has nominally the highest proportion of pauses and dysfluent events including filled pauses, false starts, and repetitions in their speech and thus, may complement the standard verbal category fluency and the Boston Naming tests in distinguishing between the 4 FTLD variants.

Our findings for the semiautomated psycholinguistic measures show that pronoun-to-noun ratio was the highest in subjects with the SD variant and lowest in subjects with the PNFA variant. The former condition is characterized by an impaired ability to access names of objects, and thus may lead to the observed tendency for increased pronoun use, whereas the latter condition involves difficulties with speech production but not necessarily picture naming. In earlier studies, patients with SD variant of FTLD were found to be significantly more impaired on a picture naming test as compared with the PNFA and bvFTD

variants.[52,53] Patients with PNFA also produced more errors on the BNT test than healthy controls; however, these errors were predominantly phonologic in nature, suggesting intact semantic store in this group.[52] Overuse of pronouns (eg, "empty" speech) is a prominent feature of progressive aphasia found in later stages of Alzheimer disease[28,54] and is likely to be a manifestation of semantic deficits in FTLD that may also lead to impaired use of nouns and verbs in this population.[16,55–57] Similar to these earlier studies, our data also suggest involvement of verbs and nouns. The proportion of both verbs and nouns was lower in patients with SD, albeit not significantly so. The fact that the ratio of pronouns to nouns showed a significant difference indicates that the use of closed class (pronouns) and open class (nouns, verbs, adjectives, etc.) words diverge in this population, which is consistent with earlier findings in fluent and nonfluent progressive aphasia variants.[22]

Pause-to-word ratio measures the hesitancy of speech, and the group diagnosed with the PNFA variant of FTLD had the highest pause-to-word ratio mean of 0.59. This ratio indicates that more than half of the audio recordings of picture descriptions by these patients consisted of silence. These findings are also consistent with earlier studies of nonfluent progressive aphasias and the PNFA variant of FTLD[15,52,53] that showed decreased performance in this group on the letter fluency test and spontaneous speech fluency assessments.[3,15,17–19]

The results of our exploratory factor analysis on the semiautomated measurements indicate that the measurements may be grouped into 4 composite variables roughly corresponding to the length/duration of the picture description, speech hesitancy, empty speech (preponderance of pronouns and false starts), and grammaticality. Although these categories are subjectively determined and do not capture all aspects of the components, the fact that the analysis resulted in 4 major components whose nature (albeit subjective) is consistent with described phenomenology of the progressive aphasias.

The precision of word-initial and word-final boundary semiautomated alignment was better than expected, given the conversational nature of the discourse, quality of the recording, and the population with impaired speech. In a study of automatic alignment accuracy on spontaneous speech obtained in conversational dialogues, Chen et al[49] reported word-initial and word-final differences between automatic and manual alignments on entire conversational turns in excess of 2.3 seconds. These differences were greatly reduced to less than 50 ms when the dialogues were presegmented on silences of greater than 500 ms resulting in shorter utterances. A major drawback of segmenting dialogues on silences, however, is that the transcripts must also be segmented, which is a manual and labor-intensive process. Thus, we did not use this technique in our study. The alignment of spontaneous dialogues is inherently a more difficult task than the picture description and contains multiple points of overlap in which both speakers talk at the same time. The audio used in our study had fewer instances of cross-talk and also fewer speaker turns, thus resulting in smaller alignment differences. The accuracy of automatic part-of-speech tagging was also consistent with earlier reported results. Brants had earlier trained and evaluated the part-of-speech tagger that was used in this study on a corpus of Wall Street Journal articles manually tagged for part-of-speech. The tagger was found to be 97% accurate on predicting the part-of-speech of the words that were present in the training data and 86% accurate on new words present only in the test data.[58] Our results show that this tagger is 84% accurate, likely owing to the differences between the data used to train the tagger (Wall Street Journal) and the conversational discourse of the picture descriptions resulting in new vocabulary for which the tagger had not been trained. Whereas the accuracy of 84% is good (approximately 1 out of 10 words is mislabeled), it can be further improved by adapting the tagger specifically to the language used in picture description tests.

Characterization of speech in the progressive aphasias has important implications for diagnosis. There is increasing recognition that the different subtypes of progressive aphasia, including PNFA, SD, and PLA, have different anatomic and biochemical bases. Proper identification of the expressive speech disorder plays an important role in differential diagnosis. 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 PNFA and the TDP-43 proteinopathy associated with semantic dementia. Although automated speech analysis could not replace clinicians, the automated approach offers a standardized way of characterizing expressive speech and could serve as a means of classifying subjects for a clinical trial, either by supporting or calling into question a clinical diagnosis.

Limitations and Future Directions

Certain limitations must be acknowledged to enable the interpretation of the results of this study. First, the sample size used in this study is relatively small. This is particularly important for the interpretation of the factor analysis results, which are preliminary and suggestive rather than conclusive. Many more samples will be required for a more comprehensive analysis that may include clustering of speech and language characteristics to define FTLD variants. FTLD is a relatively rare condition, which limits obtainable sample size; however, we continue to verify our results as we obtain more samples. Second, the semiautomated approach to language and speech analysis resulted in some loss of precision in both time alignments and part-of-speech identification. The former is owing to the quality of the available audio, whereas the latter is likely owing to the fact that the statistical model used for automatic part-of- speech tagging was trained on a publicly available manually labeled corpus of written language (Penn Treebank—Wall Street Journal).[58] Retraining the model on a suitable corpus of spontaneous speech manually labeled for part-of-speech may yield higher accuracy. The audio samples used in this study were initially collected for traditional manual rather than automatic analysis using an analog tape recorder. Despite the relatively poor quality of the audio, we found alignment accuracy within 500 ms. Going forward, we are optimizing the technology and procedures used for audio collection to obtain high quality digital audio to enable more precise alignment of audio and text. A pilot test conducted on picture descriptions by 5 healthy younger adults from a different study showed that fully manual and semiautomatic alignments were within 80 ms at word onsets and 230 ms at word endings. Third, the current implementation of the system does not identify phrasal or sentential boundaries— our syntactic analysis is currently limited to automatic part-of-speech identification and part-of-speech perplexity calculation. Being able to distinguish intrautterance from interutterance pauses and hesitations may improve the sensitivity of our prosodic measurements in addition to providing more detailed information on syntactic violations indicative of grammaticality. Fourth, the acoustic model of the speech recognizer used in this study was trained on spontaneous speech of English speakers. Thus, the generalizability of our results to speakers of other languages, English as a second language, and various social and regional dialects remains to be determined. From the technological standpoint, our system is extensible to other languages, provided that appropriate acoustic, language, and part-of-speech models are available or can be created. Fifth, similarly to the model used by the automatic part-of-speech tagger, the statistical models used for calculating part-of-speech and word level perplexity were trained on general English text. Retraining these models on picture descriptions by healthy controls of the same age and socioeconomic background as patients may change the results. Finally, the system described in this manuscript is semiautomated. As such, it still requires human input in the form of verbatim transcription of the speech samples. To make the system operate in a fully automated fashion, it will be necessary to develop an ASR engine that will carry out speech to text transcription rather than alignment of manually transcribed speech with the

audio. The challenge will be to train a system that can operate on impaired speech and to determine the acceptable level of the accuracy of the system. This study lays the foundation for these future investigations.

Acknowledgments

References

1. Rascovsky K, Hodges JR, Kipps CM, et al. Diagnostic criteria for the behavioral variant of frontotemporal dementia (bvFTD): current limitations and future directions. Alzheimer Dis Assoc Dis. 2007; 21:S14–S18.

2. Gorno-Tempini ML, Brambati SM, Ginex V, et al. The logopenic/phonological variant of primary progressive aphasia. Neurology. 2008; 71:1227–1234. [PubMed: 18633132]

3. Gorno-Tempini ML, Dronkers NF, Rankin KP, et al. Cognition and anatomy in three variants of primary progressive aphasia. Ann Neurol. 2004; 55:335–346. [PubMed: 14991811]

4. Josephs KA, Whitwell JL, Duffy JR, et al. Progressive aphasia secondary to Alzheimer disease vs FTLD pathology. Neurology. 2008; 70:25–34. [PubMed: 18166704]

5. Kertesz A, McMonagle P, Blair M, et al. The evolution and pathology of frontotemporal dementia. Brain. 2005; 128(Pt 9):1996–2005. [PubMed: 16033782]

6. Knibb JA, Xuereb JH, Patterson K, et al. Clinical and pathological characterization of progressive aphasia. Ann Neurol. 2006; 59:156–165. [PubMed: 16374817]

7. Knopman DS, Kramer JH, Boeve BF, et al. Development of methodology for conducting clinical trials in frontotemporal lobar degeneration. Brain. 2008; 131:2957–2968. [PubMed: 18829698]

8. Mesulam M, Wicklund A, Johnson N, et al. Alzheimer and frontotemporal pathology in subsets of primary progressive aphasia. Ann Neurol. 2008; 63:709–719. [PubMed: 18412267]

9. Rohrer JD, Knight WD, Warren JE, et al. Word-finding difficulty: a clinical analysis of the progressive aphasias. Brain. 2008; 131(Pt 1):8–38. [PubMed: 17947337]

10. Goodglass, H.; Kaplan, E. The Assessment of Aphasia and Related Disorders. Philadelphia: Lea and Febiger; 1983.

11. Beele KA, Davies E, Muller DJ. Therapists' views on the clinical usefulness of four aphasia tests. Brit J Dis Commu. 1984; 19:169–178.

12. Hosom JP, Shriberg L, Green JR. Diagnostic assessment of childhood apraxia of speech using automatic speech recognition (ASR) methods. J Med Speech-Language Pathol. 2004; 12:167–171.

13. Roark, B.; Hosom, J.; Mitchell, M., et al. Automatically derived spoken language markers for detecting Mild Cognitive Impairment. Paper presented at the International Conference on Technology and Aging (ICTA); 2007.

14. Hodges J, Davies R, Xuereb J, et al. Clinicopathological correlates in frontotemporal dementia. Annals Neurol. 2004; 56:399–406.

15. Ash S, Moore P, Vesely L, et al. Non-fluent speech in frontotemporal lobar degeneration. J Neuroling. 2008; 2:370–383.

16. Bird H, Lambon Ralph MA, Patterson K, et al. The rise and fall of frequency and imageability: noun and verb production in semantic dementia. Brain Lang. 2000; 73:17–49. [PubMed: 10872636]

17. Grossman M. Frontotemporal dementia: a review. J Int Neuropsychol Soc. 2002; 8:566–583. [PubMed: 12030310]

18. Peelle J, Cooke A, Moore P, et al. Syntactic and thematic components of sentence processing in progressive nonfluent aphasia and nonaphasic frontotemporal dementia. J Neuroling. 2007; 20:482–494.

19. Weintraub S, Rubin NP, Mesulam M. Primary progressive aphasia. Longitudinal course, neuropsychological profile, and language features. Arch Neurol. 1990; 47:1329–1335. [PubMed: 2252450]

20. Hodges J, Patterson K, Oxbury S, et al. Semantic dementia. Progressive fluent aphasia with temporal lobe atrophy. Brain. 1992; 115(Pt 6):1783–1806. [PubMed: 1486461]

21. Neary D, Snowden JS, Gustafson L, et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. Neurology. 1998; 51:1546–1554. [PubMed: 9855500]

22. Thompson CK, Ballard KJ, Tait ME, et al. Patterns of language decline in non-fluent primary progressive aphasia. Aphasiology. 1997; 11:297–321.

23. Josephs KA, Duffy JR, Strand EA, et al. Clinicopathological and imaging correlates of progressive aphasia and apraxia of speech. Brain. 2006; 129(Pt 6):1385–1398. [PubMed: 16613895]

24. Adlam AL, Patterson K, Rogers TT, et al. Semantic dementia and fluent primary progressive aphasia: two sides of the same coin? Brain. 2006; 129(Pt 11):3066–3080. [PubMed: 17071925]

25. Cooke A, DeVita C, Gee J, et al. Neural basis for sentence comprehension deficits in frontotemporal dementia. Brain Lang. 2003; 85:211–221. [PubMed: 12735939]

26. Murray R, Koenig P, Antani S, et al. Lexical acquisition in progressive aphasia and frontotemporal dementia. Cogn Neuropsychol. 2007; 24:48–69. [PubMed: 18416483]

27. Ash S, Moore P, Antani S, et al. Trying to tell a tale: discourse impairments in progressive aphasia and frontotemporal dementia. Neurology. 2006; 66:1405–1413. [PubMed: 16682675]

28. Almor A, Kempler D, MacDonald M, et al. Why do Alzheimer patients have difficulty with pronouns? Working memory, semantics, and reference in comprehension and production in Alzheimer's disease. Brain Lang. 1999; 67:202–227. [PubMed: 10210631]

29. Almor A, MacDonald M, Kempler D, et al. Comprehension of long distance number agreement in probable Alzheimer's disease. Lang Cogn Processes. 2001; 16:35–63.

30. Glosser G, Deser T. Patterns of discourse production among neurological patients with fluent language disorders. Brain Lang. 1991; 40:67–88. [PubMed: 2009448]

31. Gordon J. A quantitative production analysis of picture description. Aphasiology. 2006; 20:188–204.

32. Berndt, RS.; Waylannd, S.; Rochon, E., et al. Quantitative Production Analysis: A Training Manual for the Analysis of Aphasic Sentence Production. Hove, UK: Psychology Press; 2000.

33. Saffran EM, Berndt RS, Schwartz MF. The quantitative analysis of agrammatic production: procedure and data. Brain Lang. 1989; 37:440–479. [PubMed: 2804622]

34. Gordon J. Measuring the lexical semantics of picture description in aphasia. Aphasiology. 2008; 22:839–852. [PubMed: 22399832]

35. 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; 36:338–350. [PubMed: 8487525]

36. Yorkston KM, Beukelman DR. An analysis of connected speech samples of aphasic and normal speakers. J Speech Hearing Dis. 1980; 45:27–36. [PubMed: 7354627]

37. Artstein R, Poesio M. Inter-Coder Agreement for computational linguistics. Comput Linguist. 2008; 34:555–596.

38. Krippendorff K. Reliability in content analysis: some common misconceptions and recommendations. Human Commu Res. 2004; 30:411–433.

39. Poesio M, Vieira R. A corpus-based investigation of definite description use. Comput Linguist. 1998; 24:183–216.

40. Libon DJ, Xie SX, Moore P, et al. Patterns of neuropsychological impairment in frontotemporal dementia. Neurology. 2007; 68:369–375. [PubMed: 17261685]

41. Forman MS, Farmer J, Johnson JK, et al. Frontotemporal dementia: clinicopathological correlations. Ann Neurol. 2006; 59:952–962. [PubMed: 16718704]

42. Delis, DC.; Kramer, JH.; Kaplan, E., et al. California Verbal Learning Test. 2. San Antonio, TX: The Psychological Crop; 2000.

43. Mohs RC, Knopman D, Petersen RC, 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–S21. [PubMed: 9236948]

44. Wechsler, DA. Wechsler Memory Scale-Revised. New York: Psychological Corporation; 1987.

45. Stroop JR. Studies of interference in serial verbal reactions. J Experi Psychol. 1935; 18:643–662.

46. Wechsler, DA. Wechsler Adult Intelligence Scale-Revised. New York: Psychological Corporation; 1981.

47. Benton, AL.; Hamsher, K.; Sivan, AB. Multilingual Aphasia Examination. 3. Iowa City, IA: AJA Associates; 1983.

48. Kaplan, E.; Goodglass, H.; Weintraub, S. The Boston Naming Test. Boston: Lea & Fabiger; 1978.

49. Chen, L.; Liu, Y.; Harper, M., et al. Evaluating factors impacting the accuracy of forced alignments in a multimodal corpus. Paper presented at the Language Resource and Evaluation Conference (LREC); 2004.

50. Letz R. Continuing challenges for computer-based neuropsychological tests. Neurotoxicology. 2003; 24:479–489. [PubMed: 12900061]

51. Clark DG, Charuvastra A, Miller BL, et al. Fluent versus nonfluent primary progressive aphasia: a comparison of clinical and functional neuroimaging features. Brain Lang. 2005; 94:54–60. [PubMed: 15896383]

52. Nestor PJ, Graham NL, Fryer TD, et al. Progressive non-fluent aphasia is associated with hypometabolism centred on the left anterior insula. Brain. 2003; 126(Pt 11):2406–2418. [PubMed: 12902311]

53. Libon DJ, Xie SX, Wang X, et al. Neuropsychological decline in frontotemporal lobar degeneration: a longitudinal analysis. Neuropsychology. 2009; 23:337–346. [PubMed: 19413447]

54. Kempler, D. Language changes in dementia of the Alzheimer's type. In: Lubinski, R., editor. Dementia and Communication: Research and Clinical Implications. San Diego: Singular; 1995. p. 98-114.

55. Hillis AE, Oh S, Ken L. Deterioration of naming nouns versus verbs in primary progressive aphasia. Ann Neurol. 2004; 55:268–275. [PubMed: 14755731]

56. Price CC, Grossman M. Verb agreements during on-line sentence processing in Alzheimer's disease and frontotemporal dementia. Brain Lang. 2005; 94:217–232. [PubMed: 15896395]

57. Silveri MC, Salvigni BL, Cappa A, et al. Impairment of verb processing in frontal variant-frontotemporal dementia: a dysexecutive symptom. Dement Geriatr Cogn Disord. 2003; 16:296–300. [PubMed: 14512727]

58. Brants, T. TnT—A Statistical Part-of-Speech Tagger. Paper presented at the Sixth Applied Natural Language Processing Conference (ANLP 2000); Seattle, WA. 2001.

59. Young, S.; Kershaw, D.; Odell, J., et al. The HTK Book Version 3.4. Cambridge, England: Cambridge University; 2006.

60. Heeman PA, Allen JF. Speech repairs, intonational phrases, and discourse markers: modeling speakers' utterances in spoken dialogue. Comput Linguist. 1999; 25:527–571.

61. Forrest K, Weismer G, Turner GS. Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults. J Acoustical Soc Am. 1989; 85:2608–2622.

62. Levelt, WJ. Speaking: From Intention to Articulation. Cambridge, MA: MIT Press; 1989.

63. Boersma P. Praat, a system for doing phonetics by computer. Glot International. 2001; 5:341–345.

64. Godfrey, JJ.; Holliman, E. Switchboard-1 Release 2. 1997. Retrieved 11/03/2008, from Linguistic Data Consortium: http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC97S62

65. Bahl, L.; Baker, J.; Jelinek, E., et al. Perplexity–a measure of the difficulty of speech recognition tasks. Paper presented at the 94th Meeting of the Acoustical Society of America; 1977.

66. Wightman, C.; Ostendorf, M. Automatic recognition of prosodic phrases. Paper presented at the International Conference on Acoustics, Speech, and Signal Processing; 1991.

67. Brookshire RH, Nicholas LE. Speech sample size and test-retest stability of connected speech measures for adults with aphasia. J Speech Hearing Res. 1994; 37:399–407. [PubMed: 8028321]

68. Nicholas LE, Brookshire RH. Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. J Speech Hearing Res. 1995; 38:145–156. [PubMed: 7731205]

APPENDIX A: AUTOMATIC ALIGNMENT

The ASR engine used for alignment was based on the Hidden Markoff Model Toolkit (HTK 3.4) developed at Cambridge University Labs, UK.[59] We trained a speaker-independent acoustic model using speech and transcripts obtained from the TRAINS corpus containing samples of 6.5 hours of speech from 91 spontaneous dialogues containing 55,000 transcribed words.[60]

Language models were automatically constructed from transcripts for each of the 38 speech samples in this study and were represented by simple deterministic word-level networks. Each network contained all words from the picture description transcript with nonoptional transitions between words, false starts, and filled pauses, and optional silences inserted at each word boundary. A standard US English pronunciation dictionary was used to bridge the phoneme-level acoustic model and the word-level language model. To address instances of dysarthric or otherwise distorted speech, the transcribers were instructed to represent distorted speech as closely as possible to the original orthographically. For example, in 1 instance the patient had trouble with saying "cookie jar" where the first voiceless stop (k) in "cookie" was pronounced as a voiceless fricative (ch) and the word "jar" was not pronounced. This instance was transcribed as "chookie." As this word is not found in the standard English dictionary, its pronunciation was automatically constructed on-the-fly resulting in a simplified (consonant- vowel-consonant-vowel) representation in which consonants represented by phoneme (t) and vowels were represented by the midcentral (ah). Completely uninterpretable speech is also represented in transcription orthographically as closely as possible to the sound and represented on-the-fly as a sequence of "default" sounds [(t) and (ah)] in the dictionary used by the speech recognizer. Although this method does not allow us to access the content of the word (phonologic or semantic), it allows us to identify the speech/silence boundaries necessary to compute prosodic measurements, such as the pause-to-word ratio.

APPENDIX B: SEMIAUTOMATED PSYCHOLINGUISTIC MEASUREMENTS

Pause-to-word Ratio

This variable represents a simple ratio of pauses to words. A silent pause is defined as a silent segment longer than 150 ms. This cutoff was chosen conservatively to avoid counting phonetically conditioned pauses such as the release phase in the phonation of a word-final stop consonant that may last up to 100 ms in duration.[61,62] All nonsilent segments with the exception of filled pauses (um's and ah's) were treated as words in calculating this measure.

Fundamental Frequency Variance

Fundamental frequency (FF) is the lowest frequency in the harmonic series produced by an instrument with resonant properties including the human windpipe and mouth. In speech analysis, fundamental frequency is associated with pitch. Changes in FF that occur over the span of a single phoneme, word, or utterance are treated as indicative of the intonation. Thus, we use the variability in FF over the entire duration of nonsilent segments produced by patients with FTLD as an indicator of the variation in pitch or intonation. Lack of this variation may be indicative of "flat affect" and create an impression of reduced prosody. In our experiments, FF variance was calculated using the pitch-tracking tools available as part of the Praat system.[63]

Part-of-speech Perplexity

The distribution of parts-of-speech in normal language is not random and can be captured with a probabilistic language model by calculating the conditional probability of a part-of-speech occurring in a specified position given 1, 2, 3 or more preceding parts-of-speech. Such part-of-speech language model captures the notion of grammaticality in which, in normal spoken English, the likelihood of seeing an adjective immediately after a noun (eg, boy/noun little/adjective) is much smaller than the likelihood of the reverse (little/adjective boy/noun). In general terms, the language model "perplexity" is a measure of how well the part-of-speech sequence obtained from the picture description task "fits" the language model constructed from a set of reference English utterances.

For this study, we used a corpus of spontaneous telephone conversations (SWITCHBOARD) obtained from native English speakers to construct the reference part-of-speech language model. The corpus comprised 2430 6-minute conversations between 500 male and female speakers from all major American dialect groups.[64] The parts-of-speech in the transcripts of the Cookie-Theft descriptions for this study were determined using an automatic part-of-speech tagger-TnT.[58] The tagger operates by calculating the probability of a part-of-speech occurring in the context of surrounding words and parts-of-speech. For example, in "cookie in his hand," the word hand would be tagged as a noun because statistically that is a more likely category to occur after a preposition (in) and a possessive pronoun (his) than a verb. However, in "the boy wants to hand his sister a cookie," the word hand will be tagged as a verb because that is a more likely category in the context of another verb (wants) followed by an infinitive (to) than a noun. The details of this statistical part-of-speech algorithm can be found in the article by Brants, 2001.[58]

Word-level Perplexity

This measure captures the perplexity of a language model constructed the same way as measuring part-of-speech perplexity, with the exception that the perplexity is calculated using sequences of words rather than their parts-of-speech. For example, a language model may capture the fact that the word sequence "washing dishes" is more likely to be spoken by a healthy adult native speaker of English than "washing cookie." The likelihood of these word sequences is determined by counting how often the words "dishes" and "cookie" follow the word "washing" in a representative collection of naturally occurring English utterances. Once the language model is trained on a collection of utterances, we can measure how well it will predict the words in new utterances that were not used in the initial training. A model that is more efficient in predicting the words in the new utterances is said to have lower perplexity.[65] Lower perplexity indicates better generalizability of the model to new utterances and represents greater similarity between the language used to train the model and the language of the new (test) utterances.

Pronoun-to-noun Ratio

After the part-of-speech for each word in the text of the picture description is automatically labeled using the TnT part-of-speech tagger, we compute the ratio of words tagged as pronouns to words tagged as nouns.

Word Count

This measure represents a count of all words found in the picture description transcript. Silent and filled pauses and other vocalizations such as breaths, sighs, and noise were excluded.

Total Duration of Speech in the Sample

This measure was computed by adding up the durations of all nonsilent elements excluding filled pauses, noises, and false starts in the audio alignment of the picture description.

Mean Prosodic Phrase Length

Spontaneous speech produced by healthy speakers has a certain rhythm that is made up of changes in speaking rate, stress, and hesitation patterns. A sharp change in rhythm may be indicative of a rhythmic or prosodic phrase boundary, distinct from syntactic boundaries. [62] To identify changes in speech rhythm, we followed the methodology proposed by Wightman and Ostendorf[66] that relies on calculating the average normalized duration of phonemes. Wightman and Ostendorf[66] used the difference in the average normalized duration of the last 3 syllables in the coda of the preceding words from the first 3 syllables in the onset of the word after the word boundary as a predictive feature for detecting prosodic phrase boundaries. We used Wightman and Ostendorf's methodology for computing speaker normalized segment durations, with the simplifying exception that we averaged the segment duration over the entire word rather than only its coda or onset. We then computed the difference in averaged normalized segment durations for each pair of adjacent words in the input. If the difference was greater than 1 standard deviation computed over the entire speech sample, we marked that location as a rhythmic boundary. We were then able to use these locations to calculate the average length of a prosodic phrase.

Correct Information Unit Count

We developed an automated approach to counting Correct Information Units (CIU) as defined by Brook-shire and Nicholas.[67,68] CIUs represent words and phrases that reflect the conceptual content of the Cookie Theft picture. In our implementation, the text of the transcript obtained in the picture description task is searched electronically for sequences of 1,2,3, and 4 words to find matches to a predefined list of word sequences representing CIUs. The list of CIUs and word sequences was compiled using Yorkston and Beukelman's[36] list of concept units as a starting point. The list was expanded to include lexical and morphologic word variants (eg, falling, fall, fell). The complete list used in this study is provided in Appendix C.

Normalized Long Pause Count

This variable represents a normalized count of all silent pauses greater than 400 ms in duration. This measure and the subsequent measures in 11–14 were normalized by dividing their value by the total length of speech in the sample.

Normalized Filled Pause Count

Our system distinguishes between 2 types of pause fillers—the shorter ones without nasalization [eg, (ah), (eh)] and the longer ones with nasalization [eg, (uhm)]; however, for this study we collapsed the 2 filled pause types into 1.

Normalized Silent Pause Count

Only silences greater than 150 ms in duration were counted as silent pauses as described earlier.

Normalized False Start Count

This is a normalized count of false starts in which the speaker begins to speak a word but does not finish the word.

Normalized Repetition Count

Sequences of 2 and/or 3 words adjacent to each other were counted as repetitions.

Normalized Dysfluent Even Count (Filled Pauses, False Starts, and Repetitions)

This is a combined normalized count of all filled pauses, repetitions, and false starts.

APPENDIX C: LIST OF INFORMATION UNITS AND THEIR VARIANTS

two

children

little

boy

girl

sister

brother

kid

kids

standing

by the boy

by her brother

near the boy

near her brother

reaching up

reach up

reach

on stool

wobbling

off balance

three legged

3-legged

fall over

fall off

falling over

falling off

tip over

tipping over

by boy

on the floor

hurt himself

reach up

reaching up

reaching

reach

taking

take

took

stealing

pull

pulling

stole

steal

cookie

cookies

for himself

for his sister

pouring

from the jar

cookie jar

on the high shelf

on the top shelf

on the shelf

from the high shelf

from the top shelf

from the shelf

in the cupboard

with the open door

handing to sister

handing to his sister

handing cookies

to his sister

to the sister

asking for cookie

asking for a cookie

asking for cookies

trying to get

ask for cookie

ask for a cookie

ask for cookies

has finger to mouth

finger to mouth

finger to her mouth

pressed to her mouth

saying shhh

keeping him quiet

keeping the boy quiet

trying to help

laugh

laughing

mother

woman

lady

children behind her

children behind the woman

children behind the mother

behind the mother

standing

by the sink

washing

doing

dishes

on the counter

drying

faucet is on

full blast

ignoring

daydreaming

oblivious

paying attention

water

overflowing

onto the floor

on the floor

on the shoes

on her shoes

onto the shoes

onto her shoes

on her feet

onto her feet

feet are getting wet

getting wet

dirty dishes left

puddle

in the kitchen

indoors

inside

outside

disaster

lawn

road

path

tree

trees

driveway

house

next door

garage
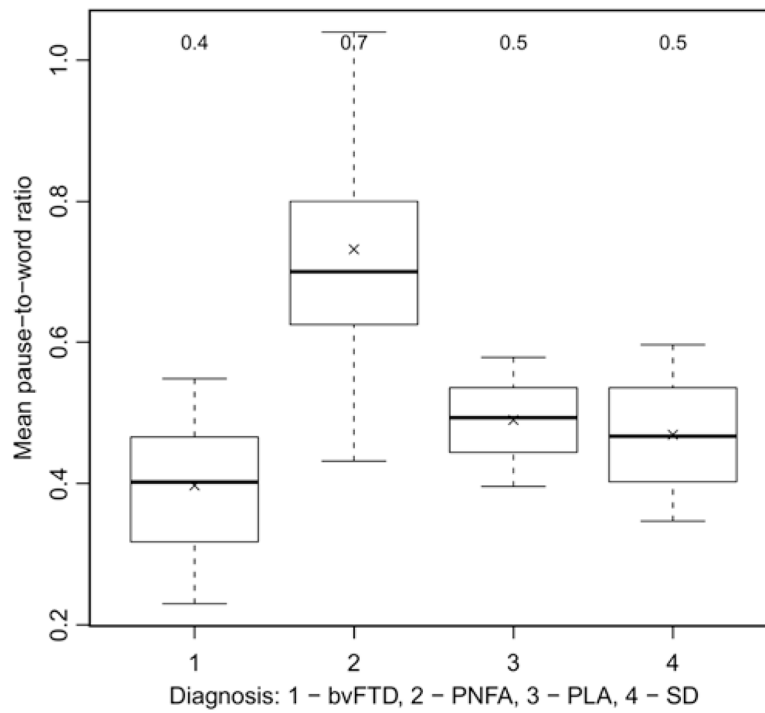
open window

window is open

curtains

drapes

draperies

plate

towel
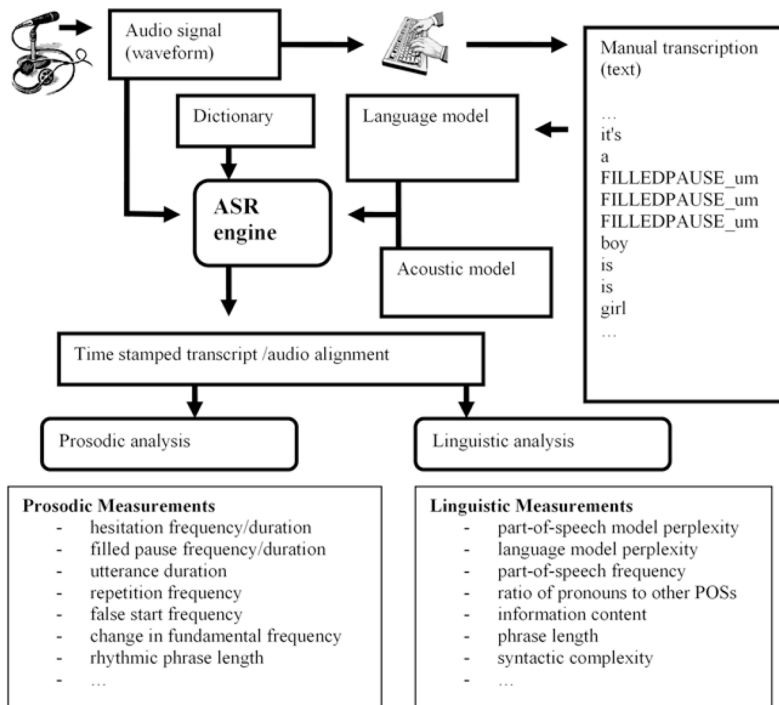
FIGURE 1.
Flowchart illustrating the operation of the computerized system for speech and language assessment.
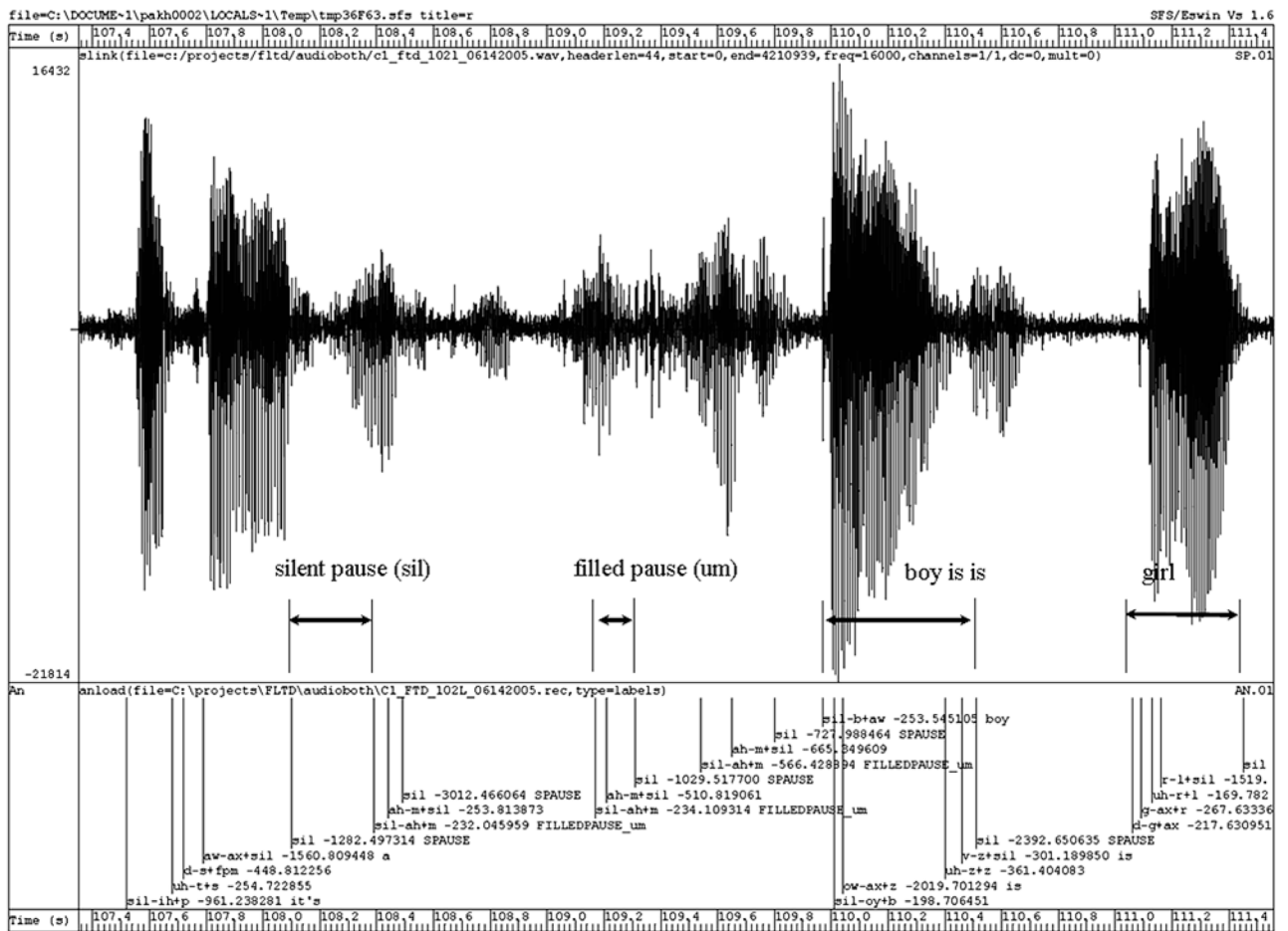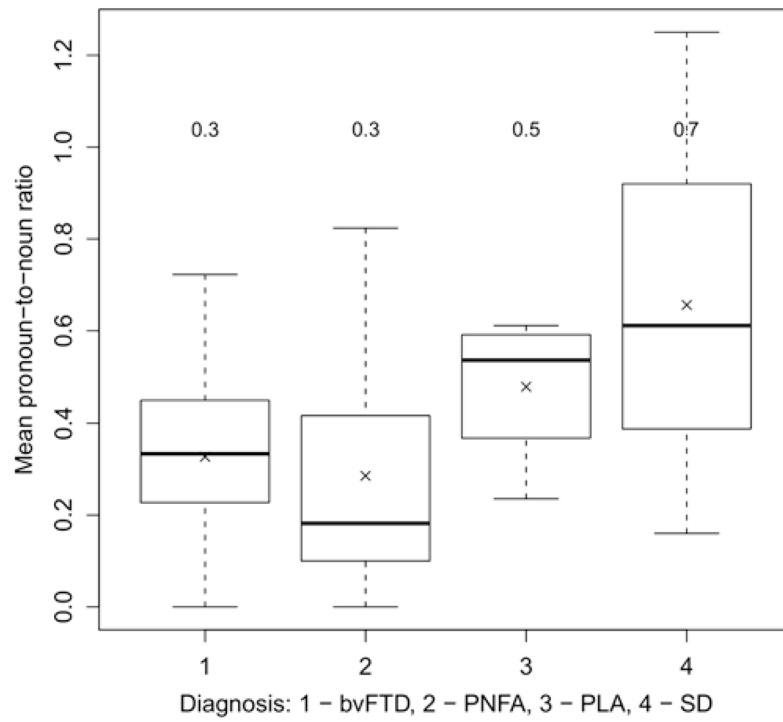
FIGURE 2.
An example of time-aligned portion of the transcription from a Cookie-Theft picture description task.

FIGURE 3.
Pause-to-word ratio means in 4 frontotemporal lobar degeneration variants. Group means by diagnosis are indicated with (×) with mean values at the top of each boxplot.
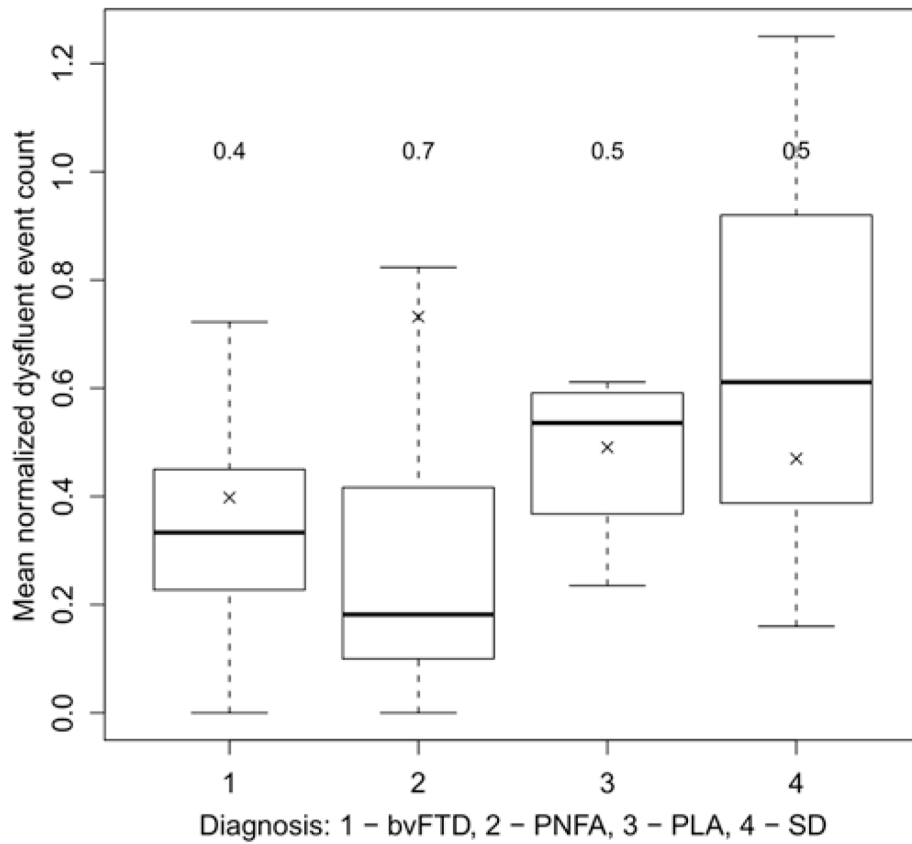
FIGURE 4.
Normalized dysfluent event means in 4 frontotemporal lobar degeneration variants. Group means by diagnosis are indicated with (×) with mean values at the top of each boxplot.

FIGURE 5.
Pronoun-to-noun ratio means in 4 frontotemporal lobar degeneration variants. Group means by diagnosis are indicated with (×) with mean values at the top of each boxplot.

TABLE 1

Rotated Component Matrix Obtained With Principal Component Analysis of the Semiautomated Psycholinguistic Measurements on all 38 Picture Description Samples [*]

	Component			
	1	**2**	**3**	**4**
Variable	**Length**[†]	**Hesitancy**	**Empty Content**	**Grammaticality**
Pause-to-word ratio	−0.148	0.803	0.132	−0.020
Fundamental frequency variance	0.042	−0.162	−0.075	0.798
POS perplexity	−0.041	0.380	0.279	0.726
Word perplexity	−0.407	0.529	0.024	0.213
Pronoun-to-noun ratio	0.108	−0.375	0.729	0.223
Word count	0.932	0.002	0.000	0.126
Speech duration (ms)	0.864	−0.063	−0.027	0.245
Mean prosodic Phrase length	−0.498	−0.550	−0.013	0.006
Correct Information unit count	0.726	0.053	−0.390	−0.217
Long pause count	−0.322	0.288	0.844	−0.135
Filled pause count	0.195	0.651	0.143	0.206
Pause count	−0.182	0.426	0.830	−0.158
False start count	0.329	0.368	0.427	0.248
Pause-to-word ratio	0.181	0.403	−0.091	0.650

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

[*] Rotation converged in 7 iterations.

[†] Coefficients in bold represent the items used in subjective labeling of components with values exceeding 0.6.