Computerized analysis of speech and language to identify psycholinguistic correlates of fronto-temporal lobar degeneration

Serguei VS Pakhomov, Ph. D.†, Glenn E. Smith, PhD. ††, Dustin Chacon†, Yara Feliciano‡, Neill Graff-Radford, M.D. †††, Richard Caselli, M.D.*, David S. Knopman, M.D. ††

† 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

Funding: Our work 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.

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
e-mail: pakh0002@umn.edu
Abstract

Objective: To evaluate the use of a semi-automated computerized system for measuring speech and language characteristics in patients with fronto-temporal lobar degeneration (FTLD).

Background: FTLD is a heterogeneous disorder comprising at least three 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 three participating medical centers were elicited using the Cookie Theft stimulus. Each patient underwent a battery of neuropsychological tests. The audio was analyzed by the computerized system to measure 14 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: Semi-automated 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 four subjectively defined components (length, hesitancy, empty content, grammaticality) corresponding to the phenomenology of FTLD variants.

Conclusion: Semi-automated language and speech analysis is a promising novel approach to neuropsychological assessment that offers a valuable contribution to the toolbox of researchers in dementia and other neurodegenerative disorders.

Keywords: spontaneous speech; language; prosody; fronto-temporal lobar degeneration; automated speech analysis;
Introduction

The need for improved definition of syndromes and phenotypes is a key theme in dementia research in general, and fronto-temporal lobar degeneration (FTLD) in particular (Rascovsky, et al., 2007). Currently, FTLD comprises 3 syndromes: behavioral variant frontotemporal dementia (bvFTD), progressive non-fluent aphasia (PNFA) and semantic dementia (SD). The inclusion of progressive logopenic aphasia (PLA) as either a variant of FTLD or Alzheimer’s is currently being debated (Gorno-Tempini, et al., 2008; Gorno-Tempini, et al., 2004; Josephs, et al., 2008; Kertesz, McMonagle, Blair, Davidson, & Munoz, 2005; Knibb, Xuereb, Patterson, & Hodges, 2006; Knopman, et al., 2008; Mesulam, et al., 2008). These syndromes are diagnosed using standard clinical criteria, neuropsychological 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” (Rohrer, et al., 2008) 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 (Goodglass & Kaplan, 1983). The assessments with these instruments, however, are traditionally performed manually, which is subjective and may not have the detail and precision of measurements necessary to define neither the full range of syndromes, nor their nuances. Detailed measurements are particularly important to the assessment of syntax, semantics and prosody – the three areas identified in a survey of clinicians’ views on the clinical usefulness of aphasia test batteries (Beele, Davies, & Muller, 1984).

Recent advances in computerized natural language processing (NLP) and automatic speech recognition (ASR) (Hosom, Shriberg, & Green, 2004; Roark, Hosom, Mitchell, & Kaye, 2007) make it possible to develop objective and precise instruments for automated or semi-automated 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 as well as other neurodegenerative disorders affecting language. While 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 semi-automated 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 fronto-temporal dementia

Over half of all patients with symptoms of FTLD exhibit language-related manifestations on initial presentation (J. Hodges, et al., 2004). 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 (Ash, et al., 2008; Bird, Lambon Ralph, Patterson, & Hodges, 2000; Gorno-Tempini, et al., 2004; Grossman, 2002; Peelle, Cooke, Moore, Vesely, & Grossman, 2007; Weintraub, Rubin, & Mesulam, 1990). The SD variant involves multi-modal non-verbal as well as verbal naming and recognition deficits with relatively preserved grammar (J. R. Hodges, Patterson, Oxbury, & Funnell, 1992; Neary, et al., 1998). However, despite these differences between the non-fluent and fluent aphasic variants of FTLD, 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. For example, Josephs et al. (Josephs, et al., 2006) 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. (Adlam, et al., 2006) suggests that the distinction between these two classifications is a matter of emphasis rather than differences in the underlying pathophysiology of the phenomenon.

The PLA variant of FTLD has been introduced into the diagnosis of FTLD relatively recently. While bvFTD, PNFA and SD syndromes are likely to represent FTLD pathologically (Knopman, et al., 2008), the grouping of the PLA syndrome with FTLD or Alzheimer’s 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 pauses (Gorno-Tempini, et al., 2008). Some of the cases of primary progressive aphasia distinct from SD and PNFA in Josephs et al. study also exhibited these altered prosodic characteristics of speech with relatively preserved grammar and could possibly be classified as PLA (Josephs, et al., 2006).

Language-related deficits in patients diagnosed with bvFTD tend to be observed at the higher level of discourse characteristics rather than syntax, phonology and semantics found with other FTLD variants. Studies of patients with bvFTD (or social/dysexecutive variant) demonstrated impaired working memory that manifests itself through deficits in sentence comprehension (Cooke, et al., 2003), thematic role processing in verbs (Murray, Koenig, Antani, McCawley, & Grossman, 2007; Peelle, et al., 2007) and altered discourse characteristics (Ash, et al., 2006). The latter, including discourse coherence, cohesion, and “empty” speech (overuse of pronouns) are also sensitive to diffuse neural degradation characteristic of Alzheimer’s disease on tasks involving elicitation of spontaneous speech (Almor, Kempler, MacDonald, & Andersen, 1999; Almor, MacDonald, Kempler, Andersen, & Tyler, 2001; Glosser & Deser, 1991).

A number of diverse speech and language features have been identified and used to characterize fluent PPA and semantic dementia in general, and the SD variant of FTLD in particular. Gordon (Gordon, 2006) used a Quantitative Production Analysis (QPA) protocol (Berndt, Waylannd, 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 QPA 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. While these measures correlated with those obtained with the QPA protocol and were sensitive to aphasia severity, they also failed to distinguish between fluent and non-fluent 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 two 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 subject/patient’s performance on a set of language dimensions (e.g., speaking rate, hesitancy, speech sound distortions, telegraphic speech, grammaticality). The other approach consists manual psycho-linguistic analysis of speech and language samples in terms of their phonological, 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, while 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 (Artstein & Poesio, 2008; Krippendorff, 2004; Poesio & Vieira, 1998). While 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.

Neuropsychological tests and FTLD assessment

In addition to the analysis of spontaneous speech and language, research in FTLD has also relied on standard neuropsychological 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 (Gorno-Tempini, et al., 2004; J. Hodges, et al., 2004; Knopman, et al., 2008; Libon, et al., 2007). However, one of the issues with the use of standardized neuropsychological tests in FTLD research has to do with the lack of consistency in test selection and scoring across different clinical sites (Forman, et al., 2006). These and other studies indicate that standard neuropsychological tests are certainly capable of distinguishing between FTLD variants as well as between FTLD and AD. However, some studies found that the group differences identified with standard neuropsychological tests did not occur consistently across different tests. For example, Thompson et al. (Thompson, et al., 1997) suggested that neuropsychological 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 neuropsychological test scores in obtaining a more complete characterization of FTLD variants. Computerized approaches to speech and language analysis based on computational linguistics and natural language processing technology that we explore in the current study may help quantify some of these qualitative measures. For example, some of the phonological 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 one of the 3 FTLD syndromes (bvFTD, PNFA, SD) and PLA were recruited from 3 academic medical centers. All aspects of this study have been approved by the IRBs at each of the medical centers as well as the University of Minnesota. All 38 participants underwent a neuropsychological test battery that included the Boston Diagnostic Aphasia Examination Cookie-Theft Picture Description Task (Goodglass & Kaplan, 1983). This assessment was part of a larger study and minimizing subject burden was a key concern. The details of administering the neuropsychological test battery and the spontaneous speech elicitation procedures were previously reported (Knopman, et al., 2008). Picture descriptions of fifteen of the 38 participants were randomly selected to be manually assessed using the criteria detailed in the Manual psycholinguistic assessments section.

Diagnostic criteria: Diagnostic and exclusion criteria for this study have been previously reported (Knopman, et al., 2008). In brief, we defined 4 syndromes: behavioral variant frontotemporal dementia (bvFTD), progressive nonfluent aphasia (PNFA), progressive logopenic aphasia (PLA), and semantic dementia (SD). The inclusion in this study was based on Neary criteria (Neary, et al., 1998). In addition, all patients were required to have imaging studies demonstrating focal cerebral atrophy of at least one of the following: the anterior temporal lobes, frontal lobes, insula or caudate nuclei.

Progressive Non-fluent aphasia (PNFA) was diagnosed with expressive speech characterized by at least 3 of the following: nonfluency (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 fluent aphasia with anomia but intact word meaning and object recognition, where these symptoms constitute the principal deficits and the initial presentation.

*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.

*Behavioral variant fronto-temporal 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.

**Cognitive Instruments**

**Neuropsychological assessments:** All participants were administered a standard neuropsychological test battery consisting of the following tests: California Verbal Learning Test (CVLT) Free Recall(Delis, Kramer, Kaplan, & Ober, 2000), Number Cancellation(Mohs, et al., 1997), Digits Backward from Wechsler Memory Scale-Revised(Wechsler, 1987), Stroop Test(Stroop, 1935), Digit-Symbol Substitution(Wechsler, 1981), Verbal Fluency for Letters and Categories(Benton, Hamsher, & Sivan, 1983), Boston Naming Test(Kaplan, Goodglass, & Weintraub, 1978). 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 a previous publication(Knopman, et al., 2008).

**Computerized psycholinguistic assessments:** Natural Language Processing (NLP) and Automatic Speech Recognition (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 semi-automated language and speech analysis based on NLP and ASR technology (illustrated in Figure 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 sub-second level alignment (illustrated in Figure 2) enables precise measurement and quantification of durational and frequency characteristics of the input at the level of utterances (two or more co-articulated words), words as well as individual phonemes. While 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 annotations, we defined the following 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 (PCA) with Varimax rotation was used to examine the relationships between various semi-automated psycholinguistic measurements and to reduce the number of variables. One-way ANOVA was used to evaluate the differences in measurements using different FTLD variants as factors with Tukey’s 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 semi-automated vs. detailed speech and text alignments. All statistical calculations were carried out with SPSS 13.0 statistical software package.

Precision of semi-automated alignment and part-of-speech tagging

Measurements of speech and language characteristics used by our semi-automated approach rely on the alignment between the transcribed text of the picture description and the audio signal depicted in Figure 2. The precision of semi-automated alignment was estimated in terms of word beginning (WBBS) and word ending boundary shifts (WEBS) (Chen, Liu, Harper, Maia, & McRoy, 2004) as compared to 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 perform 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 performed subsequent to the verbatim transcription that was carried out for the semi-automated 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 performed by a linguist specifically trained for this task (DC).

Results

Reduction of variables to principal components

Based on the results of the exploratory factor analysis performed on all 38 samples and all 14 semi-automatic psycholinguistic variables, we identified four components that cumulatively accounted for 71% of the total variance in all variables. Based on 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 four FTLD variants (bvFTD, PNFA, PLA and SD) used as factors for several neuropsychological, and automatic psycholinguistic variables.

Neuropsychological assessments: Of all tests evaluated in this study, only two 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’s 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 as well as the PNFA groups were highly significant (p<0.0001). The difference between PLA and SD was also significant (p = 0.029).

Semi-automated 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 three measurements is presented in Figure 3, Figure 4 and Figure 5, respectively. Tukey’s 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 three 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 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.

Precision of semi-automated alignment and part-of-speech tagging

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

Discussion
Our computerized approach to quantifying language use characteristics obtained from spontaneous speech removes the subjectivity inherent in manual assessments and thus may improve the reliability as well as the comparability of measurements across different studies. Current computer-based neuropsychological 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 as well as open up a wide range of new possibilities for neuropsychological testing (Letz, 2003). 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 semi-automated 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 neuropsychological testing) to perform tasks that are currently beyond the reach of computers (e.g., 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 (e.g., 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 as well as 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 (Roark, et al., 2007) and aphasia in children (Hosom, et al., 2004). We found that speech hesitancy characterized by the ratio of silent pauses to words as well as the ratio of disfluent events to words are sensitive indicators of PNFA. Fluency of spontaneous speech has been previously found to be significantly decreased in PNFA patients as compared to controls and other FTLD groups (Ash, et al., 2008). The same study found that speech fluency was generally decreased in all FTLD variants as compared to healthy controls with partially overlapping neuroanatomical sources associated with fluency.

Two clinical measures (verbal category fluency and correctness on the Boston Naming Test) and two semi-automated psycholinguistic measures (pronoun-to-noun ratio and pause-to-word ratio) showed significant differences between the means among the four 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 (Bird, et al., 2000) and FTLD (Knopman, et al., 2008; Libon, et al., 2007). 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. While these results are in keeping with some of the previous studies (Clark, Charuvastra, Miller, Shapira, & Mendez, 2005; Nestor, et al., 2003), other studies have found slightly more impaired performance on this test in the PNFA group (Libon, et al., 2009) possibly due 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 to the PLA (7.75) and SD
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 previous 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 semantic dementia and anoma. Both of these tests fail to distinguish between the behavioral and the non-fluent 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 four FTLD variants.

Our findings for the semi-automated 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, while the latter condition involves difficulties with speech production but not necessarily picture naming. In previous studies, patients with SD variant of FTLD were found to be significantly more impaired on a picture naming test as compared to the PNFA and bvFTD variants (Libon, et al., 2009; Nestor, et al., 2003). Patients with PNFA also produced more errors on the BNT test than healthy controls; however, these errors were predominantly phonological in nature suggesting intact semantic store in this group (Nestor, et al., 2003). Overuse of pronouns (a.k.a. “empty” speech) is a prominent feature of progressive aphasia found in later stages of Alzheimer’s disease (Almor, et al., 1999; Kempler, 1995) 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 (Bird, et al., 2000; Hillis, Oh, & Ken, 2004; Price & Grossman, 2005; Silveri, Salvigni, Cappa, Della Vedova, & Puopolo, 2003).
Similarly 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 sowed 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 non-fluent progressive aphasia variants (Thompson, et al., 1997).

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 audiorecordings of picture descriptions by these patients consisted of silence. These findings are also consistent with previous studies of non-fluent progressive aphasias and the PNFA variant of FTLD (Ash, et al., 2008; Libon, et al., 2009; Nestor, et al., 2003) that showed decreased performance in this group on the letter fluency test as well as spontaneous speech fluency assessments (Ash, et al., 2008; Gorno-Tempini, et al., 2004; Grossman, 2002; Peelle, et al., 2007; Weintraub, et al., 1990).

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

The precision of word-initial and word-final boundary semi-automated alignment was better that 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. (Chen, et al., 2004) 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 pre-segmented 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 where 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 previously reported results. Brants had previously 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 (Brants, 2001). Our results show that this tagger is 84% accurate which is likely due 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. While the accuracy of 84% is good (approximately one out of ten 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 progressive nonfluent aphasia, semantic dementia and progressive logopenic aphasia 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 progressive nonfluent aphasia and the TDP-43 proteinopathy associated with semantic dementia. While
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 semi-automated approach to language and speech analysis resulted in some loss of precision in both time alignments and part-of-speech identification. The former is due to the quality of the available audio, while the latter is likely due 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)(Brants, 2001). Re-training 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 semi-automatic alignments were within 80 ms at word onsets and 230 ms and 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 intra-utterance from inter-utterance 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 semi-automated. 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 automatic speech recognition engine that will perform 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 system’s accuracy. The present study lays the foundation for these future investigations.

**Acknowledgements**

Our work 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.
References


Table 1 Rotated component matrix obtained with principal component analysis of the semi-automated psycholinguistic measurements on all 38 picture description samples

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pause-to-word ratio</td>
<td>-.148</td>
<td>.803</td>
<td>.132</td>
<td>-.020</td>
</tr>
<tr>
<td>Fundamental Frequency Variance</td>
<td>.042</td>
<td>-.162</td>
<td>-.075</td>
<td>.798</td>
</tr>
<tr>
<td>POS Perplexity</td>
<td>-.041</td>
<td>.380</td>
<td>.279</td>
<td>.726</td>
</tr>
<tr>
<td>Word Perplexity</td>
<td>-.407</td>
<td>.529</td>
<td>.024</td>
<td>.213</td>
</tr>
<tr>
<td>Pronoun-to-noun ratio</td>
<td>.108</td>
<td>-.375</td>
<td>.729</td>
<td>.223</td>
</tr>
<tr>
<td>Word Count</td>
<td>.932</td>
<td>.002</td>
<td>.000</td>
<td>.126</td>
</tr>
<tr>
<td>Speech Duration (msec.)</td>
<td>.864</td>
<td>-.063</td>
<td>-.027</td>
<td>.245</td>
</tr>
<tr>
<td>Mean Prosodic Phrase Length</td>
<td>-.498</td>
<td>-.550</td>
<td>-.013</td>
<td>.006</td>
</tr>
<tr>
<td>Correct Information Unit Count</td>
<td>.726</td>
<td>.053</td>
<td>-.390</td>
<td>-.217</td>
</tr>
<tr>
<td>Long Pause Count</td>
<td>-.322</td>
<td>.288</td>
<td>.844</td>
<td>-.135</td>
</tr>
<tr>
<td>Filled Pause Count</td>
<td>.195</td>
<td>.651</td>
<td>.143</td>
<td>.206</td>
</tr>
<tr>
<td>Pause Count</td>
<td>-.182</td>
<td>.426</td>
<td>.830</td>
<td>-.158</td>
</tr>
<tr>
<td>False Start Count</td>
<td>.329</td>
<td>.368</td>
<td>.427</td>
<td>.248</td>
</tr>
<tr>
<td>Pause-to-word ratio</td>
<td>.181</td>
<td>.403</td>
<td>-.091</td>
<td>.650</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Rotation converged in 7 iterations.

Rotation converged in 7 iterations.
b Coefficients in bold represent the items used in subjective labeling of components with values exceeding 0.6
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 four FTLD variants. Group means by diagnosis are indicated with (x) with mean values at the bottom of each boxplot.
Figure 4. Normalized dysfluent event means in four FTLD variants. Group means by diagnosis are indicated with (×) with mean values at the bottom of each boxplot.
Figure 5 Pronoun-to-noun ratio means in four FTLD variants. Group means by diagnosis are indicated with (x) with mean values at the bottom of each boxplot.
Appendix

A. Automatic alignment

The ASR engine used for alignment was based on the Hidden Markov Model Toolkit (HTK 3.4) developed at Cambridge University Labs, UK (Young, Kershaw, Odell, Ollason, & Valtchev, 2006). 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 (Heeman & Allen, 1999).

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 non-optional 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 one 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”. Since this word is not found in a standard English dictionary, its pronunciation was automatically constructed on-the-fly resulting in a simplified [consonant-vowel-consonant-vowel] representation where consonants represented by phoneme [t] and vowels were represented by the mid-central [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. While this method does not allow us to access the content of the word (phonological or semantic), it does allow us to identify the speech/silence boundaries necessary to compute prosodic measurements such as the pause-to-word ratio.

**B. Semi-automated psycholinguistic measurements**

1. **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 milliseconds. 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 (Forrest, Weismer, & Turner, 1989; Levelt, 1989). All non-silent segments with the exception of filled pauses (um’s and ah’s) were treated as words in calculating this measure.

2. **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 non-silent 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 (Boersma, 2001).

3. **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 one, two, three, or more preceding parts-of-speech. Such part-of-speech language model captures the notion of grammaticality where, in normal spoken English, the likelihood of seeing an adjective immediately following a noun (e.g., 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 comprises 2,430 six minute conversations between 500 male and female speakers from all major American dialect groups (Godfrey & Holliman, 1997). 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 (Brants, 2001). 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

4. Word-level perplexity
This measure captures the perplexity of a language model constructed the same way as in 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 (Bahl, Baker, Jelinek, & Mercer, 1977). 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.

5. 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.

6. Word count

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

7. Total duration of speech in the sample
This measure was computed by adding up the durations of all non-silent elements excluding filled pauses, noises and false starts in the audio alignment of the picture description.

8. 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 (Levelt, 1989). To identify changes in speech rhythm, we followed the methodology proposed by Wightman and Ostendorf (Wightman & Ostendorf, 1991) that relies on calculating the average normalized duration of phonemes. Wightman and Ostendorf (Wightman & Ostendorf, 1991) used the difference in the average normalized duration of the last three syllables in the coda of the preceding words from the first three syllables in the onset of the word following 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 one 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.

9. Correct Information Unit count
We developed an automated approach to counting correct information units (CIU) as
defined by Brookshire and Nicholas (Brookshire & Nicholas, 1994; Nicholas & Brookshire, 1995). 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 pre-defined list of word sequences representing CIUs. The list of CIUs and word sequences was compiled using Yorkston and Beukelman’s (Yorkston & Beukelman, 1980) list of concept units as a starting point. The list was expanded to include lexical and morphological word variants (e.g. falling, fall, fell). The complete list used in this study is provided in Appendix C.

10. Normalized long pause count
This variable represents a normalized count of all silent pauses greater than 400 milliseconds in duration. This measure as well as the following measures in 11-14 were normalized by dividing their value by the total length of speech in the sample.

11. Normalized filled pause count
Our system distinguishes between two types of pause fillers – the shorter ones without nasalization (e.g., [ah], [eh]) and the longer ones with nasalization (e.g., [uhm]); however, for this study we collapsed the two filled pause types into one.

12. Normalized silent pause count
Only silences greater than 150 milliseconds in duration were counted as silent pauses as described earlier.

13. Normalized false start count
This is a normalized count of false starts where the speaker begins to speak a word but does not finish the word.
14. Normalized repetition count

Sequences of two and/or three words adjacent to each other were counted as repetitions.

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

This is a combined normalized count of all filled pauses, repetitions and false starts.
C. List of Information Units and their variants

<table>
<thead>
<tr>
<th>English</th>
<th>Variant 1</th>
<th>Variant 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>two</td>
<td>for his sister</td>
<td>doing</td>
</tr>
<tr>
<td>children</td>
<td>pouring</td>
<td>dishes</td>
</tr>
<tr>
<td>little</td>
<td>from the jar</td>
<td>on the counter</td>
</tr>
<tr>
<td>boy</td>
<td>cookie jar</td>
<td>drying</td>
</tr>
<tr>
<td>girl</td>
<td>on the high shelf</td>
<td>faucet is on</td>
</tr>
<tr>
<td>sister</td>
<td>on the top shelf</td>
<td>full blast</td>
</tr>
<tr>
<td>brother</td>
<td>on the shelf</td>
<td>ignoring</td>
</tr>
<tr>
<td>kid</td>
<td>from the high shelf</td>
<td>daydreaming</td>
</tr>
<tr>
<td>kids</td>
<td>from the top shelf</td>
<td>oblivious</td>
</tr>
<tr>
<td>standing</td>
<td>from the shelf</td>
<td>paying attention</td>
</tr>
<tr>
<td>by the boy</td>
<td>in the cupboard</td>
<td>water</td>
</tr>
<tr>
<td>by her brother</td>
<td>with the open door</td>
<td>overflowing</td>
</tr>
<tr>
<td>near the boy</td>
<td>handing to sister</td>
<td>onto the floor</td>
</tr>
<tr>
<td>near her brother</td>
<td>handing to his sister</td>
<td>on the floor</td>
</tr>
<tr>
<td>reaching up</td>
<td>handing cookies</td>
<td>on the shoes</td>
</tr>
<tr>
<td>reach up</td>
<td>to his sister</td>
<td>on her shoes</td>
</tr>
<tr>
<td>reach</td>
<td>to the sister</td>
<td>onto the shoes</td>
</tr>
<tr>
<td>on stool</td>
<td>asking for cookie</td>
<td>onto her shoes</td>
</tr>
<tr>
<td>wobbling</td>
<td>asking for a cookie</td>
<td>on her feet</td>
</tr>
<tr>
<td>off balance</td>
<td>asking for cookies</td>
<td>onto her feet</td>
</tr>
<tr>
<td>three legged</td>
<td>trying to get</td>
<td>feet are getting wet</td>
</tr>
<tr>
<td>3-legged</td>
<td>ask for cookie</td>
<td>getting wet</td>
</tr>
<tr>
<td>fall over</td>
<td>ask for a cookie</td>
<td>dirty dishes left</td>
</tr>
<tr>
<td>fall off</td>
<td>ask for cookies</td>
<td>puddle</td>
</tr>
<tr>
<td>falling over</td>
<td>has finger to mouth</td>
<td>in the kitchen</td>
</tr>
<tr>
<td>falling off</td>
<td>finger to mouth</td>
<td>indoors</td>
</tr>
<tr>
<td>tip over</td>
<td>finger to her mouth</td>
<td>inside</td>
</tr>
<tr>
<td>tipping over</td>
<td>pressed to her mouth</td>
<td>outside</td>
</tr>
<tr>
<td>by boy</td>
<td>saying shhh</td>
<td>disaster</td>
</tr>
<tr>
<td>on the floor</td>
<td>keeping him quiet</td>
<td>lawn</td>
</tr>
<tr>
<td>hurt himself</td>
<td>keeping the boy quiet</td>
<td>road</td>
</tr>
<tr>
<td>reach up</td>
<td>trying to help</td>
<td>path</td>
</tr>
<tr>
<td>reaching up</td>
<td>laugh</td>
<td>tree</td>
</tr>
<tr>
<td>reaching</td>
<td>laughing</td>
<td>trees</td>
</tr>
<tr>
<td>reach</td>
<td>mother</td>
<td>driveway</td>
</tr>
<tr>
<td>taking</td>
<td>woman</td>
<td>house</td>
</tr>
<tr>
<td>take</td>
<td>lady</td>
<td>next door</td>
</tr>
<tr>
<td>took</td>
<td>children behind her</td>
<td>garage</td>
</tr>
<tr>
<td>stealing</td>
<td>children behind the</td>
<td>open window</td>
</tr>
<tr>
<td>pull</td>
<td>woman</td>
<td>window is open</td>
</tr>
<tr>
<td>pulling</td>
<td>children behind the</td>
<td>curtains</td>
</tr>
<tr>
<td>stole</td>
<td>mother</td>
<td>drapes</td>
</tr>
<tr>
<td>steal</td>
<td>behind the mother</td>
<td>draperies</td>
</tr>
<tr>
<td>cookie</td>
<td>standing</td>
<td>plate</td>
</tr>
<tr>
<td>cookies</td>
<td>by the sink</td>
<td>towel</td>
</tr>
<tr>
<td>for himself</td>
<td>washing</td>
<td></td>
</tr>
</tbody>
</table>