Journal of Alzheimer's Disease 49 (2016) 407–422
DOI 10.3233/JAD-150520
IOS Press

Linguistic Features Identify Alzheimer's Disease in Narrative Speech

Kathleen C. Fraser[a], Jed A. Meltzer[b] and Frank Rudzicz[a,c,*]
[a]*Department of Computer Science, University of Toronto, Toronto, Canada*
[b]*Rotman Research Institute, Toronto, Canada*
[c]*Toronto Rehabilitation Institute-UHN, Toronto, Canada*

Handling Associate Editor: Peter Garrard

Accepted 20 August 2015

Abstract.
Background: Although memory impairment is the main symptom of Alzheimer's disease (AD), language impairment can be an important marker. Relatively few studies of language in AD quantify the impairments in connected speech using computational techniques.
Objective: We aim to demonstrate state-of-the-art accuracy in automatically identifying Alzheimer's disease from short narrative samples elicited with a picture description task, and to uncover the salient linguistic factors with a statistical factor analysis.
Methods: Data are derived from the DementiaBank corpus, from which 167 patients diagnosed with "possible" or "probable" AD provide 240 narrative samples, and 97 controls provide an additional 233. We compute a number of linguistic variables from the transcripts, and acoustic variables from the associated audio files, and use these variables to train a machine learning classifier to distinguish between participants with AD and healthy controls. To examine the degree of heterogeneity of linguistic impairments in AD, we follow an exploratory factor analysis on these measures of speech and language with an oblique promax rotation, and provide interpretation for the resulting factors.
Results: We obtain state-of-the-art classification accuracies of over 81% in distinguishing individuals with AD from those without based on short samples of their language on a picture description task. Four clear factors emerge: semantic impairment, acoustic abnormality, syntactic impairment, and information impairment.
Conclusion: Modern machine learning and linguistic analysis will be increasingly useful in assessment and clustering of suspected AD.

Keywords: Automatic data processing, factor analysis, geriatric assessment, heterogeneity, language, statistical

INTRODUCTION

Our aging society faces the rising incidence of Alzheimer's disease (AD). In the absence of a cure, researchers have mitigated the disease's impact through better and earlier diagnosis, and effective symptomatic treatment. Spoken language is a rich source of information on an individual's cognitive status. Unfortunately, the utility of naturalistic spoken language as a quantitative measure has been limited, due to the time-consuming nature of manual analysis and human subjectivity. Recent progress in computational linguistics has brought powerful tools to bear, showing that fully automated analyses of speech and language can reliably distinguish patients with dementia from controls [1–4], and can differentiate between specific forms of dementia, such as early versus advanced stages of AD [5] and non-fluent progressive aphasia versus semantic dementia [6].

Although memory impairment due to medial temporal lobe damage is the characteristic symptom of AD,

*Correspondence to: Frank Rudzicz, 550 University Avenue, Room 12-175, Toronto, Ontario M5G 2A2, Canada. Tel.: +1 416 597 3422/Ext. 7971; E-mail: frank@cs.toronto.edu.

ISSN 1387-2877/16/$35.00

language problems are also prevalent. Patients with AD frequently exhibit naming deficits, and all aspects of language become affected as the disease progresses; typically starting with semantics before proceeding to syntax and phonology [7]. Since lesions confined to the medial temporal lobe are typically associated only with mild impairment of high-level language [8], the deterioration of distinct micro-linguistic aspects of language is most likely attributable to the spread of pathology throughout the cortex. Cortical involvement in AD is highly variable across individuals and certain distinct variants of AD have been described, differing in both age-of-onset and cortical atrophy [9–11]. Among these variants is logopenic progressive aphasia, characterized by left parietal damage and impairments in phonology and repetition [12, 13]. Other distinct dementia syndromes, such as progressive non-fluent aphasia and semantic dementia, show clear dissociations between phonological, syntactic, and semantic impairments related to the location of cortical pathology, although these syndromes are caused by different pathologies than AD [14].

Given that cortical damage is heterogeneous in AD as the disease progresses, both within and across individuals, it is expected that linguistic symptomatology would be similarly heterogeneous.

We have two aims. First, we test classification accuracy given automated analysis of narrative speech samples from a picture description task using machine learning methods. Second, we explore the heterogeneity of linguistic impairment among participants with AD. Using exploratory factor analysis, we extract four factors that account for the majority of the variance among speakers. An oblique factor rotation method (promax) is used, as this method allows the underlying factors to be correlated, rather than guaranteeing orthogonality. Thus, we are able to identify the degree to which linguistic abilities decline in parallel through dementia, as opposed to varying independently between subjects that experience varying extents of cortical degeneration.

Background

Although memory impairment is the main symptom of AD, language impairment can be an important marker. Faber-Langendoen et al. [15] found that 36% of mild AD patients and 100% of severe AD patients had aphasia, according to standard aphasia testing protocols. Ahmed et al. [16] found that two-thirds of their participants showed subtle, but significant, changes in connected speech production up to a year before their diagnosis of probable AD. Weiner et al. [17], in a study of 486 AD patients, reported a significant correlation between dementia severity and a number of different linguistic measures, including confrontation naming, articulation, word-finding ability, and semantic fluency.

Declining performance on naming tasks can occur early in the disease progression [7, 18–20]. Krishner et al. [19] found that all the AD participants in their study were impaired on a naming task, even when their language functioning was normal by other measures. Individuals with AD can have difficulty retrieving the names of people and places [21], and may substitute generic terms for more specific ones [7, 18, 22]. Numerous studies have reported a greater impairment in category naming fluency (e.g., naming animals or tools) relative to letter naming fluency (e.g., naming words that start with the letter R) [23–25], and this finding was supported by a meta-analysis of 153 studies [26]. There is also some evidence that patients with AD may have more difficulty naming verbs than nouns. Robinson et al. [27] found that AD participants performed worse on a picture-naming task for verbs than nouns, even when the verbs and nouns were spelled and pronounced the same, and matched for frequency. As a result of word-finding difficulties and a reduction in working vocabulary, the language of AD patients can seem 'empty' [28, 29], and was described by Appell et al. [30] as "verbose and circuitous, running on with a semblance of fluency, yet incomplete and lacking coherence."

Macro-linguistic language functions, such as understanding metaphor and sarcasm, also tend to deteriorate in AD [31]. Thematic coherence, or the ability to maintain a theme throughout a discourse, may also be impaired. In a study comparing 9 AD participants to healthy controls and participants with fluent aphasia, Glosser and Deser [32] found that the AD participants showed a reduction in global coherence in a structured interview task. Blonder et al. [33] reported a similar result when interviewing five AD participants and their healthy spouses.

The effect of AD on syntax is controversial. Some researchers have reported syntactic impairments in AD, while others claim that any apparent deficits are in fact due to difficulties with memory and semantics [18]. Several studies have found evidence for a decrease in the syntactic complexity of language in AD [34–36]. Croisile et al. [34] compared oral and written picture descriptions from 22 AD patients and matched controls, and found that the AD patients produced fewer subordinate clauses than controls. Ehrlich et al. [35] reported a reduced utterance length on narrative tasks administered to 16 AD participants and controls. In a study comparing language production in AD and semantic dementia, Sajjadi et al. [36] found that their 20 patients with mild AD tended to produce fewer complex syntactic units on both a picture description task and an interview. On the other hand, Kempler et al. [37] found that 10 individuals with AD used a range of syntactic constructions with the same frequency as control participants in spontaneous conversation, despite showing signs of lexical impairment. Glosser and Deser [32] similarly did not find any difference in syntactic complexity or correctness between AD patients and controls in spontaneous speech.

There is evidence that language decline in AD is heterogeneous. Hodges and Patterson [38] divided 52 AD patients into three different categories based on dementia severity and assessed their semantic impairment on a number of different tasks. They reported a wide range of performance in the "minimal" and "mild" AD groups. Duong et al. [39] had 46 AD participants produce narratives based on a single picture and a series of pictures. A cluster analysis subsequently revealed a number of different discourse patterns rather than a single characteristic pattern of impairment. Ahmed et al. [16] contrasted their findings of heterogeneous language decline in connected speech from 15 AD patients with the more predictable patterns of decline seen in primary progressive aphasia (PPA).

Related computational work

A relatively small subset of studies on language in AD attempt to quantify the impairments in connected speech using computational techniques. Bucks et al. [1] conducted a linear discriminant analysis of spontaneous speech from 8 AD participants and 16 healthy controls. They considered eight linguistic features, including part-of-speech (POS) tag frequencies and measures of lexical diversity, and obtained a cross-validation accuracy of 87.5%.

Thomas et al. [5] classified spontaneous speech samples from 95 AD patients and an unspecified number of controls by treating the problem as an authorship attribution task, and employing a "common *N*-grams" approach. They were able to distinguish between patients with severe AD and controls with a best accuracy of 94.5%, and between patients with mild AD and controls with a best accuracy of 75.3%. They suggested that closed-class words were particularly informative in their analysis.

Guinn and Habash [3] built classifiers to distinguish between AD and non-AD language samples using 80 conversations between 31 AD patients and 57 cognitively normal conversation partners. They found that features such as POS tags and measures of lexical diversity were less useful than measuring filled pauses, repetitions, and incomplete words, and achieved a best accuracy of 79.5%.

Meilán et al. [2] distinguished between 30 AD patients and 36 healthy controls with temporal and acoustic features alone, obtaining an accuracy of 84.8%. For each participant, their speech sample consisted of two sentences read from a screen. The five most discriminating features were percentage of voice breaks, number of voice breaks, number of periods of voice, shimmer, and noise-to-harmonics ratio.

Jarrold et al. [4] used acoustic features, POS features, and psychologically-motivated word lists to distinguish between semi-structured interview responses from 9 AD participants and 9 controls with an accuracy of 88%. They also confirmed their hypothesis that AD patients would use more pronouns, verbs, and adjectives and fewer nouns than controls.

Rentoumi et al. [40] considered a slightly different problem: they used computational techniques to differentiate between picture descriptions from AD participants with and without additional vascular pathology ($n = 18$ for each group). They achieved an accuracy of up to 75% when they included frequency unigrams and excluded binary unigrams, syntactic complexity features, measures of vocabulary richness, and information theoretic features.

Orimaye et al. [41] obtained F-measure scores up to 0.74 using a relatively restricted feature set on transcripts from DementiaBank, although they combined participants with different etiologies, rather than focusing on AD. Other related work has considered the automatic analysis of language in other types of dementia, including semantic dementia [42], mild cognitive impairment (MCI) [43, 44], and PPA [6].

Our study differs from previous work in several ways. Along with Orimaye et al. [41], we consider a much larger sample size than most previous work, giving a more representative sample for machine learning. We also consider a larger number of features to help capture the array of different language impairments that can be seen in AD, and conduct factor analysis to characterize patterns of heterogeneity.

MATERIALS AND METHODS

Our data are derived from the DementiaBank corpus[1], which is part of the larger TalkBank project [45]. These data were collected between 1983 and 1988 as part of the Alzheimer Research Program at the University of Pittsburgh. Information about the study cohort is available from Becker et al. [46]. Participants were referred directly from the Benedum Geriatric Center at the University of Pittsburgh Medical Center, and others were recruited through the Allegheny County Medical Society, local neurologists and psychiatrists, and public service messages on local media. To be eligible for inclusion in the study, all participants were required to be above 44 years of age, have at least 7 years of education, have no history of nervous system disorders or be taking neuroleptic medication, have an initial Mini-Mental State Exam (MMSE) score of 10 or greater, and be able to give informed consent. Additionally, participants with dementia were required to have a relative or caregiver to act as an informant. All participants received an extensive neuropsychological and physical assessment (see [46] for complete details). Participants were assigned to the "patient" group primarily based on a history of cognitive and functional decline, and the results of a mental status examination. In 1992, several years after the study had ended, the final diagnosis of each patient was reviewed on the basis of their clinical record and any additional relevant information (in some cases, autopsy).

From the "Dementia" group, we include participants with a diagnosis of "possible AD" or "probable AD", resulting in 240 samples from 167 participants. We also include control participants, resulting in 233 additional files from 97 speakers. Demographics are given in Table 1. We compute averages over individual sessions instead of individual participants in order to capture intra-speaker variation over the five years these data were collected. The two groups are not matched for age and education, which is one limitation of these data.

Narrative speech was elicited using the "Cookie Theft" picture description task from the Boston Diagnostic Aphasia Examination [47]. This protocol instructs the examiner to show the picture to the patient and say, "*Tell me everything you see going on in this picture*." The examiner is permitted to encourage the patient to keep going if they do not produce very many words. Each speech sample was recorded then manually transcribed at the word level following the TalkBank CHAT (Codes for the Human Analysis of Transcripts) protocol [48]. Narratives were segmented into utterances and annotated with filled pauses, paraphasias, and unintelligible words.

Table 1
Demographics of DementiaBank data

	AD ($n = 240$)	Control ($n = 233$)
Age (years)	71.8 (8.5)	65.2 (7.8)
Education (years)	12.5 (2.9)	14.1 (2.4)
Gender (male/female)	82/158	82/151
Mini-Mental State Exam	18.5 (5.1)	29.1 (1.1)

From the CHAT transcripts, we keep only the word-level transcription and the utterance segmentation. We discard the morphological analysis, dysfluency annotations, and other associated information, as our goal is to create a fully automated system that does not require the input of a human annotator. Before tagging and parsing the transcripts, we automatically remove short false starts consisting of two letters or fewer (e.g., *The c-cookie jar* would become *The cookie jar*) and filled pauses such as *uh*, *um*, *er*, and *ah* (e.g., *The um um boy* would become *The boy*). All other dysfluencies (including repetitions, revisions, paraphasias, and comments about the task) remain in the transcript. The AD participants produce an average of 104.3 (SD: 59.0) words per narrative, while the control participants produce an average of 114.4 (SD: 59.5) words per narrative, although the distribution in both cases is somewhat right-skewed.

Each transcript has an associated audio file, allowing for lexical and acoustic analysis in parallel, which we convert from MP3 to 16-bit mono WAV format with a sampling rate of 16 kHz.

Features

We consider a large number (370) of features to capture a wide range of linguistic phenomena. Here we provide a brief description of the different feature types; a full explanation of each feature can be found in the Supplementary Material.

Part-of-speech

Some language changes in AD may be detected by measuring the production of different POS. For example, Ahmed et al. [16] reported changes in the number of pronouns and verbs, and Bucks et al. [1] and Jarrold et al. [4] reported a decrease in the proportion of nouns and an increase in the proportion of pronouns, adjectives, and verbs. We extract POS

[1]Downloaded November 22, 2013.

information automatically using the Stanford tagger[2]. We compute the frequency of occurrence of different POS, normalized by the total number of words in each utterance. We also compute ratios, e.g., nouns to verbs, pronouns to nouns. As described in the previous section, we remove the filled pauses *uh*, *um*, *er*, and *ah*, but we keep track of the frequency of occurrence of each type, as previous work has suggested that they may serve different purposes [49]. We also tag words which do not appear in the English dictionary (i.e., "not in dictionary", or NID), which include paraphasias (e.g., *the sink's overflown*) or word fragments that were not removed in the pre-processing phase (e.g., *In her kitch-kitchen I suppose*). Paraphasias or fragmented words that result in the unintentional production of another English word are not detected by our system.

Although we do not consider the manually annotated POS tags from the DementiaBank corpus in our analysis, we can use that information to test the performance of the Stanford tagger on this data. The task is complicated by the fact that the Stanford tagger and the CHAT protocol use different tagsets, the CHAT protocol does not include dysfluencies in the morphological analysis, and the DementiaBank corpus includes some user-defined tags. With those caveats in mind, we find the Stanford tagger has an accuracy of 85.4% on the control data, and 84.8% on the AD data (over the entire data sample).

Syntactic complexity

As discussed above, the degree to which syntactic complexity is affected in AD is uncertain. We measure the syntactic complexity of the picture descriptions using a number of well-known metrics, including the mean length of sentences, T-units, and clauses, and by calculating scores based on the parse tree, including the height of the tree and the mean, total, and maximum Yngve depth (a measure of embeddedness) [50]. Parse trees are computed using the Stanford parser[3].

Grammatical constituents

We quantify the constituents comprising the parse tree, in a set of 'context-free grammar' (CFG) features. This allows us to explore possible syntactic differences in AD in greater detail than the more general syntactic complexity metrics. Previous work has shown that these features can distinguish between participants with agrammatic aphasia and matched controls on a story-telling task [51]. To calculate these features, we compute the frequency of occurrence of different grammatical constituents, normalized by the total number of constituents in the sample. For example, we can count the number of noun phrases (NP) that consist of a determiner (DT) and a noun (NN) (i.e., NP $\longrightarrow$ DT NN), a noun without a determiner (NP $\longrightarrow$ NN), a pronoun (NP $\longrightarrow$ PRP), etc. The symbols used here are from the Penn Treebank tagset[4]. We also consider the rate, proportion, and average length of noun phrases, verb phrases (VP), and prepositional phrases (PP). This is based on work by Chae and Nenkova [52], except that rather than assess these features for each utterance in the narrative, we compute them for the entire narrative.

Psycholinguistics

A semantic impairment may manifest in an increased reliance on highly familiar words. We therefore rate each word on existing psycholinguistic norms. Specifically, we use the SUBTL frequency norms [53] and the combined Bristol and Gilhooly-Logie norms [54, 55] for familiarity, imageability, and age-of-acquisition. We compute the average of each of these for all content words, as well as for nouns and verbs separately.

Vocabulary richness

We assess the vocabulary richness (i.e., lexical diversity) of a narrative sample using a number of different metrics, including type-token ratio, moving-average type-token ratio, Brunet's index, and Honoré's statistic. Type-token ratio (TTR) is widely reported, but has also been criticized for its dependence on text length [56, 57]. Moving-average type-token ratio (MATTR) was proposed by Covington and McFall [58] as an adaptation of TTR that is independent of text length. In a study on language in aphasia, MATTR was reported to be one of the best metrics for providing an unbiased metric of lexical diversity [59]. Brunet's index and Honoré's statistic are alternative measures of vocabulary richness that have been used in previous computational studies of AD [1, 3, 5].

[2] Version 2015-01-29 http://nlp.stanford.edu/software/tagger.shtml.

[3] Version 2010-11-30, nlp.stanford.edu/software/lex-parser.shtml.

[4] For the full tag list, see: http://www.comp.leeds.ac.uk/amalgam/tagsets/upenn.html.

Information content

Previous studies of AD narratives in picture description tasks have reported decreased information content [16, 34, 36, 60, 61]. We measure this computationally by searching for relevant lexical items that point to each of the expected information units listed in Croisile et al. [34]. For example, the occurrence of the word *boy*, *son*, or *brother* all suggest that the information unit "boy" has been mentioned. Our information unit features are binary: either a word related to a given information unit has been mentioned, or not. For information units describing actions, such as "the boy or stool falling", we use the dependency structure from the Stanford parser to locate phrases with *fall* as the verb and *boy* or *stool* as the subject. There are obvious limitations to this method: if a word is used in the incorrect context (e.g., if the speaker refers to the girl as *the mother*), then it would be applied to the wrong information unit, or if a speaker refers to a concept in an unpredictable way (e.g., refers to the woman at the sink as *the nurse*), then that information unit will not be counted. However, this method provides a simple, automated estimate of informativeness. A similar method was used by Pakhomov et al. [62] to automatically score picture descriptions from participants with frontotemporal lobar degeneration, and by Hakkani-Tür et al. [63] to score picture descriptions from elderly, cognitively healthy speakers. A certified speech-language pathologist also annotated these information units over a random 5% of the data and we compared these against the automatically identified units. There was an observed agreement of 98.02%, which corresponds to a Cohen's kappa coefficient of 0.8037, which is the maximum possible given observed marginal frequencies ($p < 0.05$).

We also measure the frequency of specific words relevant to the Cookie Theft picture. For example, *mother* and *woman* both provide evidence for the "woman" information unit, as described above, but by counting the frequency of occurrence of the two words separately we can also detect if one participant group is more likely than the other to refer to the woman as a mother. We call these features "key words", and they are integer-valued frequency counts. Previous work (e.g., [40, 42]) has shown the utility of simple binary and frequency unigrams (word tokens). Rather than considering the space of all possible unigrams, we have considered only a smaller set which we have deemed to be relevant to the expected information content, to avoid problems of data sparsity, and to help improve the interpretability of the selected features.

Repetitiveness

AD patients can exhibit perseverative behavior in different areas of their lives, including in their language [64, 65]. Nicholas et al. [28] found that in a picture description task, AD patients repeated words and phrases more frequently than healthy controls and also more frequently than participants with fluent aphasia. Tomoeda et al. [65] found also that AD patients were more likely to repeat ideas in a picture description task than healthy controls, and that the frequency of repetitions was not related to severity of dementia.

Using a bag-of-words, we measure the cosine distance between each pair of utterances in the session. We remove a short list of stopwords, after observing that utterances such as *He is standing on the stool* and *He is holding the cookie* could be considered relatively similar given the common occurrences of *he*, *is*, and *the*. A distance of zero between two utterances indicates that the two utterances are identical (in word counts, not necessarily order). We detect the occurrence of repetitive content by measuring average distance, and the proportion of utterance pairs that fall below some threshold.

Acoustics

Several acoustic features are extracted from the audio. We include a number of features which are indicative of pathological speech [2, 43, 62, 66]. We also include a set of features based on Mel-frequency cepstral coefficients (MFCCs), which are nearly ubiquitously used in speech recognition research [67]. MFCCs, in some sense, encode the 'spectrum of the spectrum' in that they are the discrete cosine transform of logarithms of spectral power; this separates the source of a signal (i.e., the energy of the lungs) from the 'filter' of the signal (i.e., the upper vocal tract, in which phonemes are differentiated phonologically). The aforementioned spectrum is also mapped into the Mel scale, which approximates the sensitivity of the human ear. Individual coefficients are strictly components of a de-convolution process, but they can be *loosely* associated with articulatory phenomena in the sense that low indices correspond to lower resonances in the vocal tract. In order to compare differences across speakers, we consider the mean, variance, skewness, and kurtosis of the first 42 MFCCs through time, and the kurtosis and skewness of the means across each MFCC dimension.

Identification of AD by machine learning

Machine learning is a subfield of artificial intelligence in which statistical models are constructed from data automatically. Typically, this involves iterative refinement of models with the aim of increasing their overall accuracy. We also perform multilinear logistic regression with nominal outputs, allowing for interactions between categories and coefficients.

Our evaluation criterion is *accuracy*, which is the ratio of true positives plus true negatives over all evaluation examples. We have considered training models that identify AD based solely on textual or acoustic features, separately. However, as we have demonstrated in [69], allowing feature selection to choose from all available measures provides uniformly higher accuracy than restricting the set to only textual or acoustic features. This may be anticipated, as it lessens the expected redundancy between selected features in any single modality. As discussed in the Results, those features are a fairly balanced set of textual and acoustic ones.

We perform a 10-fold cross-validation procedure in which a unique 10% of the data (i.e., the 'test set') are used in each iteration for evaluation, and the remaining 90% (i.e., the 'training set') are used to select the most useful features (of the 370 available as described in "Features" above) and construct our models. The reported accuracy is an average across the 10 folds. In a given fold, data from any individual speaker can occur in the test set or the training set, but not both. In order to optimize the ratio of training examples to their dimensionality, we select the *N* features with the highest Pearson's correlation coefficients between each feature and the binary class. Since in each fold we are selecting features based only on a subset of the data, these features need not always be the same in each iteration, although we show in the Supplementary Material that there is very little variability between folds for $N < 100$ (see the Supplementary Material for the full details).

RESULTS

Figure 1 shows the average accuracies (and s.d.) for the logistic regression method. The maximum average accuracy (81.92%) in distinguishing between AD and controls is achieved with the 35 top-ranked features. The accuracy remains relatively constant until we choose a feature set of size 50 (accuracy = 78.72%), after which it drops off sharply. As a result, we use the top 50 features in our factor analysis. Those features, and their correlation with diagnosis, are shown in Table 2. Using all 370 features, the logistic regression method obtains 58.51% accuracy on average, which reinforces the need to do feature selection given high-dimensional feature space such as this.

Factor analysis

To help discover the underlying structure of the data, we conduct an exploratory factor analysis. Since our data do not satisfy the assumption of multivariate normality, we use the method of principal axis factors (PAF), as recommended by Fabrigar et al. [70]. We include 50 features in the factor analysis, as discussed in the previous section.

A screen test suggests that four factors are sufficient to account for the majority of the variance. To interpret the factor structure, it is customary to perform a rotation. Although varimax is the most popular rotation algorithm, it is an orthogonal rotation and is therefore guaranteed to produce uncorrelated factors. To fairly examine the degree of heterogeneity of linguistic impairments in patients with AD, we chose promax, an oblique rotation which allows factors to be correlated with each other [71].

Feature loadings on the four factors are presented in Table 2. Factor signs were deliberately set such that higher factor scores reflect greater impairment. As is customary in exploratory factor analysis, we name and present a subjective interpretation of the factors, below.

Factor 1: Semantic impairment

All of the high loadings reflect characteristics of semantically impoverished language, similar to that seen in semantic dementia, a degenerative disorder specifically affecting the temporal lobe [72]. Individuals scoring high on this factor produce many pronouns (+NP $\longrightarrow$ PRP, +pronoun ratio) and few nouns (–nouns), and are biased toward shorter (–word length) and higher frequency words (+frequency, +verb frequency). They also use a less diverse vocabulary (–Honoré's statistic) and exhibit increased repetition of content (–cosine distance).

Pronouns and high frequency words suggest empty, vague, or non-specific speech. A decrease in the proportion of nouns and an increase in the proportion of verbs is the same pattern as seen in semantic dementia [73, 74]. Individuals with a semantic impairment may have difficulty accessing more specific nouns and

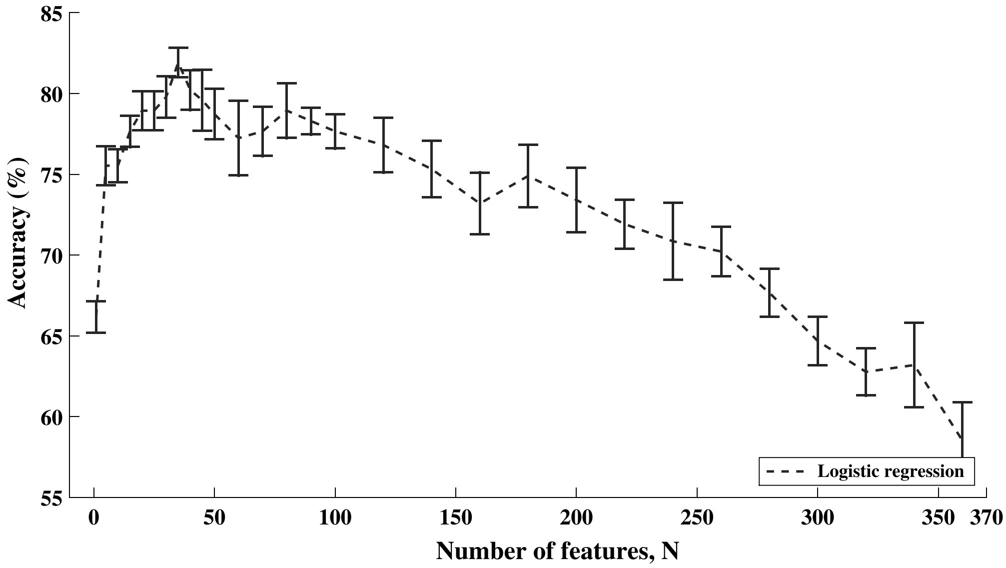

Fig. 1. Accuracies and standard errors, across the number N of predictor features used.

verbs, and as a result may replace them with generic, high-frequency substitutes (e.g., Meteyard et al. [75]). Negative Honoré's statistic suggests low lexical diversity, which has been observed in anomic aphasia [76], and negative cosine distance suggests high repetition, which bears similarity to the "stereotypic thematic perseverations" seen in semantic dementia [73]. Examples of the adverbial construction (+ADVP ⟶ RB) include "*the little girl's reaching up* ***there***" and "*a tree coming up* ***here***"; that is, the adverb serves a deictic purpose, which is more common amongst aphasics with a semantic impairment [77].

Factor 2: Acoustic abnormality

All high loadings here relate to features derived from acoustic analysis. All but one of these refer to either the skewness or kurtosis of individual Mel-frequency cepstral coefficients, whose perceptual values may not be distinguishable to humans, and whose anatomical basis depends on treating the vocal tract within a source-filter model of convolution. The remaining feature, phonation rate, is the proportion of an utterance that is vocalized; low values here refer to more time being spent silently, as in a pause.

Factor 3: Syntactic impairment

This factor appears to reflect a syntactic deficit somewhat reminiscent of such conditions as Broca's aphasia and progressive nonfluent aphasia. High-scoring patients produced fewer verbs, which is typical of agrammatic patients (e.g., Saffran et al. [78], and Thompson et al. [79]). They also produced fewer auxiliary verbs, and fewer gerunds and participles. Patients with Broca's aphasia often omit auxiliaries and use only the simplest verb tenses (e.g., *he reach* might be preferred over *he is reaching,* which requires an auxiliary and a participle) [80, 81]. Additionally, they produced more sentence fragments, and more words tagged as "not in dictionary," which could include phonological paraphasias, distortions, and unrecognizable words (note that the automatic analysis cannot distinguish between these different language phenomena). We note that while these deficits resemble Broca's aphasia and progressive nonfluent aphasia (PNFA) in their form, they are less severe, seldom reaching the point of frank agrammatism or "telegraphic" speech seen in those disorders. Presumably this reflects the fact that cortical damage to language centers in AD is less severe than in those conditions.

Factor 4: Information impairment

This factor primarily includes mention of key words and information units. Patients with high scores produced relatively uninformative picture descriptions, failing to mention key concepts. This factor differs from Factor 1 in that the relevant features do not describe generic properties of the words, such as their

Table 2
Correlations with diagnosis (first column) and promax factor loadings. Loadings less than 0.1 are excluded. Bold font indicates a loading greater than 0.3.

	r	Fac. 1	Fac. 2	Fac. 3	Fac. 4	Comm.
Pronoun:noun ratio	0.35	**1.01**		**–0.32**		0.80
NP -> PRP	0.37	**0.88**		–0.24		0.59
Frequency	0.34	**0.74**				0.56
Adverbs	0.31	**0.51**			0.19	0.39
ADVP -> RB	0.30	**0.44**		0.10		0.25
Verb frequency	0.21	**0.39**		0.13		0.21
Nouns	–0.27	**–0.97**		**0.37**		0.79
Word length	–0.41	**–0.60**		–0.13		0.39
NP -> DT_NN	0.10	**–0.52**			–0.19	0.37
Honore's statistic	–0.25	**–0.46**	–0.14	–0.14	**0.33**	0.21
Inflected verbs	–0.19	**–0.39**				0.21
Average cosine distance	–0.19	**–0.33**		–0.15	0.13	0.14
Skewness(MFCC 1)	0.22		**0.95**			0.84
Skewness(MFCC 2)	0.20		**0.87**		–0.14	0.68
Kurtosis(MFCC 5)	0.19		**0.78**			0.54
Kurtosis(VEL(MFCC 3))	0.24	–0.17	**0.44**		0.24	0.34
Phonation rate	–0.21	0.16	**–0.62**		–0.28	0.56
Skewness(MFCC 8)	–0.22		**–0.39**		–0.13	0.23
Not-in-dictionary	0.38	–0.14		**0.53**	0.26	0.35
ROOT -> FRAG	0.23	–0.15		**0.36**	0.19	0.18
Verbs	–0.29	**0.38**		**–1.05**	0.20	0.81
VP rate	–0.19	**0.37**		**–0.95**	**0.32**	0.70
VP -> AUX_VP	–0.23	–0.16		**–0.56**	0.18	0.34
VP -> VBG	–0.27	–0.28		**–0.34**	0.21	0.21
Key word: window	–0.29			0.20	**–0.79**	0.50
Info unit: window	–0.32			0.12	**–0.63**	0.36
KEY WORD: sink	–0.23				**–0.62**	0.30
KEY WORD: cookie	–0.23		0.13		**–0.61**	0.29
PP proportion	–0.21			0.18	**–0.61**	0.34
Key word: curtain	–0.25				**–0.56**	0.25
PP rate	–0.21			0.19	**–0.55**	0.33
Info unit: curtain	–0.26				**–0.53**	0.27
Key word: counter	–0.18			0.14	**–0.47**	0.19
Info unit: cookie	–0.24				**–0.46**	0.24
Info unit: sink	–0.31				**–0.43**	0.24
Info unit: girl	–0.30				**–0.42**	0.20
Info unit: girl's action	–0.25		0.13	–0.12	**–0.36**	0.14
Info unit: dish	–0.24	–0.12			–0.29	0.16
Key word: stool	–0.28	–0.15			–0.29	0.18
Key word: mother	–0.32			–0.27	–0.26	0.17
Info unit: stool	–0.32	–0.29			–0.21	0.25
Skewness(MFCC 12)	–0.19				–0.18	0.06
Info unit: woman	–0.29			–0.16	–0.18	0.14
VP -> VBG_PP	–0.34	–0.19		–0.30	–0.12	0.23
VP -> IN_S	–0.20				–0.10	0.04
VP -> AUX_ADJP	–0.19	–0.11				0.02
VP -> AUX	0.20	0.28				0.10
VP -> VBD_NP	0.19					0.03
Cosine cutoff: 0.5	0.19		0.15	0.14		0.10
INTJ -> UH	0.18	0.25				0.08

frequency or part-of-speech, but rather their appropriateness and specific semantic relevance to the task at hand. These participants also lacked prepositional phrases, which reflects a lower level of detail in their descriptions. There may be some relation between the absence of certain information units and a reduction in prepositional phrases. For example, both the information unit and key word features for "sink" are negatively weighted on this factor. When we examined the control transcripts, we found that in 57% of cases, the word *sink* appeared as the object of a prepositional phrase (e.g., *water's overflowing in the sink, the* *water is spilling out from the sink, the mother's working at the sink*). This potential connection between the omission of certain content words and a reduction in prepositional phrases will require further investigation in future work.

Relationships among factors

Figure 2 shows pairwise scatterplots of individual transcript scores on each factor. All four factors are significantly different between groups, which is not surprising given that the features comprising the factors were pre-selected for their association with diagnosis. Of greater interest is the degree of correlation among the factors, which can be estimated by the oblique promax rotation, as opposed to orthogonal rotations such as varimax that guarantee uncorrelated factors. The correlation coefficients among all samples, and limited to AD and control subjects alone, are given in each plot.

An intriguing result is the correlation between Factor 1 (semantic impairment) and Factor 3 (syntactic impairment). These factors are moderately correlated in the control group ($R = 0.42$) but much less correlated in the AD group ($R = 0.19$). This suggests that in cognitively normal individuals, semantic and syntactic abilities are somewhat linked, but when these abilities decline in AD, there can be an asymmetry to the impairment. This may be attributable to damage specific to networks responsible for distinct aspects of linguistic competence (see discussion).

Another pattern is seen when we consider Factor 4 (information impairment). Factor 1 and Factor 4 are more highly correlated in the AD group ($R = 0.49$) than the control group (0.18). Similarly, Factor 3 and Factor 4 are more highly correlated in the AD group ($R = 0.31$) than the control group (–0.097). Since information is expressed through both syntax and semantics, we hypothesize that difficulty in either Factor 1 or Factor 3 would also imply impairment in conveying information. That a similar correlation is not seen in the control data might be due to the fact that the task is generally easy for controls, and our method for scoring information units is very simplistic. For example, when we look at individual transcripts, we find a control participant who uses a number of pronouns and generic, high frequency words (*thing, something*). However, this semantic "impairment" does not prevent the participant from mentioning most of the relevant information units. Conversely, a control participant who uses more detailed language will not be able to increase their information score beyond what our simple keyword-spotting algorithm can detect.

Factor 2 (acoustic abnormality) is uncorrelated with Factor 1 (semantic impairment) and Factor 3 (syntactic impairment) in both the AD and control groups. Factor 2 is moderately correlated with Factor 4 (information impairment) in the AD ($R = 0.32$) and control ($R = 0.31$) groups. This would suggest that participants who spoke slower (phonation rate) or had some other acoustic irregularity (MFCC features) also offered less information, regardless of their clinical diagnosis.

DISCUSSION

Although language impairment is a secondary cognitive symptom of AD, numerous studies have shown that language skills become abnormal relatively early in AD [82, 83] and can serve as a sensitive index of disease severity over time. The present study employs a combination of automated quantification of language with a machine-learning classification approach to accurately distinguish between healthy controls and patients with AD on the basis of short speech samples from a picture description task.

The relatively short length (roughly 100 words) of the picture descriptions is one drawback of the current study. Other researchers have suggested that 150 words is an acceptable minimum length for language analysis [36, 78]. Our classification results indicate that there is still valuable information to be found in the short samples, confirming previous studies using picture descriptions to assess language in dementia [40, 42, 74]; however, we expect that the accuracies of each feature value would increase as the length of the sample increased. We also emphasize that our findings here do not necessarily generalize to other spoken language tasks. For example, Sajjadi et al. [36] found that the picture description was more sensitive to semantic impairment in AD, while an interview format was more sensitive to syntactic measures.

Factor analysis reveals that our relatively large set of linguistic measures can be mapped to four latent variables, broadly representing syntax/fluency, semantics, acoustic differences, and other information content. Each of these language domains has been reported separately to be altered in patients with AD/MCI [2, 84–86], but the relationships between separate domains of impairment have seldom been characterized. Many previous studies have relied upon labor-intensive manual analyses of language

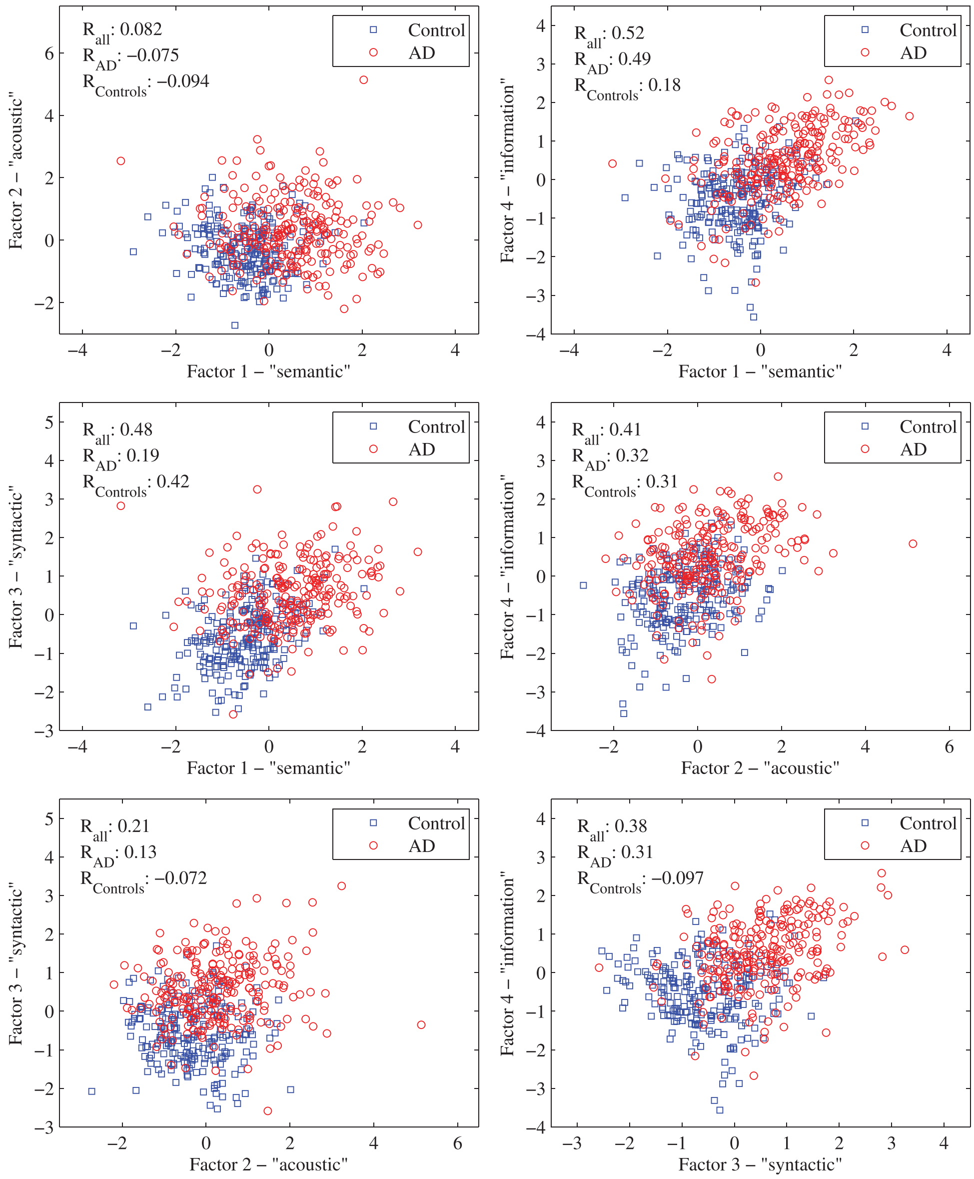

Fig. 2. Promax scores for AD and control participants.

samples, emphasizing particular aspects according to the research interests of the authors. The large heterogeneity in language impairments reported across studies leaves open the question of whether there is a single quantifiable aspect of spoken language output that is particularly diagnostic of AD. The heterogeneity of reported impairments is presumably attributable to two sources of variability: differences between individual patients, and differences in the methods and hypotheses employed by the authors of the studies.

For the present study, we aimed to capture as broad a spectrum of linguistic variables as possible using fully automated analysis of transcripts and acoustic recordings, from a relatively large sample of picture descriptions. This approach, while potentially missing some useful variables that require manual analysis, can characterize the level of heterogeneity present across individual patients tested with a consistent protocol. Previous studies of cognitive decline in MCI/AD have highlighted considerable heterogeneity among patients, but have tended to view language as a fairly unitary construct along with other cognitive domains including episodic memory and visuospatial cognition. Despite this heterogeneity, there seems to be a typical pattern of decline in AD particularly driven by impairments in episodic memory [87], with other domains affected in a more limited set of patients.

Heterogeneity in the cognitive profile of AD has been linked to different patterns of brain atrophy. Nearly all cases of AD include atrophy in the hippocampus and other parts of the medial temporal lobe, which is thought to underlie the episodic memory impairments. Involvement of the left parietal lobe has been linked to reduced speech output in the form of logopenic progressive aphasia (see below), while atrophy in the right parietal lobe and occipital cortex are associated with visuospatial impairments [9]. These cortical presentations of AD represent less common variants than the typical pattern of medial temporal atrophy. In contrast, semantic impairments are quite commonly reported in the later stages of AD, consistent with the fact that temporal lobe structures involved in semantic processing are among the most frequently affected in late-stage AD, whereas frontal-parietal regions involved in syntax and phonology are often relatively spared [88].

Variability in the cognitive presentation of AD is to be expected, given that different patients have damage to different parts of the cortex. To date, little work has been done to characterize the relationship between impairments on specific subdomains of language and cortical atrophy in AD. This situation contrasts strongly with that of PPA, a less common form of dementia in which deterioration of language is the most notable symptom. A current consensus among researchers holds that PPA can be clearly divided into at least three subtypes characterized by impairments to distinct aspects of language [72]. In PNFA, syntax and grammar are greatly affected, while semantic knowledge is spared. PNFA is linked to pathology in the left frontal cortex, and somewhat resembles the syndrome of Broca's aphasia, linked to stroke-induced damage in the same region. In semantic dementia, word knowledge is greatly reduced but grammar is largely spared. This syndrome is linked to degeneration of the anterior temporal lobes. In logopenic progressive aphasia, subtler syntactic deficits are present (mainly a simplification of spontaneous output, and reduced fluency), and repetition may be impaired. Logopenic progressive aphasia is linked to damage to the left temporoparietal junction, and unlike the other two subtypes of PPA, it is considered to be an atypical presentation of AD, as it shares the same underlying molecular pathology [12].

The double dissociation of syntactic and semantic impairments in PPA highlights the fact that distinct brain networks make unique contributions to linguistic competence. Thus, even though the episodic memory impairment dominates the cognitive profile of AD/MCI patients, variability in cortical involvement across patients should differentially impact the same subdomains of language that are affected in PPA. However, this variability in AD/MCI is likely to be more subtle than between PPA subtypes, as the patients do not fall into clearly distinguishable diagnostic categories.

In the present study, we characterized the heterogeneity of language impairment using factor analysis. The choice of rotation method is critical here: the majority of exploratory factor analyses of cognitive impairment use varimax rotation (e.g., Duong et al. [39]; Monti et al. [89]), which is guaranteed to produce factors that are uncorrelated with each other. While this can be useful for discovering the underlying domains of impairment and interpreting them, the enforced lack of correlation among the factors may bias the reader to conclude that strong heterogeneity exists. We therefore elected to use promax rotation, which allows the factors to be correlated.

In fact, the results do point to a considerable degree of heterogeneity in linguistic deficits, since the resulting factors are weakly correlated with each other, in general. These results do not support the existence of a single severity factor underlying overall linguistic impairments. Rather, patients vary in their expression of linguistic impairment in different subdomains. Based on existing knowledge of anatomical-behavioral relationships in PPA, we hypothesize that heterogeneity among AD patients may ultimately be accounted for by the variable spread of cortical involvement in the disease. Specifically, we predict that semantic impairment will be linked to atrophy of the anterior temporal lobes, and fluency reductions (reduced speech rate and simplified syntax, as seen in logopenic progressive aphasia) will be linked to the left temporoparietal junction. Deficits in information content may ultimately be linked to memory impairments, while acoustic abnormalities are more likely to reflect damage in frontal premotor circuits involved in speech. Testing of these relationships will require the collection of large numbers of anatomical images accompanied by language samples. Although numerous large-scale longitudinal studies are now underway worldwide incorporating anatomical imaging, the collection of spontaneous speech data from the same patients is still uncommon. Given that language impairments are potentially informative about the involvement of distinct brain networks in neurodegenerative disease, we hope that collection of standardized language samples will become a more common component of neuropsychological batteries linked to longitudinal anatomical studies.

Finally, although we have emphasized here the relationships between language impairments and structural degeneration in AD, functional impairment is related to factors beyond cortical atrophy, and may be amenable to intervention. Therefore, as new interventions are developed for mitigating the symptoms of AD, sensitive instruments are needed to evaluate the effectiveness of such interventions on improving cognition. Although memory impairment may be the definitive symptom for the *diagnosis* of AD, it is not necessarily the most sensitive index of cognitive function and response to intervention. Language function especially degrades as the disease progresses through moderate and severe stages [90, 91], and has been shown to improve in response to successful treatment with acetylcholinesterase inhibitors [92]. Therefore, computational analyses of naturalistic language may ultimately provide a means to monitor changes in cognitive status over the course of the disease, as well as responsiveness to interventions, and can thus serve as a useful clinical tool for purposes well beyond diagnosis.

ACKNOWLEDGMENTS

Kathleen Fraser acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC). Jed Meltzer's contribution was funded by a New Investigator Research Grant from the Alzheimer's Association. Frank Rudzicz's contribution to this work is funded by an NSERC Discovery grant (RGPIN 435874) and by a Young Investigator award by the Alzheimer Society of Canada. The original acquisition of the DementiaBank data was supported by NIH grants AG005133 and AG003705 to the University of Pittsburgh, and maintenance of the data archive is supported by NIH-NIDCD grant R01-DC008524 to Carnegie Mellon University.

Authors' disclosures available online (http://j-alz.com/manuscript-disclosures/15-0520r1).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: http://dx.doi.org/10.3233/JAD-150520.

REFERENCES

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

[2] Meilán JJG, Martínez-Sánchez F, Carro J, López DE, Millian-Morell L, Arana JM (2014) Speech in Alzheimer's disease: Can temporal and acoustic parameters discriminate dementia. *Dement Geriatr Cogn Disord* **37**, 327-334.

[3] Guinn CI, Habash A (2012) Language analysis of speakers with dementia of the Alzheimer's Type. *AAAI Fall Symposium: Artificial Intelligence for Gerontechnology*, pp. 8-13.

[4] Jarrold W, Peintner B, Wilkins D, Vergryi D, Richey C, Gorno-Tempini ML, Ogar J (2014) Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. *Proceedings of the ACL Workshop on Computational Linguistics and Clinical Psychology*, pp. 27-36.

[5] Thomas C, Keselj V, Cercone N, Rockwood K, Asp E (2005) Automatic detection and rating of dementia of Alzheimer type through lexical analysis of spontaneous speech. *Proceedings of the IEEE International Conference on Mechatronics and Automation*, pp. 1569-1574.

[6] Fraser KC, Meltzer JA, Graham NL, Leonard C, Hirst G, Black SE, Rochon E (2014) Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. *Cortex* **55**, 43-60.

[7] Kirshner HS (2012) Primary progressive aphasia and Alzheimer's disease: Brief history, recent evidence. *Curr Neurol Neurosci Rep* **12**, 709-714.

[8] MacKay DG, James LE, Taylor JK, Marian DE (2007) Amnesic HM exhibits parallel deficits and sparing in language and memory: Systems versus binding theory accounts. *Lang Cogn Process* **22**, 377-452.

[9] Lam B, Masellis M, Freedman M, Stuss DT, Black SE (2013) Clinical, imaging, and pathological heterogeneity of the Alzheimer's disease syndrome. *Alzheimers Res Ther* **5**, 1-14.

[10] Stopford CL, Snowden JS, Thompson JC, Neary D (2008) Variability in cognitive presentation of Alzheimer's disease. *Cortex* **44**, 185-195.

[11] Snowden JS, Stopford CL, Julien CL, Thompson JC, Davidson Y, Gibbons L, Pritchard A, Lendon CL, Richardson AM, Varma A, Neary D, Mann D (2007) Cognitive phenotypes in Alzheimer's disease and genetic risk. *Cortex* **43**, 835-845.

[12] Rohrer JD, Rossor MN, Warren JD (2012) Alzheimer's pathology in primary progressive aphasia. *Neurobiol Aging* **33**, 744-752.

[13] Ahmed S, de Jager CA, Haigh A-MF, Garrard P (2012) Logopenic aphasia in Alzheimer's disease: Clinical variant or clinical feature? *J Neurol Neurosurg Psychiatry* **83**, 1056-1062.

[14] Leyton CE, Villemagne VL, Savage S, Pike KE, Ballard KJ, Piguet O, Burrell JR, Rowe CC, Hodges JR (2011) Subtypes of progressive aphasia: Application of the International Consensus Criteria and validation using beta-amyloid imaging. *Brain* **134**, 3030-3043.

[15] Faber-Langendoen K, Morris JC, Knesevich JW, LaBarge E, Miller JP, Berg L (1988) Aphasia in senile dementia of the Alzheimer type. *Ann Neurol* **23**, 365-370.

[16] Ahmed S, Haigh A-MF, de Jager CA, Garrard P (2013) Connected speech as a marker of disease progression in autopsy-proven Alzheimers disease. *Brain* **136**, 3727-3737.

[17] Weiner MF, Neubecker KE, Bret ME, Hynan LS (2008) Language in Alzheimer's disease. *J Clin Psychiatry* **69**, 1223-1227.

[18] Reilly J, Troche J, Grossman M (2011) Language processing in dementia. In *The Handbook of Alzheimer's Disease and Other Dementias*. Budson AE, Kowall NW, eds. Wiley-Blackwell.

[19] Kirshner HS, Webb WG, Kelly MP (1984) The naming disorder of dementia. *Neuropsychologia* **22**, 23-30.

[20] Taler V, Phillips NA (2008) Language performance in Alzheimer's disease and mild cognitive impairment: A comparative review. *J Clin Exp Neuropsychol* **30**, 501-556.

[21] Oppenheim G (1994) The earliest signs of Alzheimer's disease. *J Geriatr Psychiatry Neurol* **7**, 116-120.

[22] Kempler D (1995) Language changes in dementia of the Alzheimer type. In *Dementia and Communication: Research and Clinical Implications*, Lubinski R, ed. Singular, San Diego, pp. 98-114.

[23] Salmon DP, Butters N, Chan AS (1999) The deterioration of semantic memory in Alzheimer's disease. *Can J Exp Psychol* **53**, 108-117.

[24] Monsch AU, Bondi MW, Butters N, Salmon DP, Katzman R, Thal LJ (1992) Comparisons of verbal fluency tasks in the detection of dementia of the Alzheimer type. *Arch Neurol* **49**, 1253-1258.

[25] Adlam A-LR, Bozeat S, Arnold R, Watson P, Hodges JR (2006) Semantic knowledge in mild cognitive impairment and mild Alzheimer's disease. *Cortex* **42**, 675-684.

[26] Henry JD, Crawford JR, Phillips LH (2004) Verbal fluency performance in dementia of the Alzheimer's type: A meta-analysis. *Neuropsychologia* **42**, 1212-1222.

[27] Robinson KM, Grossman M, White-Devine T, D'Esposito M (1996) Category-specific difficulty naming with verbs in Alzheimer's disease. *Neurology* **47**, 178-182.

[28] Nicholas M, Obler LK, Albert ML, Helm-Estabrooks N (1985) Empty speech in Alzheimer's disease and fluent aphasia. *J Speech Lang Hear Res* **28**, 405-410.

[29] Ahmed S, de Jager CA, Haigh, A-MF, Garrard P (2013) Semantic processing in connected speech at a uniformly early stage of autopsy-confirmed Alzheimer's disease. *Neuropsychology* **27**, 79-85.

[30] Appell J, Kertesz A, Fisman M (1982) A study of language functioning in Alzheimer patients. *Brain Lang* **17**, 73-91.

[31] Rapp AM, Wild B (2011) Nonliteral language in Alzheimer dementia: A review. *J Int Neuropsychol Soc* **17**, 207-218.

[32] Glosser G, Deser T (1991) Patterns of discourse production among neurological patients with fluent language disorders. *Brain Lang* **40**, 67-88.

[33] Blonder LX, Kort ED, Schmitt FA (1994) Conversational discourse in patients with Alzheimer's disease. *J Linguist Anthropol* **4**, 50-71.

[34] Croisile B, Ska B, Brabant M-J, Duchene A, Lepage Y, Aimard G, Trillet M (1996) Comparative study of oral and written picture description in patients with Alzheimer's disease. *Brain Lang* **53**, 1-19.

[35] Ehrlich JS, Obler LK, Clark L (1997) Ideational and semantic contributions to narrative production in adults with dementia of the Alzheimer's type. *J Commun Disord* **30**, 79-99.

[36] Sajjadi SA, Patterson K, Tomek M, Nestor PJ (2012) Abnormalities of connected speech in semantic dementia vs Alzheimer's disease. *Aphasiology* **26**, 847-866.

[37] Kempler D, Curtiss S, Jackson C (1987) Syntactic preservation in Alzheimer's disease. *J Speech Lang Hear Res* **30**, 343-350.

[38] Hodges JR, Patterson K (1995) Is semantic memory consistently impaired early in the course of Alzheimer's disease? Neuroanatomical and diagnostic implications. *Neuropsychologia* **33**, 441-459.

[39] Duong A, Giroux F, Tardif AA, Ska B (2005) The heterogeneity of picture-supported narratives in Alzheimer's disease. *Brain Lang* **93**, 173-184.

[40] Rentoumi V, Raoufian L, Ahmed S, de Jager CA, Garrard P (2014) Features and machine learning classification of connected speech samples from patients with autopsy proven Alzheimer's disease with and without additional vascular pathology. *J Alzheimers Dis* **42**, S3-S17.

[41] Orimaye SO, Wong JS-M, Golden KJ (2014) Learning predictive linguistic features for Alzheimer's disease and related dementias using verbal utterances. *Proceedings of the 1st Workshop on Computational Linguistics and Clinical Psychology (CLPsych)*, Baltimore, Maryland, pp. 78-87.

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

[43] Roark B, Mitchell M, Hosom J-P, Hollingshead K, Kaye J (2011) Spoken language derived measures for detecting mild cognitive impairment. *IEEE/ACM Trans Audio Speech Lang Process* **19**, 2081-2090.

[44] Lehr M, Prud'hommeaux ET, Shafran I, Roark B (2012) Fully automated neuropsychological assessment for detecting mild cognitive impairment. In *Proc 13th Ann Conf Int Speech Commun Assoc (Interspeech)*, pp. 1039-1042.

[45] MacWhinney B, Fromm D, Forbes M, Holland A (2011) AphasiaBank: Methods for studying discourse. *Aphasiology* **25**, 1286-1307.

[46] Becker JT, Boiler F, Lopez OL, Saxton J, McGonigle KL (1994) The natural history of Alzheimer's disease: Description of study cohort and accuracy of diagnosis. *Arch Neurol* **51**, 585-594.

[47] Goodglass H, Kaplan E (1983) *The Boston Diagnostic Aphasia Examination*. Lea & Febinger, Philadelphia.

[48] MacWhinney B (2000) *The CHILDES Project: Tools for analyzing talk*, 3rd edition. Lawrence Erlbaum Associates, Mahwah, New Jersey.

[49] Clark HH, Fox Tree JE (2002) Using uh and um in spontaneous speaking. *Cognition* **84**, 73-11.

[50] Yngve V (1960) A model and hypothesis for language structure. *Proc Am Physical Soc* **104**, 444-466.

[51] Fraser KC, Hirst G, Meltzer JA, Mack JE, Thompson CK (2014) Using statistical parsing to detect agrammatic aphasia. In *Proc BioNLP*, pp. 134-142.

[52] Chae J, Nenkova A (2009) Predicting the fluency of text with shallow structural features: Case studies of machine translation and human-written text. In *Proc 12th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 139-147.

[53] Brysbaert M, New B (2009) Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav Res Methods* **41**, 977-990.

[54] Stadthagen-Gonzalez H, Davis CJ (2006) The Bristol norms for age of acquisition, imageability, and familiarity. *Behav Res Methods* **38**, 598-605.

[55] Gilhooly KJ, Logie RH (1980) Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behav Res Methods* **12**, 395-427.

[56] Tweedie FJ, Baayen RH (1998) How variable may a constant be? Measures of lexical richness in perspective. *Comput Hum* **32**, 323-352.

[57] Durán P, Malvern D, Richards B, Chipere N (2004) Developmental trends in lexical diversity. *Appl Ling* **25**, 220-242.

[58] Covington MA, McFall JD (2010) Cutting the Gordian knot: The moving-average type–token ratio (MATTR). *J Quant Linguist* **17**, 94-100.

[59] Fergadiotis G, Wright HH, West TM (2013) Measuring lexical diversity in narrative discourse of people with aphasia. *Am J Speech Lang Pathol* **22**, S397-S408.

[60] Forbes-McKay KE, Venneri A (2005) Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. *Neurol Sci* **26**, 243-254.

[61] Giles E, Patterson K, Hodges JR (1996) Performance on the Boston Cookie Theft picture description task in patients with early dementia of the Alzheimer's type: Missing information. *Aphasiology* **10**, 395-408.

[62] Pakhomov SV, Smith GE, Chacon D, Feliciano Y, Graff-Radford N, Caselli R, Knopman DS (2010) Computerized analysis of speech and language to identify psycholinguistic correlates of frontotemporal lobar degeneration. *Cogn Behav Neurol* **23**, 165-177.

[63] Hakkani-Tür D, Vergyri D, Tür G (2010) Speech-based automated cognitive status assessment. In *Proceedings of the Interspeech'10, Annual Conference of the International Speech Communication Association*, pp. 258-261.

[64] Bayles KA, Tomoeda CK, McKnight PE, Helm-Estabrooks N, Hawley JN (2004) Verbal perseveration in individuals with Alzheimer's disease. *Semin Speech Lang* **25**, 335-347.

[65] Tomoeda CK, Bayles KA, Trosset MW, Azuma T, McGeagh A (1996) Cross-sectional analysis of Alzheimer disease effects on oral discourse in a picture description task. *Alzheimer Dis Assoc Disord* **10**, 204-215.

[66] Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO (2012) Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Trans Biomed Eng* **59**, 1264-1271.

[67] Rudzicz F, Hirst G, Van Lieshout P (2012) Vocal tract representation in the recognition of cerebral palsied speech. *J Speech Lang Hear Res* **55**, 1190-1207.

[68] Jaitly N, Hinton GE (2013) Using an autoencoder with deformable templates to discover features for automated speech recognition. *In Proceedings of Conference of the International Speech Communication Association (INTERSPEECH)*, pp. 1737-1740.

[69] Fraser K, Rudzicz F, Rochon E (2013) Using text and acoustic features to diagnose progressive aphasia and its subtypes. *In Proceedings of Interspeech 2013*, pp. 2177-2181.

[70] Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ (1999) Evaluating the use of exploratory factor analysis in psychological research. *Psychol Methods* **4**, 272-299.

[71] Costello AB, Osborne JW (2005) Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Pract Assess Res Eval* **10**, 1-9.

[72] Gorno-Tempini ML, Dronkers NF, Rankin KP, Ogar JM, Phengrasamy L, Rosen HJ, Johnson JK, Weiner MW, Miller BL (2004) Cognition and anatomy in three variants of primary progressive aphasia. *Ann Neurol* **55**, 335-346.

[73] Harciarek M, Kertesz A (2011) Primary progressive aphasias and their contribution to the contemporary knowledge about the brain-language relationship. *Neuropsychol Rev* **21**, 271-287.

[74] Wilson SM, Henry ML, Besbris M, Ogar JM, Dronkers NF, Jarrold W, Miller BL, Gorno-Tempini ML (2010) Connected speech production in three variants of primary progressive aphasia. *Brain* **133**, 2069-2088.

[75] Meteyard L, Quain E, Patterson K (2014) Ever decreasing circles: Speech production in semantic dementia. *Cortex* **55**, 17-29.

[76] Fergadiotis G, Wright HH (2011) Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology* **25**, 1414-1430.

[77] Varley R (1993) Deictic terms, lexical retrieval and utterance length in aphasia: An investigation of inter-relations. *Eur J Disord Commun* **28**, 23-41.

[78] Saffran EM, Berndt RS, Schwartz MF (1989) The quantitative analysis of agrammatic production: Procedure and data. *Brain Lang* **37**, 440-479.

[79] Thompson CK, Ballard KJ, Tait ME, Weintraub S, Mesulam M (1997) Patterns of language decline in non-fluent primary progressive aphasia. *Aphasiology* **11**, 297-321.

[80] Bastiaanse R, Thompson CK (2003) Verb and auxiliary movement in agrammatic Broca's aphasia. *Brain Lang* **84**, 286-305.

[81] Menn L (1995) *Non-fluent aphasia in a multilingual world* (Vol. 5). John Benjamins Publishing.

[82] Forbes-McKay KE, Venneri A, Shanks MF (2002) Distinct patterns of spontaneous speech deterioration: An early predictor of Alzheimer's disease. *Brain Cogn* **48**, 356-361.

[83] Ortiz KZ, Bertolucci PHF (2005) Language impairment in the early stages of Alzheimer's disease. *Arq Neuropsiquiatr* **63**(2A), 311-317.

[84] Snowdon DA, Kemper SJ, Mortimer JA, Greiner LH, Wekstein DR, Markesbery WR (1996) Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study. *J Am Med Assoc* **275**, 528-532.

[85] de Lira JO, Ortiz KZ, Campanha AC, Bertolucci PH, Minett TS (2011) Microlinguistic aspects of the oral narrative in patients with Alzheimer's disease. *Int Psychogeriatr* **23**, 404-412.

[86] Kemper S, Marquis J, Thompson M (2001) Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content. *Psychol Aging* **16**, 600-614.

[87] Lambon Ralph MA, Patterson K, Graham N, Dawson K, Hodges JR (2003) Homogeneity and heterogeneity in mild cognitive impairment and Alzheimer's disease: A cross-sectional and longitudinal study of 55 cases. *Brain* **126**(Pt 11), 2350-2362.

[88] Harasty JA, Halliday GM, Kril JJ, Code C (1999) Specific temporoparietal gyral atrophy reflects the pattern of language dissolution in Alzheimer's disease. *Brain* **122**, 675-686.

[89] Monti JM, Cooke GE, Watson PD, Voss MW, Kramer AF, Cohen NJ (2015) Relating hippocampus to relational memory processing across domains and delays. *J Cogn Neurosci* **27**, 234-245.

[90] Cummings JL, Benson DF, Hill MA, Read S (1985) Aphasia in dementia of the Alzheimer type. *Neurology* **35**, 394-394.

[91] Feldman H, Woodward M (2005) The staging and assessment of moderate to severe Alzheimer disease. *Neurology* **65**, S10-S17.

[92] Ferris SH, Farlow M (2013) Language impairment in Alzheimer's disease and benefits of acetylcholinesterase inhibitors. *Clin Interv Aging* **8**, 1007-1014.

[93] Hinton GE, Osindero S, Teh Y-W (2006) A fast learning algorithm for deep belief nets. *Neural Comput* **18**, 1527-1554.

[94] Lu X (2010) Automatic analysis of syntactic complexity in second language writing. *Int J Corpus Linguist* **15**, 474-496.

[95] Brunét E (1978) *Le Vocabulaire de Jean Giraudoux: Structure et Evolution*, Slatkine, Geneva.

[96] Honoré A (1979) Some simple measures of richness of vocabulary. *Assoc Lit Linguist Comput Bull* **7**, 172-177.

[97] Breedin SD, Saffran EM, Schwartz MF (1998) Semantic factors in verb retrieval: An effect of complexity. *Brain Lang* **63**, 1-31.