Using Automatic Assessment of Speech Production to Predict Current and Future Cognitive Function in Older Adults

Rachel Ostrand, PhD[1] , and **John Gunstad, PhD**[2]

Abstract

Neurodegenerative conditions like Alzheimer disease affect millions and have no known cure, making early detection important. In addition to memory impairments, dementia causes substantial changes in speech production, particularly lexical-semantic characteristics. Existing clinical tools for detecting change often require considerable expertise or time, and efficient methods for identifying persons at risk are needed. This study examined whether early stages of cognitive decline can be identified using an automated calculation of lexical-semantic features of participants' spontaneous speech. Unimpaired or mildly impaired older adults (N = 39, mean 81 years old) produced several monologues (picture descriptions and expository descriptions) and completed a neuropsychological battery, including the Modified Mini-Mental State Exam. Most participants (N = 30) returned one year later for follow-up. Lexical-semantic features of participants' speech (particularly lexical frequency) were significantly correlated with cognitive status at the same visit and also with cognitive status one year in the future. Thus, automated analysis of speech production is closely associated with current and future cognitive test performance and could provide a novel, scalable method for longitudinal tracking of cognitive health.

Introduction

Alzheimer disease (AD) and other forms of dementia are epidemic. With the aging of the baby boomer generation, the prevalence of AD will be more than triple by 2050 and generate US$1.1 trillion in annual health care costs in the United States alone.[1] With no known cure, early detection is essential for disease management. Detecting early signs of cognitive impairment often involves initial screening during a primary care or neurological evaluation, followed by comprehensive neuropsychological testing of cognitive function. Administration and interpretation of these assessments require substantial expertise and time, and thus cannot be easily conducted at regular intervals to identify and track people at risk. Additionally, screening techniques that can only be found in clinical settings may provide a barrier to individuals with reduced access to medical care or persons reluctant to discuss concerns about memory loss. New methods for early detection that can be delivered in community settings or at home could allow individuals to monitor cognitive status over time and then seek a comprehensive evaluation with their providers (e.g., neurological evaluation, neuroimaging, and neuropsychological testing) at the earliest signs of decline.

Prior work has shown that AD and other forms of dementia are associated with measurable changes in the affected person's speech production. In particular, during early stages of cognitive decline, word-finding and semantic knowledge become more difficult, likely due to a combination of the degradation of the stored meaning of words and the process of lexical access (for reviews, see studies by Kemper & Altmann,[2] Burke & Shafto,[3] and Obler & Albert[4]). This degradation induces reduced semantic specificity, including vague and empty words, higher-frequency words with less precise meanings, and indefinite articles (e.g., see studies by Bird et al.,[5] Croisile et al.,[6] Hier et al.,[7] Feyereisen et al.,[8] and Nicholas et al.[9]). In effect, early-stage dementia reduces the amount of specific content information conveyed during speech, while maintaining contextual relevance and grammaticality. In contrast, other levels of linguistic processing, including articulatory production, phonetic

---

[1] Department of Healthcare and Life Sciences, IBM Research, Yorktown Heights, NY, USA
[2] Department of Psychological Sciences & Brain Health Research Institute, Kent State University, Kent, OH, USA

Corresponding Author:
Rachel Ostrand, Department of Healthcare and Life Sciences, IBM Research, 1101 Kitchawan Road, Yorktown Heights, NY 10598, USA.
Email: ostrand.rachel@gmail.com

retrieval, and syntax, remain largely unimpaired until much more advanced stages of the disease (e.g., see studies by Hier et al.,[7] Bayles et al.,[10] Forbes-McKay et al.,[11] Salvatierra et al.,[12] Hoffmann et al.,[13] and Filiou et al.[14]).

This relative impairment of lexical-semantics is particularly true during the early stages of dementia, a time when other behavioral symptoms are often not yet noticeable. This suggests that measuring lexical-semantic properties of an at-risk person's speech production may be an effective way to identify and track early-stage cognitive decline. The present study investigates whether certain automatically calculable lexical-semantic features of spontaneous speech are reliable indicators of cognitive decline, and if spontaneous speech could be used for frequent, longitudinal assessment of people at risk for dementia. We do so by assessing whether a set of lexical-semantic features which are calculated automatically from a sample of spontaneous speech can predict current and future neuropsychological test scores.

Linguistic Changes in AD and Related Dementias

There is substantial research documenting linguistic changes in people with neurodegenerative disorders. In particular, in early-stage mild cognitive impairment (MCI) and dementia, lexical access, word retrieval, and other types of semantic knowledge become more difficult (for reviews, see studies by Burke & Shafto[3] and Obler & Albert[4]), stemming from the degradation of both the semantic meaning of particular words in the lexicon in addition to a disruption to the process of lexical access.[2] Behaviorally, this is manifest in speech production as a reduction in semantic specificity, causing an increase in vague, generic, or empty words, higher-frequency words with less precise meanings, and increased use of indefinite articles and anaphora (e.g., see studies by Bird et al.,[5] Croisile et al.,[6] Hier et al.,[7] Feyereisen et al.,[8] and Nicholas et al.[9]). This reduction in semantic specificity means that speakers produce less specific content information while still maintaining the overall semantic context and grammaticality. In contrast, other levels of linguistic processing remain largely unimpaired, as demonstrated by oral reading, writing to dictation, word repetition, and phonemic fluency tasks that are nearly on par with those of unimpaired controls.[10-12,15]

Research on the changes in speech production among people with dementia has, broadly speaking, been conducted using two types of methodologies. One method is using controlled elicitation tasks—where the participant is given some stimulus and is asked to produce small bits of speech in response. These types of tasks include confrontation naming, when the participant is shown a series of pictures and asked to give their names (e.g., the Boston Naming Test), and semantic or phonemic fluency, when the participant is given a category or a letter and asked to provide as many words as they can which belong to that category or start with that letter, among others. The constraint of controlled tasks is both a benefit and a disadvantage. It makes quantification of a participant's performance and comparison against other patients or established norms easier,

as there is a specific, predetermined metric to measure, such as the number of words produced in the given category. On the other hand, these tasks only measure a small slice of the participant's linguistic and cognitive abilities and do not measure a person's language abilities in a way that is similar to everyday language use.

The other method to assess speech production is via spontaneous speech elicitation tasks—a prompt which allows the participant to speak in a relatively unconstrained manner and produce connected speech, such as describing a picture or answering an open-ended question. With these types of tasks, the *lack* of constraint is both a benefit and a disadvantage. They can assess the participant's abilities and deficits in a situation which is much more relevant to real communication skills in their life, quantifying a major challenge with the disease,[16] and also allow a wide range of linguistic properties to be measured at multiple linguistic levels, including word choice, semantic content, syntactic structures, and acoustic properties. This measurement flexibility is a downside, however, as it is less straightforward to compare deficits between participants and to determine which metrics are the appropriate ones to measure.

Linguistic Changes on Controlled Elicitation Tasks

Controlled elicitation tasks are a good way to assess deficits on a specific linguistic characteristic. On semantic fluency tasks, where participants name as many exemplars of a category (e.g., animals) that they can, persons with AD perform significantly below the level of age-matched controls, and degree of impairment is correlated with clinically assessed dementia severity[7,11,12,15,17-20] (see study by Henry et al.[21] for a review). Patients with AD also perform significantly below the level of sex- and education-matched controls on verb fluency tasks, in which they need to specifically access words for "things that people do," and verb naming tasks, in which they are shown a picture and are asked to say what the person is doing or what is happening in the picture.[22]

People with AD are also impaired on explicit lexical access tasks. For example, patients with AD are impaired compared to healthy controls on confrontation naming, when they are shown a picture of an object and asked to produce its name.[7,10,11,15,18,20] In addition, participants with dementia score lower than healthy controls on the Wechsler Adult Intelligence Scale - 4 (WAIS) vocabulary subtest, which tests participants' ability to provide a concise definition for a given word.[7]

However, these linguistic deficits seem to be primarily focused on lexical and semantic processes. For example, although people with AD were impaired on semantic fluency tasks, they are relatively unimpaired compared to controls on similar phonemic fluency tasks, in which they are asked to produce as many words as they can which begin with a particular letter[11,12] (but see study by Sajjadi et al.[15] for different results). In a cuing task, participants were asked to write from dictation a pair of words. The second word was a homophone

(e.g., *nose/knows*); the first word provided a cue as to which lexical item of the homophone pair was intended. The cue word was either semantically (*thinks*) or syntactically (*she*) related to the target homophone (*knows*). Although controls performed the task with the same accuracy for the two types of cues, people with AD made significantly more errors when given a semantic cue rather than a syntactic cue, suggesting that semantic lexical access is particularly impaired in AD.[23] Similarly, on various neuropsychological tests, including the subtests of the Boston Diagnostic Aphasia Exam, participants with MCI and mild AD are more impaired on semantic and general fluency tasks,[24] and these tasks tend to be the most discriminative between early-stage patients and healthy controls.

Thus, evidence from controlled, experimental tasks points toward the linguistic deficits which occur in dementia, particularly in MCI and early-stage AD, as being those involving lexical-semantic processes. Research measuring patients' spontaneous speech shows similar types of deficits, as discussed in the following section.

Linguistic Changes on Spontaneous Speech Tasks

There is also some prior research which has investigated AD or MCI patients' linguistic deficits on spontaneous speech tasks (for reviews, see studies by Filiou et al.,[14] Boschi et al.,[25] Kavé & Goral,[26] Slegers et al.,[27] and Mueller et al.[28]). The most common spontaneous speech elicitation task is picture description, when the participant is shown a relatively complex scene and asked to describe in detail what is occurring. A popular picture for these types of tasks is the Cookie Theft from the Boston Diagnostic Aphasia Exam.[29] Other tasks include short writing prompts, semi-structured interviews, in which the experimenter prompts the participant with only occasional questions to keep the speech largely monologue, and retelling a story.[25] Spontaneous speech tasks demonstrate similar deficits as do controlled experimental tasks: Particularly in the early stages of cognitive decline, participants show impairment on lexical access and measures of semantic specificity, and relatively intact syntax and phonological and phonetic production.

A notable change in spontaneous speech in people with dementia is the increased production of empty or indefinite words as compared to healthy controls[6,9,26,30] or from early- to late-stage impairment.[7] These are words that are highly vague and nonspecific, like "thing" and "stuff," and their production may indicate word-finding difficulty or semantic memory degradation, as the speaker is attempting to refer to a particular entity without being able to access its name. The more general metric of this type of word-finding difficulty is manifest in average lexical frequency, as people with dementia produce more common, higher-frequency words than do healthy controls, and average lexical frequency increases as the patient's disease progresses.[5,20,31] Low-frequency words are more difficult to access from the lexicon, and less specific words (*thing*) tend to be substantially higher-frequency than a more specific counterpart (*cookie*). People with AD and cognitive decline also show reduced lexical diversity, repeating the

same words rather than producing unique words,[7,32-34] a behavior which increases with disease progression,[35,36] another manifestation of their reduced ability to easily retrieve words from their lexicon.

A related linguistic deficit in MCI and AD is the reduction of content words in spontaneous speech as compared to controls.[6,8,20,37] Content words convey semantic meaning (as contrasted with function words, which convey grammatical relationships and largely provide syntactic scaffolding) and include nouns, verbs, and adjectives (in contrast to, e.g., pronouns and prepositions). As with the previous metrics, the reduction of content word production in MCI and AD is likely an indicator of a lexical access deficit and is a metric of reduced semantic specificity. In fact, as dementia severity worsens, the production of content words is even further reduced.[38] Conversely, impaired participants produce increased function words,[15] particularly by replacing nouns (which are more specific) with pronouns (which are vaguer and less explicit), compared to controls[20,31,39] and as the degree of impairment increases.[7] People with AD also make fewer definite references to objects compared to controls,[8] a deficit which is thought to index impairments in declarative memory.[40]

Some studies report that patients produce fewer total words than controls[6,7,26] or that total word count decreases as impairment increases.[35,36] However, although many studies have found a numeric decrease in word count among impaired participants, the difference between groups or as a function of dementia severity is often not significant.[8,15,20,34,38,39,41] However, in referential communication dialogue tasks, which require the participant to adapt their speech based on learning from repeated interaction with a partner, people with AD actually produce *more* words than controls because they do not seem to learn and match their partner's language use.[8,41] Persons with AD also speak more slowly than controls in spontaneous speech, and speech rate anticorrelates with impairment severity,[13,15,41,42] with longer pauses and more word-finding delays among more impaired people.[6,13,26,43] Relatedly, people even with mild AD produce more filler words (filled pauses such as "um" and "uh") as compared to unimpaired older adults, suggesting the patients may be struggling with rapid lexical access.[15]

Lastly, patients with MCI or AD produce fewer words that are important to the discourse, and the words they do produce tend to be less relevant than those produced by elderly healthy controls. Much prior work investigating dementia patients' production of spontaneous speech (usually describing a picture which has fixed visual features) has counted the number of information or content units produced by the speaker, which can be compared against the "correct" number for that picture. Persons with dementia produce significantly fewer content units about the setting, events, characters, and the main idea of a scene compared to controls,[6,8-10,15,31,37] a deficit which is apparent even in very early-stage MCI,[34] and the degree of reduction is a function of dementia severity.[44] In both monologue and dialogue, more impaired participants produce fewer pieces of crucial information[41] and a lower density of relevant

information out of the total narrative they produce.[36,45] Conversely, people with AD produce more *irrelevant* information: words that are linguistically correct but not appropriate or useful to the current context.[6,41,43,45] Relatedly, more impaired patients' speech has lower idea density and higher uninformative output, producing fewer distinct pieces of information per number of words[7,36] and also per time unit,[41,45] and such deficits are evident in their writing as well.[38,46] This is also manifest in dialogue, when people with AD require more speaking turns to convey their message to their listener.[8] Together, these observations suggest that although impaired speakers may not reduce their total lexical output, what they do produce is wordier and less precise and conveys less information about the topic at hand. This suggests that patients, particularly with early-stage dementia, produce reduced semantic content and specificity, and these characteristics of speech could be used to quantify a person's degree of impairment. We assess this possibility in the present work, using automated calculation of lexical-semantic characteristics of spontaneous speech.

Can Spontaneous Speech Be Analyzed Automatically to Detect Cognitive Impairment?

The present work investigates whether semantic specificity in spontaneous speech can predict the degree of cognitive decline in older adults without a diagnosis of dementia. As mentioned above, current screening methods for cognitive impairment often require a visit with a clinician or trained examiner, potentially including administration of a battery of neuropsychological tests. This necessitates travel, access to medical care, substantial time for evaluation, and specially trained medical personnel. In addition, some screening tests have a restricted set of stimuli and thus frequent repetition in a short time span may artificially inflate scores. Together, these reasons encourage the development of additional methodologies for monitoring an individual patient's cognitive status over time.

The goal of the present work is not to combine language features with scores from traditional assessments to better predict future neuropsychological scores. Instead, we hope to demonstrate that lexical-semantic features of spontaneous speech can serve as a proxy measure for clinical cognitive screening instruments and thus could be administered *in between* clinical visits or before a person thinks to be evaluated the first time. Spontaneous speech can be collected in a participant's home, with minimal equipment and training, only a small outlay of time (minutes rather than hours), and using a much more varied set of materials to elicit responses from participants. Spontaneous speech thus has potential as a tool which can be collected with minimal burden on the participant. In addition, the features discussed here are straightforward to calculate automatically from transcribed speech. Thus, the present work seeks to demonstrate a predictive relationship between linguistic features and clinical cognitive status in older adults, to ultimately demonstrate that speech could itself be a useful diagnostic metric, opening the door to substantially more frequent assessment and monitoring of at-risk people in

between clinical examinations. Additionally, understanding the components of speech that are particularly informative of cognitive status can help elucidate the organization of language in the brain; in particular, demonstrating which characteristics track with damage induced by neurodegeneration over time can aid our understanding of how neurodegenerative conditions progress.

Here, we make use of the known relationship between cognitive status and speech production, except in the opposite causal direction. Rather than studying how cognitive status affects a particular linguistic characteristic, we investigate whether we can use changes in linguistic characteristics to predict changing cognitive status. In addition, in contrast to much prior research which compared different individuals' behaviors against each other to classify them as healthy or impaired, we employ a within-subjects, longitudinal design, and use a particular individual's speech production at an initial visit to predict their own cognitive status at that visit and also one year in the future. This avoids confounds of between-subjects differences in education level, socioeconomic status, language proficiency, and so on.

There is some precedent for this automatic approach in prior work, as several studies have collected spontaneous speech samples of participants with the ultimate goal of predicting participants' cognitive scores. Kavé and Dassa[47] found that the number of complete words spoken, type-token ratio, average lexical frequency, and the number of information units each individually correlated with Mini-Mental State Exam (MMSE) scores in persons with AD. Bucks and colleagues[32] collected speech samples from participants with AD and matched healthy controls via a semi-structured interview session, and found differences between AD and controls on several linguistic measures including part-of-speech counts and vocabulary richness, and their model classified participants (AD vs. control) with 87.5% accuracy. Ahmed and colleagues[37] found significant differences in spontaneous speech between healthy controls and people with autopsy-confirmed AD and also between participants at different stages of AD, on the proportion of pronouns and verbs produced, and a composite semantic and information content measure. Fraser et al.[31] took a computational approach to distinguishing participants with AD from healthy controls using spontaneous speech from picture descriptions. They calculated 370 linguistic features—lexical, semantic, information content, syntactic, and acoustic—and used machine learning to classify AD versus controls with up to 82% accuracy.

The Present Approach

However, there are a number of limitations in these prior studies using automatic assessment of spontaneous speech as a diagnostic measure of cognitive decline, which we address in several important ways. First, we use linguistic features to predict a cognitive score for all participants—both healthy and impaired—in order to characterize individuals along the continuous spectrum of cognitive decline. This is in contrast to most existing work that only predicted outcomes for

already-impaired participants (e.g., a study by Kavé & Dassa[47]) and/or merely conducted a binary by-group classification to discriminate patients from controls, rather than predicting the degree of impairment (e.g., see studies by Fraser et al.,[31] Bucks et al.,[32] Ahmed et al.,[37] and Asgari et al.[48]). In the present work, in line with the goal to be clinically relevant, we use linguistic features to assess participants at various stages of impairment (or lack thereof) and predict their neuropsychological score along a continuous scale rather than simply binning participants into the "impaired" or "unimpaired" group. This strategy would allow for a clinician to follow a participant's progression via his/her speech production and monitor cognitive decline as it occurs.

Second, in line with the goal of longitudinal screening and monitoring of cognitive function across a community, all linguistic features were calculated automatically rather than manually counted or annotated. The only component that was not automated was the speech transcription. (As automatic speech recognition technology improves, even manual transcription will become less necessary.)

Third, some previous studies (e.g., study by Kavé & Dassa[47]) investigated the relationship between disease status and each linguistic feature separately. However, this may miss important relationships between multiple linguistic features that jointly have predictive power of cognitive status, and thus we use a combination of linguistic features as predictors.

Fourth, we restricted the set of linguistic predictor variables to those which are theoretically and experimentally motivated and are human-interpretable (in contrast to, e.g., a study by Fraser et al.[31]). An automated cognitive assessment system meant to be used in conjunction with a clinician's assessment should be human-explainable to allow clinicians to understand the behavior which drove the automatic assessment.

Finally, we use the Modified Mini-Mental State Exam (3MS) rather than the MMSE as the cognitive screening test. The 3MS increases the score range compared to the MMSE, allowing for greater sensitivity to impairment by including items that test a broader range of cognitive functions. The 3MS has been shown to be better at identifying dementia than the MMSE—both higher sensitivity (detecting true positives) and also higher specificity (detecting true negatives), and is more internally consistent.[49,50]

Methods

Participants

Thirty-nine older adults living in a senior living community participated in the experiment. Inclusion criteria were chosen to produce a range of cognitive dysfunction, while excluding persons with speech deficits due to nondegenerative conditions. Specifically, the following inclusion criteria were used (ascertained by participant self-report): 60 to 90 years of age, English was their primary (or only) language, and no history of neurological disorders other than MCI or "early dementia" (e.g., no stroke or traumatic brain injury), no history of

Table 1. Demographic Characteristics of the Participant Cohort at Session 1 and Session 2.

Characteristic	Session 1 (N = 39)	Session 2 (N = 30)
Age: mean ± SD (range)	81.2 ± 6.0 (68 to 90)	81.1 ± 5.8 (70 to 91)
Gender	12 male; 27 female	11 male; 19 female
Education: mean (range)	16.6 years (12 to 21)	16.6 years (12 to 21)
3MS raw score: mean ± SD (range)	94.6 ± 7.6 (61 to 100)	94.0 ± 11.3 (48 to 100)
3MS t score: mean ± SD (range)	56.7 ± 9.4 (19 to 73)	52.1 ± 21.9 (−40 to 65)
Ppts impaired on 3MS	5.1%	3.4%
Ppts diagnosed with MCI	5.1%	3.4%
Ppts with hypertension	52.5%	58.6%
Ppts with type 2 diabetes	10.3%	20.6%
Ppts with depression	7.7%	10.3%
Ppts with family history of dementia	30.8%	37.9%

Abbreviations: MCI, mild cognitive impairment; 3MS, Modified Mini-Mental State Exam; SD, standard deviation.

developmental or severe psychiatric disorder (e.g., schizophrenia, intellectual disability), no alcohol or other substance abuse, and no advanced medical condition (e.g., current cancer, heart failure, end-stage renal disease, liver failure, etc). The average age of the participant sample was 81.2 years, was 69% female, and all participants had at least a high school-level education (see Table 1 for more detailed demographic information). At the time of enrollment, 5.1% of participants self-reported a diagnosis of MCI.

Thirty participants of the initial cohort additionally completed a second session approximately one year later (mean interval: 1.1 years, range: 0.99-1.2 years). The nine participants lost to follow-up did not differ from those who returned for the second session in demographic (age, gender, and education) or medical (MCI, hypertension, type 2 diabetes, depression, and family history of dementia) characteristics or in 3MS test performance at baseline (all $P > .05$). All participants were treated in accordance with the guidelines for ethical treatment of human subjects and provided written informed consent, approved by the Kent State University Institutional Review Board (study # 17-330).

Assessment

The duration of each testing session was approximately 75 minutes, including 3 minutes for the picture description task and 8 to 10 minutes for the expository speech task (described in more detail in the following section). Participants were administered an identical neuropsychological battery in a fixed order at both study visits under the supervision of a licensed clinical neuropsychologist. Specific clinical tests included the 3MS,[51] Hopkins Verbal Learning Test,[52] Complex Figure Test,[53,54] Digit Span,[55] Trail Making Test A and B,[56] Frontal Assessment Battery,[57] Controlled Oral Word Association Test,[58] Animal Naming,[58] and Boston Naming Test—Short Form.[59] The present analyses focus on the 3MS, a brief

measure of global cognitive abilities which assesses attention, memory, executive function, and language, providing a score of cognitive status[51] with high reliability and validity.[60,61] Raw scores were used for primary analyses (range from 0-100), with higher scores reflecting better cognitive function. Normative values adjusting for age and education[62] were used to help characterize the sample, with performances falling more than 1.5 standard deviations below the mean identified as being impaired.

Speech Tasks

Each visit consisted of several tasks to elicit open-ended, monologue speech from participants. During Session 1, participants were shown three pictures (separately) and asked to describe what was happening. The first picture was the Cookie Theft from the Boston Diagnostic Aphasia Exam.[29] The other pictures were similar types of scenes, depicting a man changing a lightbulb[63] and a kitten in a tree[64]. Participants then completed two expository speech tasks. First, they listed the four most important people in their lives and discussed the person in the second position. Second, participants described a place that was meaningful to them. Participants spoke for approximately 3 minutes for each picture description and 8 to 10 minutes for each expository task.

During Session 2, participants described the same Cookie Theft picture, then discussed another important person in their life. Finally, they looked through a picture book of the fairy tale Cinderella (with the words removed) to remember the story and then retold the story from memory (following study by Saffran et al.[65]; see also study by MacWhinney et al.[66]). (Note that in the present work, we do not analyze the speech collected at Session 2, only neuropsychological test scores.)

We included multiple speech tasks because some may be better able to elicit certain linguistic features, which in turn could be differentially indicative of cognitive status.[15,25] Picture descriptions, although somewhat artificial, elicit spontaneous speech describing a standardized input and thus may facilitate cross-participant comparisons. For example, a picture in which specific concepts are extremely salient may provide a starker contrast between healthy and impaired speakers by highlighting the differences in noun and determiner production. In contrast, expository speech tasks may elicit longer and more emotionally invested responses, and more closely resemble speech production "in the wild." For example, an expository task may reduce the topic constraints and allow for a wider range of speaking time across participants and thus better capture between-participant variability in total words produced, lexical frequency, and lexical diversity.

Sixteen lexical and semantic features which are known to be affected in persons with dementia or cognitive decline were selected from prior literature investigating linguistic changes in people with these disorders. As dementia more strongly affects lexical access, particularly in early stages of the disease (as discussed in the Introduction), these features mostly comprise linguistic metrics of semantic specificity and memory.

We use "semantic specificity" to mean the degree to which speakers use more precise, specific, and content-heavy words, as opposed to vaguer or more general words. For example, using a noun ("Mary") is more specific than using a pronoun ("she"); using a lower-frequency, more precise word ("poodle") is more specific than a higher-frequency, vaguer word ("dog" or "animal"); and using a definite article ("the house") is more specific than an indefinite article ("a house") as it refers to a particular entity as opposed to any object of that type. All features were calculated automatically from the transcriptions. Part-of-speech tags were computed using the Natural Language Toolkit (NLTK; version 3.2.1[67]) in Python (version 2.7.17) and the Penn Treebank tagset.[68] Lexical frequency was calculated using the Switchboard and Fisher corpora, a collection of spoken telephone conversations jointly containing 24 million words and 1,975 hours of speech.[69-71] The linguistic features that were calculated for each spontaneous speech task at each visit are listed in Table 2, along with a description of the feature.

Audio Recording

Speech samples were recorded using a Shure SM10A head-mounted, directional (cardioid) microphone. This setup isolates the participant's speech and ensures that the recorded audio has a high signal-to-noise ratio, excluding extraneous background noise such as other people speaking, environmental sounds, or static. Recordings were manually transcribed and time-stamped off-line by trained transcribers who were blind to the participant's cognitive status, and checked by a second transcriber.

Results

Prediction of Clinical Scores From Same-Session Linguistic Features

The first step in building a system to automatically assess cognitive decline is to determine the strength of the relationship between linguistic features and clinical scores measured at the same time point. To address this question, a multiple linear regression was conducted using the set of 16 linguistic features as predictors and 3MS score as outcome variable (see Note 1).

The two types of speech tasks (picture description and expository speech) were analyzed separately. For the picture description task, each participant's set of feature values was averaged across the three pictures to reduce the influence of any one picture. Similarly, a participant's set of feature values for the two expository speech tasks was averaged.

For the picture description task, the set of linguistic predictors was correlated with cognitive score, explaining 42% of the variance in 3MS and providing a significantly better fit than the null, intercept-only model (adjusted $R = 0.65$, $F_{16,22} = 2.75$, $P = .01$). As an exploratory measure to investigate the relationship between each linguistic feature and 3MS score, separate simple regressions were calculated between each linguistic feature and 3MS; statistically significant results are discussed

Table 2. Linguistic Features Calculated for Each Spontaneous Speech Task at Each Visit.

Feature name	Description
Total number of words	Overall count of all phonological entities spoken; including real words, nonwords, and partial words[2,6,7,26,35,36,72]
Filler words	Count of filled pauses (e.g., "uh," "um," "hmm"), as a percentage of total word count[15,73]
Empty words	Count of empty words (e.g., "thing," "place," "stuff"), as a percentage of total word count[6,7,9,26,30,43]
Lexical frequency	Mean of the log of the frequency of all real words spoken[5,20,31]
Type-token ratio	Ratio of unique words (types) to total words (tokens) spoken, used as a measure of vocabulary size and lexical diversity; higher values means the speaker produced a more varied vocabulary[7,32,33,35,36]
Honoré's statistic	Measure of lexical richness/diversity based on the number of words produced exactly once; higher values mean more diverse speech. It is calculated as: $(100 * \log(\text{tokens})) / (1 - V_1/\text{types})$, where V_1 is the number of words spoken exactly once[32]
Brunet's index	Measure of lexical richness (i.e., degree of variation in vocabulary), which is less biased by text length, calculated from the total number of words produced (tokens) and the number of unique words (types); lower values mean richer speech. It is calculated as: tokens ^ types ^ (-0.165)[32,74]
Speech rate	Count of total words divided by total elapsed time of the speech (in words per second)[13,15,41–43]
Filler rate	Count of filler words divided by total elapsed time of the speech (in words per second)[45]
Definite articles	Count of uses of "the," as a percentage of total word count[8,40]
Indefinites articles	Count of uses of "a" and "an," as a percentage of total word count[40]
Pronouns	Count of pronouns, as a percentage of total word count[2,7,15,20,32,39]
Nouns	Count of nouns, as a percentage of total word count[6,8,20]
Verbs	Count of verbs, as a percentage of total word count[2,6,22,38]
Determiners	Count of determiners, as a percentage of total word count[15]
Content words	All words that are not function words (as defined by the list of stop words in NLTK), as a percentage of total word count[20,47]

in the text and all individual-feature results are presented in Appendix A. All individual-feature Ps were corrected for multiple comparisons using the Benjamini-Hochberg method.[75] A few of the individual linguistic features were significantly correlated with the same-visit 3MS score. Average lexical frequency was negatively correlated with 3MS, meaning that participants with poorer cognitive function produced higher-frequency (more common, "easier") words ($R = -0.55$, $F_{1,37} = 15.91$, $P = .005$). The use of definite articles and determiners were each positively correlated with cognitive score, such that participants with a higher 3MS (less impaired) produced a higher proportion of definite articles and determiners in their speech (definite articles: $R = 0.50$, $F_{1,37} = 12.07$, $P = .01$; determiners: $R = 0.44$, $F_{1,37} = 8.63$, $P = .03$). Finally, the number of nouns produced as a proportion of total words was also positively correlated with cognitive score ($R = 0.41$, $F_{1,37} = 7.60$, $P = .04$).

For the expository speech task, the complete set of linguistic predictors was again significantly correlated with 3MS, explaining 51% of the variance in 3MS and providing a significantly better fit than the null, intercept-only model (adjusted $R = 0.72$, $F_{16,22} = 3.49$, $P = .004$). Of the individual features, only average lexical frequency was correlated with 3MS score, again showing a negative relationship such that more impaired participants, with a lower 3MS, produced higher-frequency words ($R = -0.53$, $F_{1,37} = 14.20$, $P = .009$). Figure 1A presents the results from the multiple regression.

There were two participants with low 3MS scores relative to the larger group. When there are a small number of outliers, the best fit line can be skewed toward those outliers to reduce the overall error, producing a relatively high R^2 value even though it does a poorer job of explaining the variance of the majority of data points. Therefore, the regression was repeated with these two cases excluded (retaining participants with a 3MS score greater than 80). Findings were similar, as lexical-semantic features of participants' speech still explained a significant amount of the variance in the concurrent 3MS scores (speech from picture descriptions: adjusted $R = 0.63$, $F_{16,20} = 2.46$, $P = .03$; speech from expository tasks: adjusted $R = 0.59$, $F_{16,20} = 2.17$, $P = .05$).

Prediction of Future Clinical Scores From Earlier Linguistic Features

The second goal of the present work was to determine whether linguistic features predict future cognitive status. To that end, we assessed whether a multiple linear regression using the set of 16 linguistic features calculated on *Session 1 speech* could predict within-participant *Session 2 3MS scores*.

The Session 1 linguistic features accounted for some of the variance of future cognitive status for both speech tasks, although the expository task's linguistic features did a better job of future 3MS prediction compared to the picture description's linguistic features. The set of linguistic features calculated on the Session 1 expository tasks explained 56% of the variance in Session 2 3MS and provided a significantly better fit than the null, intercept-only model (adjusted $R = 0.75$, $F_{16,13} = 3.35$, $P = .02$), but the set of linguistic features derived from the Session 1 picture description tasks only explained 29% of the variance in Session 2 3MS (adjusted $R = 0.54$,
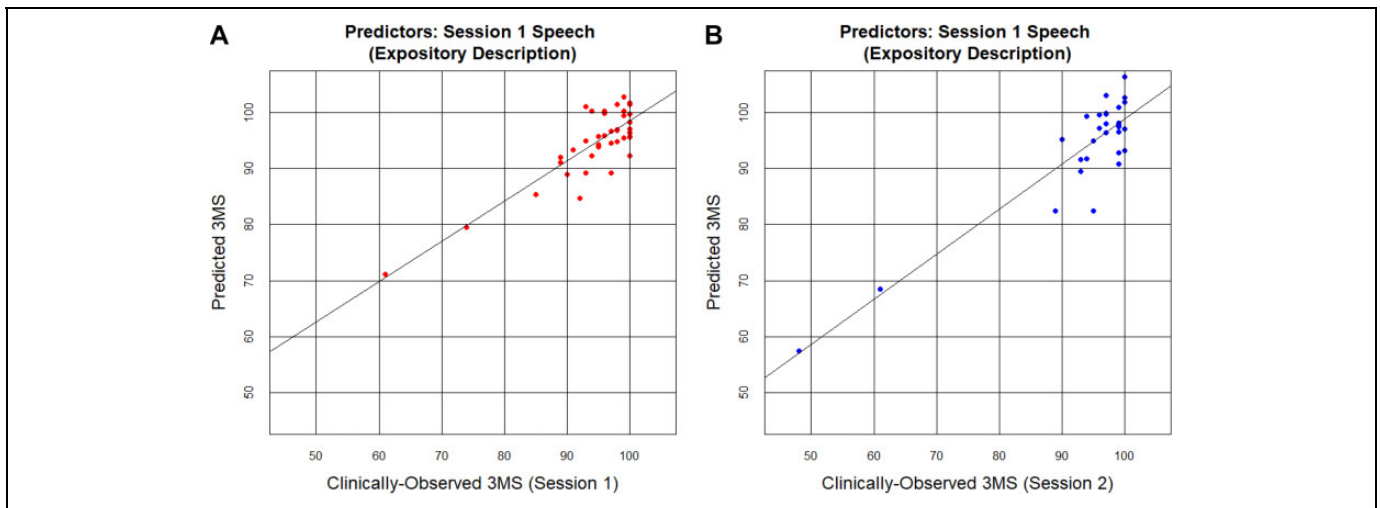
Figure 1. Actual (clinically-observed) 3MS scores plotted against predicted 3MS scores. Predicted scores were derived using the full set of 16 linguistic predictors, calculated on the expository speech tasks. Each dot represents one participant. A, Linguistic features calculated on speech from Session 1 were used as predictors for the Session 1 3MS outcome variable. B, Linguistic features calculated on speech from Session 1 were used as predictors for the Session 2 3MS outcome variable. 3MS indicates Modified Mini-Mental State Exam.

$F_{16,13} = 1.73$, $P = .16$). Notably, the Session 1 expository speech tasks explained a similar amount of variance in Session 2 clinical scores as they did for Session 1 scores (in fact, even slightly more variance), suggesting that expository speech tasks may provide an enduring window into both current and future cognitive status (see Figure 1B for the results from the multiple regression).

Similar to the relationship with concurrent cognitive status, average lexical frequency was negatively correlated with cognitive status a year into the future (speech from picture descriptions: $R = -0.51$, $F_{1,28} = 9.95$, $P = .06$; speech from expository tasks: $R = -0.57$, $F_{1,28} = 13.58$, $P = .02$). The production of definite articles showed a positive trend with future cognitive status, such that producing more definite articles today correlated with higher cognitive score one year later, but only in the picture description task (speech from picture descriptions: $R = 0.47$, $F_{1,28} = 8.09$, $P = .07$). The production of nouns, filler words, and the rate of filler word production all showed a positive trend with future cognitive status as well, but only for the expository speech tasks (nouns: $R = 0.43$, $F_{1,28} = 6.21$, $P = .08$; filler words: $R = 0.44$, $F_{1,28} = 6.74$, $P = .08$; filler rate: $R = 0.42$, $F_{1,28} = 6.10$, $P = .08$).

Discussion

The goal of the present work was to investigate whether spontaneous speech could be used as a marker of cognitive function in older adults. To do so, we automatically calculated a set of lexical-semantic features known to be impaired in persons with AD and other dementias and used them to predict concurrent and future performance on the 3MS. The linguistic features had good predictive power of the clinical scores, explaining up to 51% of the variance of the speaker's current cognitive score,

and up to 56% of the variance of the speaker's cognitive score a year in the future.

Consistent with expectations, the individual linguistic features which were the most relevant to predicting cognitive test performance were those related to the loss of semantic specificity: More impaired participants produced words with higher average lexical frequency and a lower proportion of definite articles, determiners, and nouns. The higher-frequency words produced by more impaired participants may reflect reduced vocabulary due to word-finding difficulty. Similarly, decreased use of definite articles has been suggested to indicate diminished declarative memory,[40] a hallmark of AD.

An interesting possibility arising from the present work is that some types of spontaneous speech elicitation tasks may be better-suited than others for prediction of cognitive function. In the current work, both speech tasks (picture description and personal expositories) showed high predictive abilities for concurrent 3MS scores (42% and 51%, respectively, both statistically significant). In contrast, only speech from the expository task was able to predict 3MS scores at 12-month follow-up (29% and 56%, respectively, with only the expository task statistically significant). As Sajjadi and colleagues[15] note, a picture description task is much more highly structured and constrained than is an open-ended expository question (see also study by Boschi et al.[25]). As a result, picture descriptions are likely better-suited to elicit nouns referring to the particular items shown in the picture, and definite articles to signal that a particular woman or sink or cookie jar is being referenced, and thus may more starkly show deficits on those metrics as speakers replace nouns with pronouns or related but incorrect nouns, or definite articles with indefinite articles or no determiner at all. In contrast, the expository task offers more flexibility in topic, content, and word choice, and thus may elicit a more accurate distribution

of that speaker's range of lexical frequency and syntactic abilities. As a result, an expository interview may be a better method for assessing a participant's spontaneity or lack thereof[15] and thus have better predictive ability of *future* cognitive status. Relatedly, producing a picture description likely imposes lower load on the cognitive abilities which are affected in MCI and AD (e.g., working memory, executive function, and episodic memory[76,77]) than do more open-ended narrative speech tasks. As a result, the expository task may be more sensitive to early-stage impairment as it is more taxing to the relevant cognitive systems. In future work with a larger participant sample and a longer time horizon, it will be important to continue to explore the relative predictive ability of these different speech elicitation tasks.

A related consideration for future work is that different linguistic features may be correlated with cognitive test performance at different stages of decline. For example, past work suggests that in early stages of AD, lexical access is particularly impaired, but in more advanced stages, syntactic production begins to show impairment as well.[35,37,46] One potential application of this pattern is quantifying and monitoring these changes within individuals over time from preclinical throughout multiple disease stages, and using different linguistic features or tasks for people at different stages of the disease, as some features may be more predictive in separating healthy speakers from people with early-stage dementia, while others might be more predictive for placing speakers along the cognitive status continuum at the more impaired end. This approach would help with early detection but also inform the effectiveness of interventions and the need for additional services in persons with dementia (e.g., assisted living and guardianship). Studies with larger samples and longer follow-up intervals are needed to help examine this possibility, particularly those that perform comprehensive assessments to document clinical status.

Identifying which features of spontaneous speech are predictive of neuropsychological status could also further our understanding of how different diseases affect the organization of linguistic knowledge in the brain. To that end, it is possible that different profiles of speech deficits may correspond to specific neurodegenerative disorders. Future between-subject group studies comparing healthy controls to patients with well-characterized clinical conditions—such as those which informed the selection of the linguistic features used here—could reveal which linguistic features are impacted by which neurodegenerative disorders and at what stages of disease progression. For example, persons with various forms of frontotemporal dementia and chronic traumatic encephalopathy exhibit early changes in spontaneous speech[78,79] that could likely be detected using similar speech analyses. Similarly, future work should also examine the association between spontaneous speech indices and findings from advanced neuroimaging. As an example, the present work shows that average lexical frequency is a strong indicator of future cognitive status among relatively unimpaired older adults. A future study clarifying the extent to which this relationship is attributable to global or selective atrophy (e.g., the hippocampus), cerebrovascular disease, or amyloid deposition[80,81] would provide important insight into the underlying neurological processes for speech production in older adults and clarify the predictive value of spontaneous speech for future neurological outcomes.

In addition to these scientific benefits, the present methodology has the potential to be applied in concert with existing clinical methods to increase screening frequency and improve monitoring of individuals at risk for cognitive decline. Unlike the administration of a traditional cognitive screening test, it is possible to collect speech data on a monthly or even weekly basis with little participant burden.[82] Spontaneous speech could be assessed in an ongoing manner and utilized as an early-warning system: If the clinical score predicted on the basis of a person's linguistic features drops below a threshold, their doctor or family members could be notified, allowing the person to undergo formal, clinical assessment. Moreover, speech deficits are associated with numerous poor outcomes across a range of neurological conditions, including reduced or impaired activities of daily living, greater depression, and poorer quality of life.[83-88] Early detection and monitoring of speech may lead to improved outcomes across these domains. In particular, loss of language and communication abilities by patients with MCI and AD is especially challenging for family members. In fact, communication loss is the most frequently generated response from family members when asked about difficulties in caring for impaired relatives, even outweighing the need to perform the activities of daily living for their relative.[16] The ability to objectively quantify this breakdown via frequent, automated analyses, and use that quantification to inform both clinical care and strategies used by family members may ultimately help to ease caregiver burden.

The present research demonstrates an initial proof-of-concept that open-ended narrative speech can be collected and, after transcription, be analyzed entirely automatically with significant explanatory power of the speaker's present and future cognitive status. Further research must be conducted with a larger sample size to validate the predictive ability of the present linguistic features and to expand the automated analyses to transcription as well. This approach may lead to earlier detection and intervention for conditions like AD, MCI, and other dementias, potentially influencing their trajectory.

Appendix A

Table A1. Correlation Between Session 1 Linguistic Features in Each of the Two Types of Speech Tasks With Session 1 3MS Score.[a]

	Session 1 picture description		Session 1 expository task	
	R	P	R	P
Total words	0.37	.07[♦]	0.30	.19
Filler words	0.28	.14	0.34	.18
Empty words	0.02	.90	−0.03	.96
Definite articles	0.50	.01**	0.21	.38
Indefinite articles	0.18	.31	0.01	.96
Pronouns	0.20	.31	0.19	.40
Nouns	0.41	.04*	0.34	.18
Verbs	0.19	.31	0.26	.26
Determiners	0.44	.03*	0.19	.40
Content words	0.33	.10[♦]	0.29	.19
Lexical frequency	−0.55	.01**	−0.53	.01**
Type-token ratio	−0.32	.10[♦]	0.06	.96
Honoré's statistic	0.31	.10[♦]	0.17	.44
Brunet's index	0.35	.07[♦]	0.04	.96
Speech rate	0.09	.63	0.01	.96
Filler rate	0.19	.31	0.32	.19

[a]All P values were corrected for multiple comparisons using the Benjamini-Hochberg[75] method.
[♦]$P \leq .10$.
*$P \leq .05$.
**$P \leq .01$.

Table A2. Correlation Between Session 1 Linguistic Features in Each of the Two Types of Speech Tasks With Session 2 3MS Score.[a]

	Session 1 picture description		Session 1 expository task	
	R	P	R	P
Total words	0.23	.50	0.29	.33
Filler words	0.18	.57	0.44	.08[♦]
Empty words	−0.04	.95	0.14	.63
Definite articles	0.47	.07[♦]	0.18	.57
Indefinite articles	0.12	.74	0.04	.89
Pronouns	0.02	.97	0.14	.63
Nouns	0.29	.45	0.43	.08[♦]
Verbs	0.04	.95	0.23	.50
Determiners	0.40	.16	0.19	.54
Content words	0.19	.55	0.30	.33
Lexical frequency	−0.51	.06[♦]	−0.57	.02*
Type-token ratio	−0.21	.53	−0.04	.89
Honoré's statistic	0.28	.45	0.22	.50
Brunet's index	0.24	.50	0.03	.89
Speech rate	0.01	.97	−0.09	.77
Filler rate	0.11	.74	0.42	.08[♦]

[a]All P values were corrected for multiple comparisons using the Benjamini-Hochberg[75] method.
[♦]$P \leq .10$.
*$P \leq .05$.

ORCID iD

Rachel Ostrand ⓘ https://orcid.org/0000-0001-5491-7656

Note

1. There were not enough participants and data points to allow for feature selection on an isolated training set to be applied to a test set, and thus, as we had a priori expectations of a possible relationship between each of the linguistic variables and cognitive status, we opted to include the entire set of 16 linguistic features together as predictors. There is substantial collinearity between some of the linguistic predictors. However, this works *against* the strength of the correlation result, as it means that more predictors may have been included in the correlation than are needed to explain the explainable variance. The statistical measure used, adjusted R^2, penalizes for each additional predictor. Therefore, including more collinear predictors means that the reported correlation values likely *underrepresent* the amount of variance that is explained by the predictors. Additionally, the collinearity between predictors means that the β values in the multiple regression are not necessarily meaningful indicators of each predictor's contribution to the outcome variable, and thus we opted to additionally compute individual regressions between each linguistic predictor variable and the outcome variable.

References

1. Hebert LE, Weuve J, Scherr PA, Evans DA. Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. *Neurology*. 2013;80(19):1778-1783. doi:10.1212/WNL.0b013e31828726f5
2. Kemper S, Altmann LJP. Dementia and Language. In: *Reference Module in Neuroscience and Biobehavioral Psychology*. Elsevier; 2017. doi:10.1016/b978-0-12-809324-5.01884-8

3. Burke DM, Shafto MA. Language and aging. In: *The Handbook of Aging and Cognition*. 3 rd ed. Psychology Press; 2008:373-443.

4. Obler L, Albert ML. Language in aging. In: *Clinical Neurology of Aging*. 1st ed. Oxford University Press; 1984:245-253.

5. Bird H, Lambon Ralph MA, Patterson K, Hodges JR. The rise and fall of frequency and imageability: noun and verb production in semantic dementia. *Brain Lang*. 2000;73(1):17-49. doi:10.1006/brln.2000.2293

6. Croisile B, Ska B, Brabant MJ, et al. Comparative study of oral and written picture description in patients with Alzheimer's disease. *Brain Lang*. 1996;53(1):1-19. doi:10.1006/brln.1996.0033

7. Hier DB, Hagenlocker K, Shindler AG. Language disintegration in dementia: effects of etiology and severity. *Brain Lang*. 1985;25(1):117-133. doi:10.1016/0093-934x(85)90124-5

8. Feyereisen P, Berrewaerts J, Hupet M. Pragmatic skills in the early stages of Alzheimer's disease: an analysis by means of a referential communication task. *Int J Lang Commun Disord*. 2007;42(1):1-17. doi:10.1080/13682820600624216

9. Nicholas M, Obler LK, Albert ML, Estabrooks NH. Empty speech in Alzheimer's disease and fluent aphasia. *J Speech Hear Res*. 1985;28(3):405-410. doi:10.1044/jshr.2803.405

10. Bayles KA, Tomoeda CK, Trosset MW. Relation of linguistic communication abilities of Alzheimer's patients to stage of disease. *Brain Lang*. 1992;42(4):454-472. doi:10.1016/0093-934x(92)90079-t

11. McKay KF, Shanks MF, Venneri A. Profiling spontaneous speech decline in Alzheimer's disease: a longitudinal study. *Acta Neuropsychiatr*. 2013;25(6):320-327. doi:10.1017/neu.2013.16

12. Salvatierra J, Rosselli M, Acevedo A, Duara R. Verbal fluency in bilingual Spanish/English Alzheimer's disease patients. *Am J Alzheimers Dis Other Demen*. 2007;22(3):190-201. doi:10.1177/1533317507301792

13. Hoffmann I, Nemeth D, Dye CD, Pákáski M, Irinyi T, Kálmán J. Temporal parameters of spontaneous speech in Alzheimer's disease. *Int J Speech Lang Pathol*. 2010;12(1):29-34. doi:10.3109/17549500903137256

14. Filiou RP, Bier N, Slegers A, Houzé B, Belchior P, Brambati SM. Connected speech assessment in the early detection of Alzheimer's disease and mild cognitive impairment: a scoping review. *Aphasiology*. 2019:1-33. doi:10.1080/02687038.2019.1608502

15. Sajjadi SA, Patterson K, Tomek M, Nestor PJ. Abnormalities of connected speech in semantic dementia vs Alzheimer's disease. *Aphasiology*. 2012;26(6):847-866. doi:10.1080/02687038.2012.654933

16. Murray J, Schneider J, Banerjee S, Mann A. EUROCARE: a cross-national study of co-resident spouse carers for people with Alzheimer's disease: II—a qualitative analysis of the experience of caregiving. *Int J Geriatr Psychiatry*. 1999;14(8):662-667. doi:10.1002/(SICI)1099-1166(199908)14:8<662:: AID-GPS993>3.0.CO;2-4

17. Hart S. Language and dementia: a review. *Psychol Med*. 1988;18(1):99-112. doi:10.1017/s0033291700001938

18. Martin A, Fedio P. Word production and comprehension in Alzheimer's disease: the breakdown of semantic knowledge. *Brain Lang*. 1983;19(1):124-141. doi:10.1016/0093-934x(83)90059-7

19. Taler V, Phillips NA. Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review. *J Clin Exp Neuropsychol*. 2008;30(5):501-556. doi:10.1080/13803390701550128

20. Kavé G, Goral M. Word retrieval in picture descriptions produced by individuals with Alzheimer's disease. *J Clin Exp Neuropsychol*. 2016;38(9):958-966. doi:10.1080/13803395.2016.1179266

21. Henry JD, Crawford JR, Phillips LH. Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis. *Neuropsychologia*. 2004;42(9):1212-1222. doi:10.1016/j.neuropsychologia.2004.02.001

22. Beber BC, da Cruz AN, Chaves ML. A behavioral study of the nature of verb production deficits in Alzheimer's disease. *Brain Lang*. 2015;149:128-134. doi:10.1016/j.bandl.2015.07.010

23. Kempler D, Curtiss S, Jackson C. Syntactic preservation in Alzheimer's disease. *J Speech Lang Hear Res*. 1987;30(3):343-350. doi:10.1044/jshr.3003.343

24. Tsantali E, Economidis D, Tsolaki M. Could language deficits really differentiate mild cognitive impairment (MCI) from mild Alzheimer's disease? *Arch Gerontol Geriatr*. 2013;57(3):263-270. doi:10.1016/j.archger.2013.03.011

25. Boschi V, Catricalà E, Consonni M, Chesi C, Moro A, Cappa SF. Connected speech in neurodegenerative language disorders: a review. *Front Psychol*. 2017;8:269. doi:10.3389/fpsyg.2017.00269

26. Kavé G, Goral M. Word retrieval in connected speech in Alzheimer's disease: a review with meta-analyses. *Aphasiology*. 2018;32(1):4-26. doi:10.1080/02687038.2017.1338663

27. Slegers A, Filiou RP, Montembeault M, Brambati SM. Connected speech features from picture description in Alzheimer's disease: a systematic review. *J Alzheimers Dis*. 2018;65(2):519-542. doi:10.3233/JAD-170881

28. Mueller KD, Hermann B, Mecollari J, Turkstra LS. Connected speech and language in mild cognitive impairment and Alzheimer's disease: a review of picture description tasks. *J Clin Exp Neuropsychol*. 2018;40(9):917-939. doi:10.1080/13803395.2018.1446513

29. Goodglass H, Kaplan E. *The Assessment of Aphasia and Related Disorders*. 2nd ed. Lea & Febiger; 1983.

30. Ehrlich JS, Obler LK, Clark L. Ideational and semantic contributions to narrative production in adults with dementia of the Alzheimer's type. *J Commun Disord*. 1997;30(2):79-99. doi:10.1016/0021-9924(95)00053-4

31. Fraser KC, Meltzer JA, Rudzicz F. Linguistic features identify Alzheimer's disease in narrative speech. *J Alzheimers Dis*. 2016;49(2):407-422. doi:10.3233/jad-150520

32. Bucks RS, Singh S, Cuerden JM, Wilcock GK. Analysis of spontaneous, conversational speech in dementia of Alzheimer type: evaluation of an objective technique for analysing lexical performance. *Aphasiology*. 2000;14(1):71-91. doi:10.1080/026870300401603

33. Dijkstra K, Bourgeois MS, Allen RS, Burgio LD. Conversational coherence: discourse analysis of older adults with and without dementia. *J Neurolinguistics*. 2004;17(4):263-283. doi:10.1016/S0911-6044(03)00048-4

34. Mueller KD, Koscik RL, Turkstra LS, et al. Connected language in late middle-aged adults at risk for Alzheimer's disease. *J Alzheimers Dis*. 2016;54(4):1539-1550. doi:10.3233/JAD-160252

35. Davis BH, Maclagan M. Examining pauses in Alzheimer's discourse. *Am J Alzheimers Dis Dementiasr*. 2009;24(2):141-154. doi:10.1177/1533317508328138

36. Dijkstra K, Bourgeois M, Petrie G, Burgio L, Allen-Burge R. My recaller is on vacation: discourse analysis of nursing-home residents with dementia. *Discourse Process*. 2002;33(1):53-76. doi:10.1207/S15326950DP3301_03

37. Ahmed S, Haigh AMF, de Jager CA, Garrard P. Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain*. 2013;136(12):3727-3737. doi:10.1093/brain/awt269

38. Kemper S, LaBarge E, Ferraro FR, Cheung H, Cheung H, Storandt M. On the preservation of syntax in Alzheimer's disease. Evidence from written sentences. *Arch Neurol*. 1993;50(1):81-86. doi:10.1001/archneur.1993.00540010075021

39. Almor A, Kempler D, MacDonald MC, Andersen ES, Tyler LK. Why do Alzheimer patients have difficulty with pronouns? Working memory, semantics, and reference in comprehension and production in Alzheimer's disease. *Brain Lang*. 1999;67(3):202-227.

40. Duff MC, Gupta R, Hengst JA, Tranel D, Cohen NJ. The use of definite references signals declarative memory: evidence from patients with hippocampal amnesia. *Psychol Sci*. 2011;22(5):666-673. doi:10.1177/0956797611404897

41. Carlomagno S, Santoro A, Menditti A, Pandolfi M, Marini A. Referential communication in Alzheimer's type dementia. *Cortex*. 2005;41(4):520-534. doi:10.1016/S0010-9452(08)70192-8

42. Ash S, Moore P, Vesely L, Grossman M. The decline of narrative discourse in Alzheimer's disease. *Brain Lang*. 2007;103(1):181-182. doi:10.1016/j.bandl.2007.07.105

43. Forbes KE, Venneri A, Shanks MF. Distinct patterns of spontaneous speech deterioration: an early predictor of Alzheimer's disease. *Brain Cogn*. 2002;48(2):356-361. doi:10.1006/brcg.2001.1377

44. Giles E, Patterson K, Hodges JR. Performance on the Boston Cookie theft picture description task in patients with early dementia of the Alzheimer's type: missing information. *Aphasiology*. 1996;10(4):395-408. doi:10.1080/02687039608248419

45. Murray LL. Distinguishing clinical depression from early Alzheimer's disease in elderly people: can narrative analysis help? *Aphasiology*. 2010;24(6-8):928-939. doi:10.1080/02687030903422460

46. Snowdon DA, Kemper SJ, Mortimer JA, Greiner LH, Wekstein DR, Markesbery WR. Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun study. *JAMA*. 1996;275(7):528-532.

47. Kavé G, Dassa A. Severity of Alzheimer's disease and language features in picture descriptions. *Aphasiology*. 2018;32(1):27-40. doi:10.1080/02687038.2017.1303441

48. Asgari M, Kaye J, Dodge H. Predicting mild cognitive impairment from spontaneous spoken utterances. *Alzheimers Dement Transl Res Clin Interv*. 2017;3(2):219-228. doi:10.1016/j.trci.2017.01.006

49. McDowell I, Kristjansson B, Hill GB, Hébert R. Community screening for dementia: the Mini Mental State Exam (MMSE) and Modified Mini-Mental State Exam (3MS) compared. *J Clin Epidemiol*. 1997;50(4):377-383. doi:10.1016/s0895-4356(97)00060-7

50. Tombaugh TN, McDowell I, Kristjansson B, Hubley AM. Minimental state examination (MMSE) and the modified MMSE (3MS): a psychometric comparison and normative data. *Psychol Assess*. 1996;8(1):48-59.

51. Teng EL, Chui HC. The Modified Mini-Mental State (3ms) Examination. *J Clin Psychiatry*. 1987;48(8):314-318.

52. Brandt J, Benedict RHB. *Hopkins Verbal Learning Test–Revised: Professional Manual*. Psychological Assessment Resources; 2001.

53. Meyers JE, Meyers KR. *Rey Complex Figure Test and Recognition Trial—Professional Manual*. Psychological Assessment Resources; 1995.

54. Berry DTR, Allen RS, Schmitt FA. Rey-Osterrieth complex figure: psychometric characteristics in a geriatric sample. *Clin Neuropsychol*. 1991;5(2):143-153. doi:10.1080/13854049108403298

55. Wechsler D. *Wechsler Adult Intelligence Scale—Third Edition*. The Psychological Corporation; 2008.

56. Reitan RM. Validity of the trail making test as an indicator of organic brain damage. *Percept Mot Skills*. 1958;8(3):271-276. doi:10.2466/pms.1958.8.3.271

57. Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: a frontal assessment battery at bedside. *Neurology*. 2000;55(11):1621-1626. doi:10.1212/WNL.55.11.1621

58. Lezak MD, Howieson DB, Loring DW. *Neuropsychological Assessment*. 4 ed. Oxford University Press; 2004.

59. Williams BW, Mack W, Henderson VW. Boston naming test in Alzheimer's disease. *Neuropsychologia*. 1989;27(8):1073-1079. doi:10.1016/0028-3932(89)90186-3

60. Bassuk SS, Murphy JM. Characteristics of the Modified Mini-Mental State Exam among elderly persons. *J Clin Epidemiol*. 2003;56(7):622-628. doi:10.1016/S0895-4356(03)00111-2

61. Grace J, Nadler JD, White DA, et al. Folstein vs Modified Mini-Mental State Examination in geriatric stroke: stability, validity, and screening utility. *Arch Neurol*. 1995;52(5):477-484. doi:10.1001/archneur.1995.00540290067019

62. Jones TG, Schinka JA, Vanderploeg RD, Small BJ, Graves AB, Mortimer JA. 3MS normative data for the elderly. *Arch Clin Neuropsychol*. 2002;17(2):171-177. doi:10.1016/s0887-6177(00)00108-6

63. Marshall RC, Wright HH. Developing a clinician-friendly aphasia test. *Am J Speech Lang Pathol*. 2007;16(4):295-315. doi:10.1044/1058-0360(2007/035)

64. Nicholas LE, Brookshire RH. A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *J Speech Hear Res*. 1993;36(2):338-350. doi:10.1044/jshr.3602.338

65. Saffran EM, Berndt RS, Schwartz MF. The quantitative analysis of agrammatic production: procedure and data. *Brain Lang*. 1989;37(3):440-479. doi:10.1016/0093-934X(89)90030-8

66. MacWhinney B, Fromm D, Holland A, Forbes M, Wright H. Automated analysis of the Cinderella story. *Aphasiology* 2010;24:856-868. doi:10.1080/02687030903452632

67. Bird S, Klein E, Loper E. *Natural Language Processing with Python*. 1st ed. O'Reilly Media; 2009.

68. Marcus MP, Santorini B, Marcinkiewicz MA. Building a large annotated corpus of English: the Penn Treebank. *Comput Linguist*. 1993;19(2):313-330.

69. Godfrey J, Holliman E. *Switchboard-1 Release 2 LDC97S62*. Linguistic Data Consortium; 1993.

70. Cieri C, Graff D, Kimball O, Miller D, Walker K. *Fisher English Training Speech Part 1 Transcripts LDC2004T19*. Linguistic Data Consortium; 2004.

71. Cieri C, Graff D, Kimball O, Miller D, Walker K. *Fisher English Training Part 2, Transcripts LDC2005T19*. Linguistic Data Consortium; 2005.

72. Tomoeda CK, Bayles KA, Trosset MW, Azuma T, McGeagh A. Cross-sectional analysis of Alzheimer disease effects on oral discourse in a picture description task. *Alzheimer Dis Assoc Disord*. 1996;10(4):204-215.

73. Bortfeld H, Leon SD, Bloom JE, Schober MF, Brennan SE. Disfluency rates in conversation: effects of age, relationship, topic, role, and gender. *Lang Speech*. 2001;44(2):123-147. doi:10.1177/00238309010440020101

74. Brunet É. *Le Vocabulaire de Jean Giraudoux : structure et évolution : statistique et informatique appliquées à l'étude des textes à partir des données du Trésor de la langue française*. Slatkine; 1978.

75. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol*. 1995;57(1):289-300.

76. Kirova AM, Bays RB, Lagalwar S. Working memory and executive function decline across normal aging, mild cognitive impairment, and Alzheimer's disease. *BioMed Res Int*. doi:10.1155/2015/748212

77. Albert M, Moss MB, Blacker D, Tanzi R, McArdle JJ. Longitudinal change in cognitive performance among individuals with mild cognitive impairment. *Neuropsychology*. 2007;21(2):158-169. doi:10.1037/0894-4105.21.2.158

78. Olney NT, Spina S, Miller BL. Frontotemporal dementia. *Neurol Clin*. 2017;35(2):339-374. doi:10.1016/j.ncl.2017.01.008

79. Meehan W, Mannix R, Zafonte R, Leone AP. Chronic traumatic encephalopathy and athletes. *Neurology*. 2015;85(17):1504-1511. doi:10.1212/WNL.0000000000001893

80. Ossenkoppele R, CohnSheehy BI, Joie RL, et al. Atrophy patterns in early clinical stages across distinct phenotypes of Alzheimer's disease. *Hum Brain Mapp*. 2015;36(11):4421-4437. doi:10.1002/hbm.22927

81. Verfaillie SCJ, Witteman J, Slot RER, et al. High amyloid burden is associated with fewer specific words during spontaneous speech in individuals with subjective cognitive decline. *Neuropsychologia*. 2019;131:184-192. doi:10.1016/j.neuropsychologia.2019.05.006

82. Ostrand R, Berger S, Cecchi G. Personalized, longitudinal assessment of cognitive disorders using spontaneous speech production at home. Paper presented at the: Northeast Computational Health Summit; April 2019; Providence, RI.

83. Wolff L, Benge J. Everyday language difficulties in Parkinson's disease: caregiver description and relationship with cognition, activities of daily living, and motor disability. *Am J Speech Lang Pathol*. 2019;28(1):165-173. doi:10.1044/2018_AJSLP-18-0091

84. Kashibayashi T, Ikeda M, Komori K, et al. Transition of distinctive symptoms of semantic dementia during longitudinal clinical observation. *Dement Geriatr Cogn Disord*. 2010;29(3):224-232. doi:10.1159/000269972

85. Gialanella B, Prometti P, Vanoglio F, Comini L, Santoro R. Aphasia and activities of daily living in stroke patients. *Eur J Phys Rehabil Med*. 2016;52(6):782-790.

86. Lee H, Lee Y, Choi H, Pyun SB. Community integration and quality of life in aphasia after stroke. *Yonsei Med J*. 2015;56(6):1694-1702. doi:10.3349/ymj.2015.56.6.1694

87. Lazar RM, Boehme AK. Aphasia as a predictor of stroke outcome. *Curr Neurol Neurosci Rep*. 2017;17(11):83. doi:10.1007/s11910-017-0797-z

88. Jang J, Jang J, Cushing N, Clemson L, Hodges JR, Mioshi E. Activities of daily living in progressive non-fluent aphasia, logopenic progressive aphasia and Alzheimer's disease. *Dement Geriatr Cogn Disord*. 2012;33(5):354-360. doi:10.1159/000339670