JOURNAL OF
EMERGING INVESTIGATORS

Article

Decline in vocabulary richness in individuals with Alzheimer's disease

Andi Liu[1], Tanya Dash[2]
[1] Louis Reil School, Calgary, Canada
[2] Department of Communication Sciences and Disorders, University of Alberta, Edmonton, Canada

SUMMARY

Alzheimer's disease (AD) is a progressive neurodegenerative disease that is projected to affect 955,900 Canadians by 2030. AD causes the buildup of amyloid-beta plaques and tau tangles in the brain. As a result, individuals diagnosed often face language impairments such as word finding difficulties (anomia), increased repetition, perseverations, diminished vocabularies, echolalia, and verbal apraxia. Characterizing language impairment in AD may aid in the development of non-invasive diagnostic tools using machine learning. In this study, we assessed vocabulary richness in individuals with and without AD. We hypothesized that individuals with AD would have decreased vocabulary richness compared to those without due to language impairment. Our study used publicly available data from DementiaBank, specifically the Pitt corpus, which contains interview transcripts of 136 participants describing the "Cookie Theft" picture from the Boston Diagnostic Aphasia Examination. Results indicate that Type Token Ratio, a measure of vocabulary richness, was lower in participants with AD when compared to participants without AD based on picture description transcripts ($p=0.007$). However, the effect size was found to be small to medium ($d=-0.46$). Language impairments occurred less frequently in individuals with AD than hypothesized likely because the picture description task did not sufficiently engage memory. Even though the results we observed came from tasks that were not memory dependent, vocabulary richness may still be a useful indicator of AD and may be useful in automated detection of AD through language.

INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative condition that is characterized by progressive cognitive impairment, a decline in cognition and independence in daily activities, and abnormally impulsive and violent behavior (1–3). AD is projected to affect 955,900 Canadians by 2030, and one in three Canadians are currently caregivers of someone with the disease (4).

AD progression can be classified into four stages. In the preclinical stage, there are no noticeable symptoms and no signs of impairment. In the early stage, symptoms such as loss of concentration and memory, disorientation, mood swings, and depression occur. The moderate AD stage is characterized by trouble recognizing family and friends, loss of impulse control, and difficulty reading and writing. Severe AD occurs when the disease has spread to the entire cortex, at which point the patient cannot recognize their family and has difficulties swallowing and urinating. The most common cause of death from AD is respiratory illness (5-6).

Individuals with AD often have language impairments, including difficulties finding words, diminished vocabulary, anomia (difficulty retrieving words), echolalia (repetition of words and phrases), and verbal apraxia (inability to articulate words) (7-10). Language deficits can lead to increased isolation, depression, and anxiety. In more moderate phases of the disease, the affected person is not able to remember their thoughts long enough to express them. They may repeat words, sounds, or sentences to make themselves understood. The most common types of linguistic impairments are vocabulary impairments. Individuals tend to substitute words with superordinate words and use circumlocutory speech (11). In addition, a reduction in semantic granularity (informativeness of a word) and an increase in semantic variability (conceptual closeness of successive words) can be observed as the disease progresses (12). AD patients were found to have significantly higher variability scores than healthy controls. AD patients also exhibited more use of low granularity words than healthy individuals (13).

Seventy-five percent of AD cases worldwide are undiagnosed (14). However, patients who are diagnosed early can benefit more from clinical trials and lifestyle modifications to avoid risk factors, potentially delaying the progression of the disease (15). While there are also disease-modifying drugs available for AD, they are often are not effective in treating AD because the drugs are administered when symptoms have already progressed (16). Accumulations of amyloid-beta and tau, which are hallmarks of AD, occur decades before symptoms of AD are apparent. Disease-modifying drugs may not be able to reverse neurodegeneration by the time AD is diagnosed (16). Early diagnosis of AD not only can make patients eligible to participate in clinical trials and research that provide therapies intended to halt AD or lessen symptoms, but also can relieve the stress family members and caregivers may experience following the diagnosis (17). Diagnoses at a later stage complicate treatment because the brain may be irreparably damaged by the time a diagnosis is received (18). With the economic burden of AD in Canada reaching 40.1 million CAD in 2020, there is a growing need for effective and accessible methods to detect AD (19). As the prevalence of AD continues to rise, it is crucial to understand current diagnostic methods and design new ones that are more cost-effective and accessible.

https://doi.org/10.59720/24-289

Currently, the most common diagnostic strategies for AD use positron emission tomography (PET), magnetic resonance imaging (MRI), and biomarkers in cerebrospinal fluid (CSF) (20). These methods are useful in assessing the levels of amyloid-beta and tau proteins, which are strongly associated with the development of AD. Detection of AD diagnostic biomarkers that rely on pathologic features related to amyloid-beta and tau require highly specialized training and personnel, rendering these strategies expensive and inaccessible to healthcare centers (20).

Numerous alternatives have been proposed to the above mentioned diagnostic techniques (20). Machine learning is used to diagnose AD from MRI scans by detecting structural changes in the brain. Models such as convolutional neural networks convert the pixels of MRI scans into image features like edges. The model then decides the diagnosis based on the feature extracted from the scan. Biomarker prediction uses models to detect AD from clinical variables, such as gender, age, body mass index (BMI), sleep cycles, and the amount of tau and amyloid-beta found in CSF (21-23). While these methods can be effective in detecting AD, we chose to focus on diagnosis from speech impairment because of its accessibility. MRI scans, PET scans, and CSF analysis used in machine learning are labor-intensive, expensive, uncomfortable for the patient, and not practical for large scale screening of AD. Language is a cheaper and more convenient medium to diagnose AD (24).

Assessment of spontaneous speech is one of the most promising approaches for early detection of cognitive decline. It allows for automated analysis of lexical-semantic features that are closely associated with both current and future cognitive status and is a scalable and efficient method for tracking cognitive health over time (20). Spontaneous speech is speech produced with the intention of communicating a message by the speaker. Spontaneous speech tasks are common in detecting language impairments because they reflect how impairments are presented in everyday conversations (25). Lexical-semantic features measure word finding and semantic knowledge in language. Examples of features include: the amount of information conveyed in speech, the number of vague words, the number of words with a more general meaning, and the number of indefinite articles (20). Language impairment is often apparent before dementia is diagnosed. Language ability is also associated with memory and attention. Spontaneous speech has the potential to indicate an individual's cognitive performance, and the use of machine learning techniques may be a diagnostic tool for AD detection (26).

Common parts of speech features used in speech detection include the number of unique words, the number of filler words, the number of incomplete words, and the Type Token Ratio (TTR) of a speech sample. Filler words, such as "uh", "er", and "um" represent a speaker's lapse in thought. Theoretically, individuals with AD would use filler words more often than individuals without AD because of increased cognitive impairment. Incomplete words indicate a similar lapse (27). TTR is defined by the equation Type/Token, where Type is the number of unique words and Token is the number of total words in a body of text (28). Language impairment includes declining vocabulary, so the justification for using TTR to diagnose AD is that individuals with AD will have poorer vocabulary richness as a symptom of their cognitive decline. Supporting this, Guinn et al. analyzed conversational samples of participants with and without AD and found that TTR was decreased in participants with AD (mean=0.406 ± 0.114) compared to healthy participants (mean TTR= 0.414 ±0.140) (27).

Yamada et al. measured linguistic features in the language of 2 AD and 13 control participants and found that topic (d=-1.76) and word repetition (d=-1.67) in two different conversations over two separate days, and topic (d=-1.08) and word repetition (d=-0.80) in single conversations, had a large effect size (29). Effect size measures the magnitude of difference between groups, and the results presented by Yamada et al. indicate that there can be a large magnitude of difference in topic and word repetition between individuals with AD and individuals without AD. Features with a larger effect size may be a better indicator of impairment than others because the difference between the two groups is more obvious, so comparing effect sizes is comparing the effectiveness of TTR to other features (29).

Our project was the continuation of a previous project. We found that individuals with AD had greater vocabulary richness when writing blog posts compared to when speaking (unpublished data). We decided to use this project to gain a more thorough understanding of the exact nature of vocabulary richness in spoken language.

One of the biggest challenges in machine learning is the number of features. Large amounts of unnecessary features cause overfitting, where a model may perform well during testing but poorly in the real world and increase the computation time. Feature selection is where developers remove features that are redundant or have low predictive power. To build more effective diagnostic models, researchers need to know what language features are affected and indicative of AD. Measuring vocabulary richness in AD determines if TTR is a good predictor of AD and if it should be used as a feature when training AD diagnosis tools (30).

Building on the importance of using effective features in machine learning based diagnosis, the objective of this study was to characterize the decline in vocabulary richness in individuals with AD. We hypothesized that vocabulary richness would be significantly higher in patients without compared to participants with AD, with the expectation that the effect size would be medium to large. Given that TTR is influenced by word repetition, we predicted that the effect size of mean TTR would be large between the two groups. We included effect size in our hypothesis and results to compare the effect size between different lexical-semantic features.

We found that vocabulary richness was significantly reduced in AD participants compared to healthy adults. The effect size of TTR between the two groups was smaller than other language features measured in other studies due to the use of memory aids by participants. Our results demonstrate that TTR should be used as a training feature in machine learning models. They support the use of non-invasive language-based diagnosis of AD and may improve its accuracy and therefore accessibility.

To investigate vocabulary richness in AD, we analyzed the Pitt corpus (DementiaBank) , an English language database containing audio and text data of participants with and without AD (31). Participants sat down with an interviewer and were asked to describe what they saw in the "Cookie Theft" illustration from the Boston Diagnostic Aphasia Examination

https://doi.org/10.59720/24-289

(32). The illustration depicts a mother drying dishes by a sink in the kitchen while a boy stands on a chair to take cookies from a cupboard accompanied by a girl. Participants were instructed to describe events occurring in the picture. This task is commonly used with adults with AD, aphasia, right-hemisphere stroke, and (Mild Cognitive Impairment) MCI because of its ability to reveal linguistic impairments (33). Speech from the interview was later hand-transcribed into interview transcripts. Transcripts and the associated demographic data of the participant were loaded into a comma-separated file (CSV). Unique words were measured as an indicator of vocabulary richness.

RESULTS

Sixty-eight participants were in the AD group and 68 were in the control group. Both groups were matched on age and education to avoid bias. The AD group was slightly older than the control group; the mean age for the AD group was 69.3 ± 7.8 years, and the mean age for the control group was 66.9 ± 7.3 years. Education in both groups was point-to-point matched. The male-to-female ratio was 29:39 and 24:44 in the AD and control groups, respectively (**Table 1**). The correlation between the Mini-Mental State Examination (MMSE) score and TTR was -0.11 ($p=0.41$) in the AD group and 0.03 ($p=0.86$) in the control group.

Both groups were confirmed to have a normal distribution of TTR (Shapiro-Wilk test, $W=0.99$ for AD and 0.96 for HC, $p>0.05$) **(Figure 1).** The closer to 1 the W value is, the more normally distributed the group is, so TTR in the AD group was slightly more normally distributed than the control group (34). The closeness of both values to 1 indicated that our use of parametric tests, such as Student's t-test and Cohen's d, to assess the relationships in our data was appropriate.

When assessing the vocabulary richness used to describe the "Cookie Theft" illustration from the Boston Diagnostic Aphasia Examination, we found that AD participants repeated words or phrases more often than healthy participants (**Table 2**). The control group had a significantly higher mean TTR than the AD group (mean TTR=0.45 vs. mean TTR=0.40; $p=0.007$) **(Figure 2)**. This indicates that 45% of speech in the control group was made up of unique words, while only 40% of speech in the AD group was made up of unique words. The absolute value of Cohen's d was 0.46, which falls between the small (absolute $d=0.02$) and medium (absolute $d = 0.05$) effect size threshold. Therefore, the effect size of the mean TTR is small to medium (12).

	Total	AD Group	Control Group
# of individuals	136	68	68
Sex (Male: Female)	53:83	29:39	24:44

	Mean	AD Group	Control Group
Mean Age (years)	68	69	67 p=0.64
Mean Years of Education	13.2	13.2	13.2 p=1.0

Table 1. Demographic data of participants. Number of participants in each demographic group. Participants were age and education matched. Student's t-test: $p>0.05$ for mean age and education level in years.

DISCUSSION

We set out to explore the utility of linguistic characteristics to distinguish between participants with or without AD. Our findings demonstrated that individuals with AD exhibited reduced vocabulary richness compared to participants without AD. Individuals without AD had a higher mean TTR than individuals with AD, indicating that healthy adults had increased vocabulary richness compared to adults with AD ($p<0.05$). The relative effect size of TTR between the two groups was medium.

The observed decrease in vocabulary richness of individuals with AD appeared to be primarily due to the verbal perseverations in the AD group. Verbal perseverations, which are the inappropriate repetitions of words and phrases, were evident in the individual samples of individuals with AD (35). Easily retrievable words are more likely to be repeated. A study by Mizzo et al. found that, of 392 participants, 74% of individuals with AD exhibited preservation rates above the average rate of healthy controls (1–3% for healthy controls) (35). It is possible that participants were able to retrieve easy-to-generate words in the image, such as "cookie", "boy", and "water" and were unable to retrieve less obvious words, such as "teacup" and "grass". It may also have been more difficult for participants with AD to keep track of already mentioned words, leading to repetition.

We found a medium relative difference in TTR effect size between control and AD participants, contrary to the initial hypothesis of a large effect size. We speculate that this difference may have been due to participants not engaging memory when completing the picture description task. Difficulty recalling words from memory is a hallmark of AD, with AD patients making more errors when recalling words than controls (36). In a study with 168 participants, 32% of participants with AD, but only 16% of healthy participants, made intrusion errors on a word list memory task (37). On the contrary, visual aids like images and illustrations have been shown to enhance verbal abilities in AD patients (38). Participants of the Pitt corpus completed the picture description task with the illustration in front of them as opposed to having to recall the image from memory, in some cases, the interviewer referenced the image itself. For example, one investigator said, "Can you tell me what this is?" Participants therefore did not need to engage in long- or short-term memory while completing the task. The fact that we did not see a larger difference in TTR with the picture description task could indicate that the presence of visual cues can minimize the manifestation of language impairments in individuals with AD.

The effect size of TTR ($d=-0.46$) was also smaller than metrics like word and topic repetition found in Yamada et al. Yamada et al. had participants use spontaneous speech with no memory aids, which may explain why the effect size of word and topic repetition was larger than the effect size we observed in our study (29). Despite the absence of memory recall pressure, we still detected notable differences in vocabulary richness, highlighting the impact of AD on language and indicating that TTR is capable of detecting some of the effects of AD on language capabilities. Future studies using tasks that require memory recall might reveal lower TTR in AD participants and a correspondingly larger

JOURNAL OF
EMERGING INVESTIGATORS
https://doi.org/10.59720/24-289

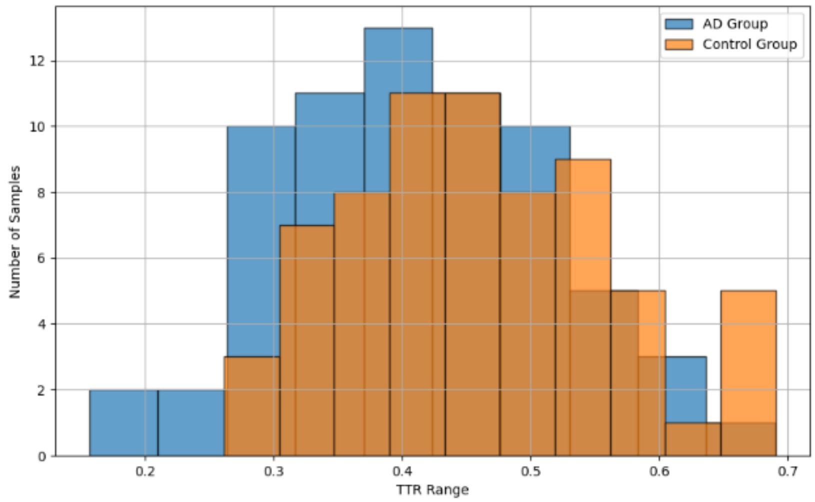

Figure 1. Type-token ratio (TTR) distribution in the Alzheimer's disease (AD) and control group. Number of transcripts in each TTR bin. TTR was measured with the Natural Language Processing Toolkit. Shapiro-Wilk test confirms that both groups have an approximately normal distribution (p>0.05). The orange bars indicate the TTR distribution of the control group, and the blue bars indicate the TTR distribution of the AD group.

effect size of TTR, thereby improving the effectiveness of TTR as a language feature for diagnosing AD. Our findings suggest that differences in vocabulary richness are present but influenced by task design and other contextual factors.

Our results support current literature on TTR decline in AD. Kave and Goral found that AD participants had a lower TTR of 0.73 versus 0.87 in control participants when describing the Boston "Cookie Theft" task (39). The difference in TTR between our findings and Kave and Goral's may have been due to interviewers asking for elaboration from participants. Interviewers in Kave and Goral's study prompted participants to elaborate on their statements by asking questions, while interviewers in the Pitt corpus responded to participants with "ok" and "hm" (39). Participants in Kave and Goral's study may have produced more complex sentences, making the language deficits of AD participants more apparent. Our findings support the consensus that language in AD individuals become limited in vocabulary (39). In contrast to previous literature, we found no significant correlation between TTR and MMSE score in the AD group (39). Lower vocabulary richness was not linked with greater cognitive impairment. We think this may be due to having a large proportion of data missing (35 participants, or over a quarter of the cohort, had unavailable MMSE data).

Our primary limitation was the use of Pitt corpus, which lacked important demographic data, such as socioeconomic status, language history, primary language, and information on visual and auditory impairments. Unaccounted demographic data could have impacted TTR in participants. For example, visual and auditory impairments may have made it harder for the participant to understand the task at hand, causing them to become confused and indicate lower TTR even in healthy individuals; individuals whose first language is not English may produce less diverse vocabulary, regardless of group. These variables may also influence the current findings and need to be addressed in future research. Furthermore, including additional demographic data would lead to further insight into language impairment, such as the relationship between vocabulary richness and socioeconomic status. An

Participants with AD
"kids are trying to get a. its full of its full of mistakes . its full of mistakes . its full of mistakes . he is changing taking cookie jar. thats all . the mothers just drying the dishes. from the from . this is …"
"this boy is getting cookies out of this jar . .. well why why they must they they must . no you've gotta you gotta get the you gotta get . yeah I guess so ."
Participant without AD
"there's a child reaching for a cookie . the stool is tilted . the girl is possibly moving her finger towards her mouth and waters running out of the sink . the mothers drying dishes . somebody's gonna be screaming when the kid falls . I don't know if its action but mom was standing in water . that's about all I see ."

Table 2. Speech excerpts from participants with and without AD. Speech from transcripts from the AD and control group.

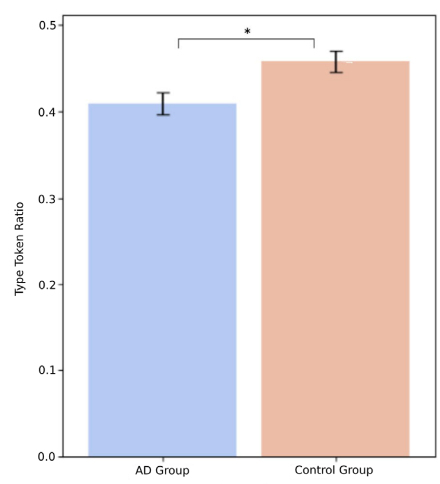

Figure 2. Mean TTR in participants with and without AD. Mean TTR distribution in the control and AD groups. *$p < 0.05$. The AD group had a significantly lower mean TTR than the control group, indicating decreased vocabulary richness.

https://doi.org/10.59720/24-289

ideal dataset would also include participants completing a recall task, to limit the effect of visual aids on word retrieval. Recall tasks would also be more true to real-life interactions, as individuals with AD do not always have visual aids on them, and may therefore be a better indicator of AD than a language test using visual aids.

In the future, we would like to conduct a more thorough review of vocabulary in AD. We would like to investigate word class differences (e.g., noun use vs. verb use) by calculating TTR sorted by parts of speech or using parts of speech specific metrics like noun index. We could also confirm our explanation of verbal perseverations by calculating the number of repeated words divided by the number of total words, or the number of consecutive repeated words. Five minutes of speech data is considered more effective than one-minute speech samples, so we would seek out a long corpus in the future (29).

Language-based AD diagnosis models can make AD diagnosis more accessible and affordable. One potential application of these models is in mobile apps. Mobile apps can diagnose AD with a 78.6% accuracy and the early stages (MCI) with 87.6% accuracy and can be self-administered (40). However, these models can be plagued by overfitting (reducing applicability in the real world) and high computational cost (models are limited in what hardware they can run on) if they are trained on poor features. Feature selection identifies language features that make the model more accurate (41). In feature selection, features are generally used if they are statistically significant between the two groups, so TTR is still an effective feature despite not having a large effect size (41). The effect size of TTR may increase by removing memory aids during tasks, further supporting the use of TTR in machine learning models. TTR, along with other features (such as the adjective to unique word ratio and the adverb to unique word ratio), makes AD diagnosis from language more applicable (42).

In summary, we found that individuals with AD had decreased vocabulary richness compared to healthy adults of the same age, as measured by TTR. Qualitative analysis indicated that individuals with AD use repetitions, like verbal perseverations and echolalia, more often than individuals without AD. There is a statistically significant change in vocabulary richness in affected individuals, which points to TTR as a training feature for AD diagnosis models. Our findings further support the evaluation of language as an effective non-invasive mode to diagnose AD. Our findings can also inform caregivers of language impairments they can expect to see in their patients with AD, which can aid them in the treatment of, and their day-to-day interactions with, the patient.

MATERIALS AND METHODS

Preprocessing

The Pitt corpus (DementiaBank) was first downloaded as a .zip file (31). The dataset was scrapped and converted to a .csv file along with the file name. Demographic data was matched up to speech data by matching the file names. Linguistic corpus was pre-processed using Regex and Natural Language Processing Toolkit (NLPK) (Python version 3.10.9). DementiaBank transcripts contain dialog made by the investigator and miscellaneous symbols. Any dialog preceded by "INV" was removed with Regex, as that dialog was from the investigator and not the participants. Any punctuation excluding periods, commas, semicolons, questions, and exclamation marks were also removed. Non-words such as "[*p:n]" and "&+w", and filler words such as "erm", "uh" and "ah were also removed to avoid interference. DementiaBank stores mispronounced words where the intended paraphrased word is written in brackets after the mispronounced word. For example: "*He is at the think [:sink]*," where "sink" was the intended word and "think" was the pronounced word. During preprocessing, the mispronounced word was removed as verbal apraxia was not in the scope of this study. The preprocessed phrase would then read "*He is at the sink,*" and this edited phrase would then be used for downstream analyzes.

NLPK was used to calculate total word count, and then the number of words that appear once. The Researchpy library (version 0.3.5) was used to calculate statistical tests. Matplotlib (version 3.6.2) was used to display graphs and charts.

Participants were classified into two groups, the AD group and the healthy control (HC) group. We were not able to control for the severity of AD of participants in the AD group because the corpus did not provide data on disease severity for individuals with an AD diagnosis. The AD group consisted of participants with a physician-confirmed diagnosis of AD, while HC group had no relevant cognitive impairment diagnoses. Of the initial 142 participants, 6 were excluded because their diagnosis was a condition other than AD or healthy control.

Statistical analysis

The Shapiro-Wilk test is a test of normal distribution in data; the closer to 1 the test is the more normally distributed the data is, $p<0.05$ indicates that the data has a skewed distribution. The two-tailed paired t-test was used to test statistical significance. The t-test is considered statistically significant when the p-value is below the alpha value ($a=0.05$). Cohen's d is the measure of the practical significance of differences of means. The test statistic is the effect size. The relative size of Cohen's d was determined by using recommended ranges (31).

Mini mental state examination (MMSE)

MMSE is a questionnaire that evaluates attention, recall, memory, calculation, drawing polygons, and language and measures cognitive impairment overall. Participants were given a score of up to 30, with a lower score indicating greater impairment (43).

Controlling for external factors

TTR tends to be biased towards shorter text samples and decreases with increasing length of text. Therefore, each text was restricted to 150 words based on previous research suggesting that this length is adequate for tracking language changes in AD (44). Using Python code, TTR was calculated by dividing the number of unique words by the number of total words per participant transcript. Outliers with a z-score above 3 were removed

Python Code

The code first split the transcript text from sentences into words to determine total word count. For every word in the transcript, the code tried to find a matching string; if a matching word cannot be found, then the word is determined to be unique. The total word count and unique word count were recorded in the "Total Word Count" and "Unique Word Count" columns, respectively. TTR was calculated by dividing the Unique Word Count column by the Total Word Count column and recorded in the "TTR" column. The code limits transcripts to 150 words by replacing words with an index number greater than 150 with blank space. Our code can be found in the following Github repository: https://github.com/texas-flavoured-canadien/Vocab_Richness_AD/blob/main/Vocabulary_Richness_AD.ipynb.

ACKNOWLEDGMENTS

We would like to thank Dr. Vaden Masrani, Joshua Chow, Mr. Balkwil, Ms. Davis, Dr. Davida Fromm and Dr. Brian MacWhinney for their support.

Received: September 13, 2024
Accepted: June 23, 2025
Published: December 4, 2025

REFERENCES

1. "Alzheimer's Disease Facts and Figures." *Alzheimer's Association.* www.alz.org/alzheimers-dementia/facts-figures. Accessed 1 Mar. 2024.
2. Klimova, Blanka, et al. "Alzheimer's disease and language impairments: social intervention and medical treatment." *Clinical Interventions in Ageing,* vol. 10, Aug. 2015, https://doi.org/10.2147/CIA.S89714
3. "What Are the Signs of Alzheimer's Disease?" *National Institute on Ageing.* www.nia.nih.gov/health/what-are-signs-alzheimers-disease. Accessed 1 March 2024
4. "Dementia numbers in Canada." *Alzheimer Society of Canada.* https://alzheimer.ca/en/about-dementia/what-dementia/dementia-numbers-canada. Accessed 1 Mar. 2024.
5. Breijyeh, Zeinab, and Rafik Karaman. "Comprehensive Review on Alzheimer's Disease: Causes and Treatment". *Molecules*, vol. 25, no. 24, Dec. 2020, p. 5789. https://doi.org/10.3390/molecules25245789.
6. Todd, S., et al. "Cause of Death in Alzheimer's Disease: A Cohort Study". *QJM: An International Journal of Medicine*, vol. 106, no. 8, Aug. 2013, pp. 747–53. https://doi.org/10.1093/qjmed/hct103.
7. Conroy, Paul J., et al. "Using Phonemic Cueing of Spontaneous Naming to Predict Item Responsiveness to Therapy for Anomia in Aphasia". *Archives of Physical Medicine and Rehabilitation*, vol. 93, no. 1, Jan. 2012, pp. S53–60. https://doi.org/10.1016/j.apmr.2011.07.205.
8. "Echolalia." *Cleveland Clinic*. https://my.clevelandclinic.org/health/symptoms/echolalia. Accessed 10 June 2025.
9. "Verbal Apraxia." Stony Brook Medicine. https://www.stonybrookmedicine.edu/sites/default/files/pe_verbal_apraxia.pdf. Accessed 10 June 2025.
10. Masrani, Vaden. "Detecting dementia from written and spoken language." *UBC Theses and Dissertations.* Jan. 2018 https://doi.org/10.14288/1.0362923
11. Aramaki, Eiji, et al. "Vocabulary Size in Speech May Be an Early Indicator of Cognitive Impairment." *PLOS ONE,* vol. 11, no. 5, pp. e0155195, 13 May 2016. https://doi.org/10.1371/journal.pone.0155195
12. Albertoni, Riccardo, et al. "Semantic Granularity for the Semantic Web" *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops,* 2006. pp. 1863–1872. http://doi.org/10.1007/11915072_109.
13. Sanz, Camila, et al. "Automated text-level semantic markers of Alzheimer's disease." *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 14, no. 1, 14 Jan 2022. https://doi.org/10.1002/dad2.12276
14. Dumas, Annette, et al. "Rethinking the detection and diagnosis of Alzheimer's disease: Outcomes of a European Brain Council project." *Aging Brain*, vol. 4, May 2023, https://doi.org/10.1016/j.nbas.2023.100093
15. Livingston, Gill, et al. "Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission." *The Lancet Commissions,* vol. 404, no. 10452, 10 Aug. 2024, pp. 572-628, https://doi.org/10.1016/S0140-6736(24)01296-0
16. Yiannopoulou, Konstantina G., et al. "Reasons for Failed Trials of Disease-Modifying Treatments for Alzheimer Disease and Their Contribution in Recent Research." *Biomedicines*, vol. 7, no. 4, Dec. 2019, p. 97, https://doi.org/10.3390/biomedicines7040097.
17. Dubois, Bruno, et al. "Timely Diagnosis for Alzheimer's Disease: A Literature Review on Benefits and Challenges" *Journal of Alzheimer's Disease,* Vol 49, No. 3, 2016, pp. 617–631. https://doi.org/10.3233/JAD-150692
18. Juganavar, Anup, et al. "Navigating Early Alzheimer's Diagnosis: A Comprehensive Review of Diagnostic Innovations". *Cureus*, vol. 15, no. 9, pp. e44937. https://doi.org/10.7759/cureus.44937.
19. "Dementia in Canada: Economic Burden 2020 to 2050." *Canadian Centre for Economic Analysis.* www.cancea.ca/wp-content/uploads/2023/07/CANCEA-Economic-Impact-of-Dementia-in-Canada-2023-01-08.pdf. Accessed 1 Mar 2024.
20. Ostrand, Rachel and John Gunstad. "Using Automatic Assessment of Speech Production to Predict Current and Future Cognitive Function in Older Adults" *Journal of Geriatric Psychiatry and Neurology*, vol. 34, no. 5, Sep. 2021. pp. 357–369. https://doi.org/10.1177/0891988720933358
21. Gaeta, Anna Michela, et al. "Predicting Alzheimer's Disease CSF Core Biomarkers: A Multimodal Machine Learning Approach". *Frontiers in Aging Neuroscience*, vol. 16, June 2024. https://doi.org/10.3389/fnagi.2024.1369545.
22. Toumaj, Shiva, et al. "Applications of Deep Learning in Alzheimer's Disease: A Systematic Literature Review of Current Trends, Methodologies, Challenges, Innovations, and Future Directions". *Artificial Intelligence Review*, vol. 58, no. 2, Dec. 2024, p. 44. https://doi.org/10.1007/s10462-024-11041-5.
23. Yao, Zhaomin, et al. "Artificial Intelligence-Based Diagnosis of Alzheimer's Disease with Brain MRI Images". *European Journal of Radiology*, vol. 165, Aug. 2023, p. 110934. https://doi.org/10.1016/j.ejrad.2023.110934.
24. Yang, Qin, et al. "Deep Learning-Based Speech Analysis for Alzheimer's Disease Detection: A Literature Review". *Alzheimer's Research & Therapy*, vol. 14, no. 1, Dec. 2022, p. 186. https://doi.org/10.1186/s13195-022-01131-3.
25. Gordon, Jean K. "Factor Analysis of Spontaneous Speech in Aphasia". *Journal of Speech, Language, and Hearing*

https://doi.org/10.59720/24-289

Research, vol. 63, no. 12, Dec. 2020, pp. 4127–47. https://doi.org/10.1044/2020_JSLHR-20-00340.
26. García-Gutiérrez, Fernando, et al. "Unveiling the sound of the cognitive status: Machine Learning-based speech analysis in the Alzheimer's disease spectrum." *Alzheimer's Research & Therapy,* vol. 16, no. 26, 2 Feb. 2024. https://doi.org/10.1186/s13195-024-01394-y.
27. Guinn, Curry and Anthony Habash. "Language Analysis of Speakers with Dementia of the Alzheimer's Type" *Artificial Intelligence for Gerontechnology*. 2021 https://cdn.aaai.org/ocs/5574/5574-23756-1-PB.pdf.
28. Zozuk, Nataliia. "Lexical diversity and language impairment." *Journal of Linguistics,* Vol. 74, No 1, 2023, pp. 301 – 309. https://doi.org/10.2478/jazcas-2023-0047.
29. Yamada, Yasunori, et al. "Atypical Repetition in Daily Conversation on Different Days for Detecting Alzheimer Disease: Evaluation of Phone-Call Data From Regular Monitoring Service". *JMIR Mental Health*, vol. 7, no. 1, Jan. 2020, p. e16790. https://doi.org/10.2196/16790.
30. Pudjihartono, Nicholas, et al. "A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction". *Frontiers in Bioinformatics*, vol. 2, June 2022, p. 927312. https://doi.org/10.3389/fbinf.2022.927312.
31. Becker, J. T., et al. "The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis." *Archives of Neurology*, vol. 51, no. 6, 1994, pp. 585-594. https://doi.org/10.1001/archneur.1994.00540180063015
32. Goodglass, Harold, et al. "BDAE The Boston Diagnostic Aphasia Examination.", edited by, 3rd ed., Lippincott Williams & Wilkins, 2001.
33. Cummings, Louise. "Describing the Cookie Theft Picture: Sources of Breakdown in Alzheimer's Dementia". *Pragmatics and Society*, vol. 10, no. 2, July 2019, pp. 153–76. https://doi.org/10.1075/ps.17011.cum.
34. "Normality Checks: The Shapiro-Wilk Test – A Comprehensive Overview." *Kaggle*. https://www.kaggle.com/general/432129. Accessed 10 June 2025.
35. Miozzo, Michele, et al. "Perseverations in Alzheimer's disease: Memory slips?" *Cortext,* 2012. http://doi.org/10.1016/j.cortex.2012.10.016
36. Baek, Min et al. "Comparison between the Story Recall Test and the Word-List Learning Test in Korean patients with mild cognitive impairment and early stage of Alzheimer's disease" *Journal of Clinical and Experimental Neuropsychology*. 2012, http://doi.org/10.1080/13803395.2011.645020
37. Chaves, Marcia L. F. and Ana L. Camozzato. "How many items from a word list can Alzheimer's disease patients and normal controls recall? Do they recall in a similar way?" *Dementia & Neuropsychologia,* vol. 1, no. 1, 2007, pp. 52–58. https://doi.org/10.1590/S1980-57642008DN10100009
38. Egan, Mary, et al. "Methods to Enhance Verbal Communication between Individuals with Alzheimer's Disease and Their Formal and Informal Caregivers: A Systematic Review" *International Journal of Alzheimer's Disease,* 10 Jun. 2010. https://doi.org/10.4061/2010/906818
39. Gitit Kavé & Ayelet Dassa." Severity of Alzheimer's disease and language features in picture descriptions" *Aphasiology*, Vol. 32, No. 1, 2018, pp. 27-40. https://doi.org/10.1080/02687038.2017.1303441
40. Yamada, Yasunori, et al. "A Mobile Application Using Automatic Speech Analysis for Classifying Alzheimer's Disease and Mild Cognitive Impairment". *Computer Speech & Language*, vol. 81, Jun. 2023, p. 101514. *ScienceDirect*, https://doi.org/10.1016/j.csl.2023.101514.
41. Banerjee, Chandradip. "P Value and Feature Selection." *Medium*. https://medium.com/@chandradip93/p-value-and-feature-selection-629bec71d828. Accessed 10 June 2025.
42. Zozuk, Nataliia Casnochova, et al. "Relationship between Language Features Extracted through NLP and Clinically Diagnosed Alzheimer's Disease and Mild Cognitive Impairment in Slovak". *Alzheimer's & Dementia : Diagnosis, Assessment & Disease Monitoring*, vol. 17, no. 2, 19 May 2025, p. e70122. *PubMed Central*, https://doi.org/10.1002/dad2.70122.
43. Arevalo-Rodriguez, Ingrid, et al. "Mini-Mental State Examination (MMSE) for the Detection of Alzheimer's Disease and Other Dementias in People with Mild Cognitive Impairment (MCI)". *The Cochrane Database of Systematic Reviews*, vol. 2015, no. 3, 27 July 2021, p. CD010783. https://doi.org/10.1002/14651858.CD010783.pub3.
44. Petti, Ulla, et al. "How Much Speech Data Is Needed for Tracking Language Change in Alzheimer's Disease? A Comparison of Random Length, 5-Min, and 1-Min Spontaneous Speech Samples" *Digital Biomarkers,* vol. 7, no. 1, 24 Nov. 2023, pp. 157–166. https://doi.org/10.1159/000533423