9-2021

# Evaluating the Role of Gender in Dementia-Related Language Deficiencies

Kelsey Bourque

*The Graduate Center, City University of New York*

# EVALUATING THE ROLE OF GENDER IN DEMENTIA-RELATED LANGUAGE DEFICIENCIES

by

KELSEY BOURQUE

A thesis submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the degree of Master of Arts, The City University of New York

2021

Evaluating the Role Of Gender in Dementia-Related Language Deficiencies

by

Kelsey Bourque

This manuscript has been read and accepted for the Graduate Faculty in Linguistics in satisfaction of the thesis requirement for the degree of Master of Arts.

| | |
|---|---|
| Date | [Rivka Levitan] |
| | Thesis Advisor |
| Date | [Cecelia Cutler] |
| | Executive Officer |

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Evaluating the Role Of Gender in Dementia-Related Language Deficiencies

by

Kelsey Bourque

Advisor: Rivka Levitan

Typically, about 60% of dementia patients are women. Researchers have historically dismissed this imbalance as a result of the life expectancy for women being longer, and since age is the primary risk factor associated with dementia, and women's longer lifespan equates to a higher percentage of the dementia patient population (Mielke, 2018). While the exact cause of dementia is unknown, researchers and clinicians have historically treated male and female populations the same, asserting that there is no significant difference between the two sexes in regards to detecting dementia. The present study aims to address this potential gap in dementia research, where newer research (as recent as 2018) also demands for differences in gender to be addressed in this field. In the present study the Pitt Corpus from DementiaBank, to attempt to find significant results in how men and women with dementia utilize language. A statistical analysis was performed using linear regression and ANOVA models, which found significant interactions between sex and linguistic features. This same data was used to train and test machine learning models in attempts to categorize utterances from both sexes accurately. Logistic regression, Naive Bayes, and SVM models were used on various forms of TF-IDF vectors, with logistic regression performing at the highest accuracy at 56%. The implication of these results aligns with the hypothesis of this study, that there is a significant difference between the linguistic markers of both sexes.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

# CHAPTER 1: INTRODUCTION TO DEMENTIA

According to the Alzheimer's Association, "Dementia is a general term for loss of memory, language, problem-solving and other thinking abilities that are severe enough to interfere with daily life." Dementia typically affects people over the age of 65: one in 14 people in this age group have dementia (Alzheimer's Society). However, dementia can also affect younger people. A common stereotype regarding dementia is that this disease is a normal part of aging, due to how common dementia is. However, this stereotype is misguided, as serious mental decline is not a normal or healthy part of aging.

Typically, about 60% of dementia patients are women. Researchers have historically dismissed this imbalance as a result of the life expectancy for women being longer, and since age is the primary risk factor associated with dementia, women's longer lifespan equates to a higher percentage of the dementia patient population (Mielke, 2018). A dementia diagnosis relies on a few different sources, with no single test to determine whether or not a patient has dementia. In order to diagnose patients, medical providers often take into account medical history, and perform a physical examination, laboratory tests, and assess changes in thinking, day-to-day function, and behavior associated with each type of dementia (Alzheimer's Association). While the exact cause of dementia is unknown, researchers and clinicians have historically treated male and female populations the same, asserting that there is no significant difference between the two sexes in regards to detecting dementia.

The present study aims to address this potential gap in dementia research, where newer research (as recent as 2018) also demands for differences in gender to be addressed in this field. In the present study, a well-known dataset collected from real dementia patients was utilized, the Pitt Corpus from DementiaBank, to attempt to find significant results in how men and women with dementia utilize language. A statistical analysis was performed using linear regression and ANOVA models, which measured interactions between sex and linguistic features. This same data was used to train and test machine learning models in attempts to categorize utterances from both sexes accurately. Logistic regression, Naive Bayes, and SVM models were used on various forms of TF-IDF vectors, with logistic regression performing at the highest accuracy at 60%. The implication of these results aligns with the hypothesis of this study, that there is a significant difference between the linguistic markers of both sexes.

## 1.1 Dementia and Language

Cognitive decline due to various forms of dementia is a well-studied area in neuroscience and gerontology, as well as in computational linguistics. Language deficits are a common symptom of dementia, through which patients often have difficulty with word finding (anomia), sentence comprehension, and experience a lack of cohesion in conversation (Kempler et al, 2008).

There are different types of dementia that have each been heavily researched in the medical field, including dementia of the Alzheimer's type (also called Alzheimer's dementia, or DAT), and vascular dementia (VD). There are also two types of frontotemporal dementia: semantic dementia (SD) and primary progressive nonfluent aphasia (PNFA). Despite all four of these subtypes falling under the umbrella of dementia diagnoses, they exhibit  differences in their clinical presentation.

Alzheimer's disease, a degenerative brain disease causing dementia, has a notable progressive deterioration of memory, accompanied by at least two other cognitive deficits, such as language, visuospatial perception, or executive function. DAT affects many areas of the brain, but particularly the hippocampus and parts of the frontal cortex (Kempler et al, 2008).

Vascular dementia is caused by cholesterol clogging the blood vessels in the brain, resulting in insufficient oxygen delivery. Small blockages can develop, causing small strokes that kill brain cells, most of which often go unnoticed. VD can also develop after a major stroke, which causes sudden mental change, sometimes accompanied by paralysis or slurred speech. Patients with VD often exhibit symptoms like confusion, slurred speech, and problems thinking and remembering (Harvard Health).

Frontotemporal dementia affects the frontal and temporal lobes of the brain, two areas that are associated with personality, behavior, and language. Semantic dementia (SD) primarily affects the temporal lobe, and primary progressive nonfluent aphasia (PNFA) primarily affects the frontal lobe. SD patients display language impairments that are considered distinct from other types of dementia. SD is characterized by fluent speech output with anomia and comprehension issues. Despite SD usually standing out from other forms of dementia, it can become clinically confused with DAT, specifically in patients with DAT who show early language impairments. The second type of frontotemporal dementia is primary progressive nonfluent aphasia (PNFA). PNFA typically exhibits nonfluent speech output and anomia, alongside decent comprehension. Unlike SD, PNFA is rarely confused with DAT due to the nonfluent speech output (Kempler et al). Although these forms of dementia are described as distinct, patients can exhibit a mixture of symptoms across different types of dementia, referred to as mixed dementia (MD).

### 1.2 Clinical Language & Cognition Tasks

Various clinical tests are required for a confident dementia diagnosis, including language tests. These tasks measure basic language and cognition skills, such as anomia and recall. One well-known test, the Cookie Theft Picture, measures a patient's language skills is the description of events occurring in a drawing. The Cookie Theft Picture features a woman washing dishes with the water overflowing in the sink, spilling onto the floor. Behind her, her children are trying to steal cookies from the cookie jar on the top shelf of a cabinet, with her son standing on an unsteady stool while her daughter looks on. This drawing is part of the Boston Diagnostic Aphasia Examination, and is meant to measure cognitive and linguistic impairments and regression in a wide range of patient populations, not just dementia (Berube et al, 2019).

While the Cookie Theft Picture is widely used in dementia diagnosis, it is not the only language task used to help diagnose and measure the progression of dementia. The Pyramids and Palm Trees (PPT) test measures semantic memory, representation, and processing. PPT consists of 52 picture sets, each one containing three words or pictures, and also three samples in order to introduce the test to subjects. The target word or picture is placed on the top of the other two words or pictures, and the subject is asked to match the target word or picture with one of the words or pictures below (Mehri et al, 2018). For example, the target picture may be a fork, and the participants are given a ladle or a spoon to choose from. Since a spoon is more closely related to a fork in that it is used for eating, while a ladle is used more for serving, participants should choose the spoon (Mehri et al, 2018). Other popular tests for language capacity in patients with dementia include the Cinderella test. The Cinderella test (TalkBank) asks patients to tell the popular fairytale "Cinderella", with the goal of producing a complex speech act. This task can be accompanied by a Cinderella picture book, and the task is usually introduced by asking the participant if they remember the story of Cinderella. The goal of the Cinderella test is to assess memory deterioration due to dementia and/or other aphasias (MacWhinney et al, 2010). One aspect that all three tests (and others) have in common is that they rely heavily on cultural context. In fact, the PPT test was adapted by Mehri et al to test their dementia patients, Persian adults, on material that they would likely be familiar with. It was translated into Persian, and some picture changes were made, including, "canoe to boat, tulip to wheat, nun to clergyman, church to mosque, bellows to hand fan, and ticket to Persian ticket." (Mehri et al, 2018). Similarly, the Cinderella story is a classic that most people in the Western world are familiar with, therefore making it an appealing task for researchers to test their patients with.

### 1.3 Background Research

DementiaBank is a widely utilized and collaborative archive of various audio, text, and video files of clinical data collected from actual dementia patients. DementiaBank is part of a larger archive, TalkBank, which serves as a database for various subjects in language studies, such as aphasia, child language development, and more. Many researchers have used the datasets from DementiaBank in machine learning experiments, due to the ease of accessibility of the corpora.

Using DementiaBank, Orimaye et al (2017) examined Alzheimer's patients specifically to find low-level linguistic features that they could use to classify language data from patients with Alzheimer's with machine learning algorithms, in order to compare them to data from healthy patients. Using annotated transcripts, the authors extracted nine syntactic features via the Stanford Parser, plus 14 lexical features, as well as bigrams and trigrams. The DementiaBank transcripts include CHAT symbols, which represent implicit and explicit features that describe each patient's lexical capabilities. Orimaye et al (2017) extracted each CHAT symbol in the transcript files and stored them according to their frequencies and positions in each sentence. The CHAT symbols mark errors and other behaviors the participants produce, such as word repetitions or verbal errors in which a correction attempt was made. This aids in the extraction of features directly from the transcripts. The notable lexical features included repetitions and revisions, plus mean length of utterances (MLU). MLU is often used to measure grammar growth in children with Specific Language Impairment, but Orimaye et al (2017) extrapolated this feature to dementia patients to determine language disorder. Other lexical features that align well with noted language deficits in dementia include trailing off indicator (CHAT symbol), word repetition, incomplete words, and filler words. First, a statistical analysis was performed on the feature set, both a Student's t-test and a Mann-Whitney test. They achieved similar two-tailed results, but they chose the Student's t-test for further statistical evaluation. The final step in this experiment was the application of machine learning algorithms to the data. This study specifically utilized Sequential Multiple Optimisation (SMO), which is a variant of the Support Vector Machines (SVM) algorithm. The diagnostic task proved to perform well with these low-level features (reduced sentence, MLU, and trailing off), with different combinations scoring mostly between 70-93%, with one outlier scoring 54% accuracy, demonstrating that machine learning algorithms can be used to accurately predict the diagnosis of probable Alzheimer's disease.

For their study on detecting linguistic characteristics of Alzheimer's disease, Karlekar et al (2018) used the same data from DementiaBank as Orimaye et al (2017). However, they relied on neural networks to determine the best features for them to distinguish the difference between dementia and control participants. Karlekar et al (2018) argue that the automatic diagnosis of dementia is extremely complicated due many variables, for both the patients and researchers. Researchers need to be highly specialized in the language tasks performed (recognizing self-corrections, incorrect words, word searching, etc.) to recognize dementia-like patterns, and patients present different linguistic biomarkers at different stages of progression in their disease. Considering these points, Karlekar et al (2018) designed an end-to-end neural model to address these shortcomings, including CNN and LSTM-RNN models, and the third model being a joint CNN-LSTM architecture. The models were tested with both POS-tagged and untagged data. The untagged data yielded lower results, each model scoring 82.8%, 83.7% and 84.9% accuracy, respectively. The tagged data produced better results, with the CNN-LSTM model producing 91.1% accuracy. An error analysis revealed that almost all Alzheimer's disease-positive data were correctly classified, and a further false negative analysis on 10% of the misclassified data showed that 36.3% were short utterances, such as "okay", "alright", etc. These forms of speech utterances were found in both participant groups, but were more common in the Alzheimer's disease group. The remaining 63.7% false negatives were utterances that were deemed ambiguous due to surrounding context. The authors also considered patient gender, which has been a topic of debate amongst researchers, as some believe that language differences between men and women account (or do not account) for notable differences between male and female dementia patients. However, Karlekar et al (2018) examined the top ten most frequent words for both male and female patients, and did not find  a significant difference between the two lists. In other words, Karlekar et al (2018) did not detect a meaningful difference between male and female patients in regards to the production of frequent words. The authors also separated the training data by gender and tested the models on each set. The male-tagged utterances achieved an 86.6% accuracy, and the female-tagged utterances 86.2% accuracy, suggesting again that the gender of the patient did not make a significant impact on their models.

The aforementioned studies demonstrate differing approaches to creating predictive models through the use of DementiaBank. Each study was successful in using these models to reliably distinguish between dementia patients and healthy participants. The current state of natural language processing (NLP) research in the specific area of dementia research commonly focuses on the differences between dementia vs. non-dementia speech, and many researchers are successful in this task, as demonstrated above. However, much of this research does not focus heavily on the differences between dementia patients based on biological, social, and cultural factors. As previously mentioned, it is known that cultural identity has an impact on dementia patients, hence the adaptation of the Pyramids and Palm Trees test. Moreover, there is space in dementia research for the consideration of how other factors, such as gender, sexual identity, and race, influence the progression and diagnosis of the disease. Gender has been examined in past studies, and has often been dismissed as having no significant bearing on differences in dementia between the two sexes. However, there are researchers that disagree with this notion, and who have attempted to contribute to this gap in our understanding of dementia. The following section reviews the relevant literature on the intersection of gender and dementia.

### 1.4 Dementia and Gender

Despite the notion  that much of the existing dementia research today does not emphasize the gender and sex differences between men and women, some researchers believe that this is an oversight. While it is known  that dementia affects both men and women, the reasons how and why women are disproportionately diagnosed with dementia have not been adequately explored. As mentioned previously, this disproportion is often related to the finding that women have a longer lifespan than men, with aging being the cited cause for dementia. However, there are researchers who are not convinced that this is the case, arguing that we have yet much to learn about dementia and how it develops in both women and men. One important distinction to note at this point is the difference between "sex" and "gender". The National Academy of Medicine definitions of sex and gender where sex refers to the biological and physiological differences between women and men, with the sex chromosomes (X and Y) contributing to these differences. Gender refers to a combination of environmental, social, and cultural influences on women and men, where gender is rooted in biology but it is primarily shaped by environment and experience (Mielke, 2018). Dr. Michele Mielke called for researchers to take a deep look at this issue in her 2018 paper titled "Sex and Gender Differences in Alzheimer's Disease Dementia". Mielke argues that while the frequency, or count, of Alzheimer's dementia is higher in women, sex differences in the incidence of dementia are less clear. She defines "incidence" as the total number of individuals developing a disease in a given time period divided by the number of people at risk of the disease. In the United

States, the majority of studies reported that the incidence of AD dementia does not differ by sex, even after the age of 85. However, in other areas of the world, women do appear to have a higher incidence of dementia. For example, studies in several European countries report that women have a higher incidence of Alzheimer's dementia after the age of 80. The Cognitive Function and Aging Study in the United Kingdom revealed contrasting results, and initially reported a higher incidence for men. This conflicting data demonstrates that research is needed on a global scale to better understand not only the sex and gender differenes in dementia, but also generational, geographical, and cultural differences as well (Mielke, 2018). Mielke explains that even if women and men have the same incidence of Alzheimer's dementia (age- and sex-specific incidence rate), the mechanisms, pathways, and risk factors can still differ. Mielke points out the many studies examining risk factors for dementia that claim to adjust for sex in the analyses, but mentions that they do not determine whether there are actual sex differences, such as whether the strength of the association between a risk factor and Alzheimer's dementia differs by sex. There are multiple scenarios by which sex and gender differences could affect the risk of Alzheimer's dementia:

- *Sex differences in risk factors:* Women are twice as likely to suffer from depression as men are. Depression can negatively affect the brain in mood and memory over a lifetime, in addition to also increasing the risk of developing dementia up to 70%. Men and women are both susceptible to depression, however, as women face a greater risk, "this greatly impacts their overall risk of dementia" (Mielke, 2018). Interestingly, more women with dementia suffer from depressive symptoms than their male counterparts. Sleep apnea and generally poor sleep patterns have also been connected to later developing dementia and overall cognitive decline. Men have a greater overall prevalence of sleep apnea, although women also have a higher risk of developing sleep apnea post-menopause. "The effect of sleep apnea affecting more men than women points to a higher risk of developing dementia" (Mielke, 2018).
- *Gender differences in risk factors:* Studies have consistently shown that lower levels of education in both men and women are a risk factor in developing dementia later on in life (Karp et al, 2004). Historical trends in access to education illustrate that men, until recently, had greater access to education than women. Recently, there have been improvements in access to education, and women are starting to outnumber men in higher education institutions. It is possible that this increase in female education rates relates to an overall decrease in women developing dementia in some countries (Mielke, 2018).

- *Sex-specific risk factors:* Menopause is a universal event for all cisgender women. The menopausal transition has been associated with a decrease in verbal memory, and early menopause (either natural or surgical) has an increased risk of dementia. Specifically, the abrupt loss of the ovarian hormones via bilateral oophorectomy has been associated with both an increased risk of dementia, and accelerated aging (Rocca et al, 2010). In contrast, prostate cancer is specific to cisgender men. An estimated 50% of all men suffering from prostate cancer are prescribed androgen-deprivation therapy (ADT). Some studies suggest ADT increases the risk of cognitive decline and dementia, although others claim it to be safe (Gonzalez et al, 2015).

While Mielke's research focused on the biological, cultural, and environmental factors and their impact on overall sex-specific risk of men and women developing dementia, Chapman et al (2011) experimented with the verbal episodic memory and recall capacities of both elderly men and women and how they compare to their dementia-suffering counterparts. They used the Logical Memory (LM) subtest of the Wechsler Memory Scale-III to determine how gender differences in AD compare to those seen in normal elderly individuals, and whether or not these differences impact assessment of AD. The LM test was administered to both an AD and a control group, each composed of 21 men and 21 women. The groups had similar demographics, where the control men had an average of 15.9 years of education, control women had an average of 14.9, AD men had an average of 15.1, and AD women had an average of 14.0 years of education. Their Mini Mental State Exam (MMSE) scores, used to gage normal cognition to mild-acute dementia, were 28.9 (average) for control women and 27.8 (average) for men (with a score of 30 being the healthiest), and AD women were 24.3 (average) and AD men were 24.4 (average), falling in the moderate range for dementia. As expected, they found a large drop in performance from non-dementia afflicted elderly individuals to those with AD. Of interest was a gender interaction whereby the women's scores dropped 1.6 times more than the men's did. Control women, on average, outperformed control men on every aspect of the test, including immediate recall, delayed recall, and learning. Conversely, AD women tended to perform worse than AD men. Much like Mielke (2018), Chapman et al acknowledge the controversy around gender-based differences in cognition and behavior in individuals with Alzheimer's disease, and cited many sources that focused on gender differences in AD. However, when in the context of their research, Chapman et al (2011) acknowledge that the implications of gender effects on the assessment of memory impairments is not well known, but how men and women might disparately perform on cognitive tests could impact the diagnoses and conclusions reached by clinicians and researchers. Additionally, the LM achieved perfect diagnostic accuracy in discriminant analysis of AD vs. control women, which was significantly higher for men. The results indicate the LM is a more powerful and reliable tool in detecting AD in women than in men.

As mentioned previously, there is a lack of consensus regarding the role of gender in the progression and diagnosis of Alzheimer's disease and other forms of dementia. However, as shown in the summary of relevant literature above, there are some findings that illustrate the importance of considering the ways in which gender may influence dementia. There are a vast array of social, demographic, physiological, and psychological variables to explore that may or may not support this claim. However, one piece of common ground amongst these works continues to be the possibility of linguistic differences between genders and how researchers can quantitatively measure and examine them, while qualitatively determining the value of these results. In this paper, this disparity is addressed  by examining the linguistic differences between men and women diagnosed with mild to moderate AD according to quantitative measures.

# CHAPTER 2: METHOD AND DATA

The data used for this study was provided by DementiaBank, a widely used academic database for dementia research, as stated in the previous section. Although DementiaBank is a multimedia resource, for this project, only the textual patient transcripts were utilized; the audio and visual media files were not leveraged. DementiaBank has a small catalog of datasets for dementia researchers to choose from, mostly in English with a few datasets being in Mandarin, Spanish, German, and Taiwanese. This project considered data from the Pitt corpus, as many other researchers cited in this paper also used the Pitt corpus for their research.

## 1.1 The Pitt Corpus

The Pitt corpus was created by Dr. James Becker and Dr. Francois Boller in the United States over the course of several years. They began recruiting participants from a variety of sources between 1983 and 1988. Every participant in the study, patients and controls, all underwent an extensive neuropsychiatric evaluation, each with approximately three sessions conducted over the course of two weeks. Every participant in the study was interviewed by a psychiatric nurse to assess their physical and cognitive limitations, as well as the workload their primary caregiver completed in their care. In addition to the psychiatric evaluations, the participants also had to complete various laboratory studies, including the standard hematological studies, blood chemistry studies, liver and thyroid function tests, vitamin levels, and rapid plasma reagin tests. Participants also underwent electroencephalogram (a non-invasive electrical monitoring used to test brain activity), a computed tomogram of the head (also known as a CT scan), and a roentgenogram of the chest (also known as a chest x-ray). After all necessary tests were performed, the study team reviewed the results. To qualify as a "patient", the individual only had to demonstrate a history of progressive cognitive and functional decline and an abnormal mental status exam (Becker, Boller et al, 1994). Both physical and mental examinations were used to rule out any participants with confounding conditions, which could have been responsible for the patient's dementia. Becker and Boller (1994) explain their patient evaluation methods and standards by pointing out that at the time of this study, the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association criteria did not exist, and the *Diagnostic and Statistical Manual of Mental Disorders, Third Edition* did not allow for differing presentations of dementia in AD patients. Therefore, greater emphasis was placed on patient history and presentation than on strict adherence to diagnostic criteria. As such, Becker and Boller (1994) also mention that one of the goals of this study was to develop and evaluate these criteria. Ultimately, after all subjects were evaluated, the researchers had chosen 102 control subjects and 204 AD patients of the 319 individuals enrolled in the study. As of March 1, 1992, 75 of the patients had died. Of these deaths, 50 of them had autopsies, and 43% were classified as definite AD. Several patients were found to have other conditions that excluded them from the study, leaving 181 total patients left. Of the 102 control subjects, 5 subjects developed AD after study entry, and were removed, despite not exhibiting any cognitive decline. Below are tables outlining the study eligibility criteria and the final breakdown of study patient characteristics.

*Figure 1: Criteria required to participate in the Becker and Boller study*

| **Study Eligibility Criteria** | |
|---|---|
| *Age* | >44 years |
| *Education* | More than 7th grade |
| *Literacy* | Able to read and write English fluently before dementia onset (patients only) |
| *Pre-existing conditions* | No history of major nervous system disorders such as cerebral trauma, stroke, meningitis, epilepsy, mental retardation, hypothyroidism, excessive drug use (including alcohol), malnutrition, and major psychiatric disorder (i.e. *Diagnostic and Statistical Manual of Mental Disorders, Third Edition*) except dementia |
| *Medication* | No maintained on any neuroleptic or other medication affecting central nervous system functions except antidepressants |
| *MMSE score* | Must be able to cooperate with neuropsychological testing, and initial Mini-Mental State Examination score of ≥10 |
| *Informed consent* | Must be able to give informed consent |
| *Informant* | Must have an informant (patients only) |

*Figure 1: Criteria required to participate in the Becker and Boller study*

*Figure 2: Patient characteristics of the Becker and Boller study*

| **Study Patient Sample Characteristics** | |
|---|---|
| *Original sample* | 204 |
| Vascular component | 7 |
| Non-AD (clinical) | 9 |
| Non-AD (pathologic) | 7 |
| *Final sample* | 181 |
| **Clinical outcomes (Non-AD)** | |
| Depression | 3 |
| Chromosomal abnormality | 1 |
| Mass lesion developed on computed tomographic scan | 1 |
| Ethanol abuse | 2 |
| Parkinson's disease | 1 |
| Not demented | 1 |
| **Neuropathology outcomes** | |
| Total deaths | 76 |
| Autopsies | 50 |
| Definite AD | 43 |
| Progressive supranuclear palsy | 1 |
| Motor neuron disease with dementia | 1 |
| Creutzfeld-Jakob disease | 1 |
| Striatonigral disease | 1 |
| No distinct histopathological features | 3 |

*Figure 2: Patient characteristics of the Becker and Boller study*

In addition to being well-curated, the Pitt corpus is an attractive choice for researchers, as it includes both dementia and control data for four language tasks: the Cookie Theft picture, verbal fluency, sentence construction, and story recall, all collected from a large longitudinal study. Because  this corpus is quite large with varying research goals, the primary focus of the present study is the Cookie Theft sub-task, where patients with dementia and their control group counterparts were asked to describe the famous Cookie Theft picture (displayed below) over time, with their scores reevaluated at each exam.

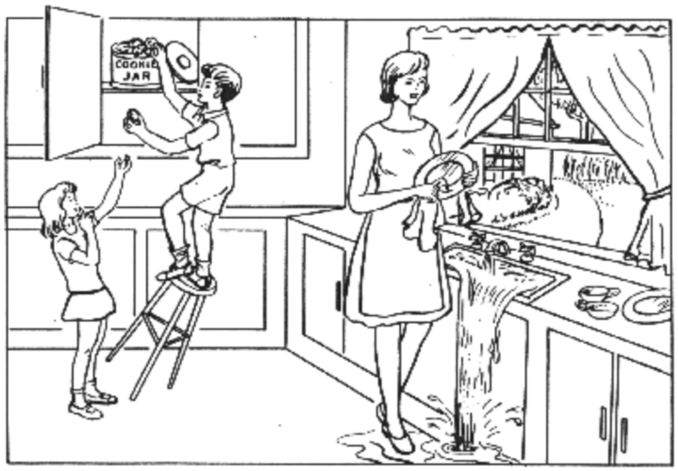


*Figure 3: the Cookie-Theft photo originally from the Boston Diagnostic Aphasia Examination*

The Cookie-Theft section of Pitt corpus evaluates participants' ability to describe the above illustration over the course of time, with evaluation scores given each time. In addition to evaluating participants with MMSE scores, Becker and Boller (1994) also evaluated their participants with other cognitive and memory exams, including the Clinical Dementia Rating scale, the Blessed-Ross Dementia Scale, and the Hamilton Rating Scale for Depression, among others.

**1.2 The Mini-Mental State Examination**

According to the Alzheimer's Society, the Mini-Mental State Examination (MMSE) is the most commonly used test for problems with memory or other mental abilities. MMSE was first published in 1975 by M. F. Folstein et al as an appendix in their study, *The Mini-mental state: A practical method for grading the cognitive state of patients for the clinician.* MMSE can be used by clinicians to help diagnose dementia and to assess its progression and severity, in conjunction with other comprehensive tests. The exam consists of a series of questions and tests, with points earned when the questions are answered correctly. The exam is performed the same way each time, with scripted questions and comments. Questions are seemingly simple, such as asking the patients the date, season, date of the week, or what town and state they're located in. Despite the questions being straightforward and simple, the scoring pays close attention to the amount and quality of information the subject is able to produce. For example, when a patient is asked for the date, the question must be posed as, *"What is the date today?"* and participants may earn up to three total points for their answer: one point for the month, one point for the day, and one point for the year. After asking for the date, an additional related question may be asked specifically for parts of the date that the participant omitted, such as *"Can you also tell me what month, year it is?"* allowing the participant to earn the additional points allotted for this question. This is the only follow up question asked for this standard question, and the evaluator must record and move on once the participant has given their answer. MMSE also includes open-ended questions, such as asking the participant to make up a sentence, and even tasks such as copying a simple drawing. MMSE scores fall on a rating scale of 0-30, and scores are grouped accordingly. 25-30 is "questionably significant", 20-25 is "mild", 10-20 is "moderate", and 0-10 is "severe". Therefore, the lower one's MMSE score is, the more profound their dementia is (M. F. Folstein et al, 1975).

### 1.3 Data Summary

Data collected from a sample of the Pitt corpus were utilized for both the statistical analysis as well as the machine learning classifier evaluation sections of this analysis. For the present study, there is a focus on those participants who fell within the mild and moderate MMSE range, i.e., who received scores ranging from 10-25. Therefore, any subjects who fell outside of this range were eliminated. In an effort to keep the data balanced between the sexes, the number of men and women were limited in the data to be 51 participants each, as to not influence one sex over the other in our experiments. Participants were also evaluated only after their first visit for the present study, as the main goal of this analysis is to focus on gender, rather than changes over time. Below is a table highlighting important features of our participant population, separated by sex.

*Figure 4: notable features of the participant data used in the present study*

| **Feature** | **Male** | **Female** |
|---|---|---|
| **Age** | 70.28 years | 70.48 years |
| **Onset Age** | 66.48 years | 66.86 years |

| | | |
|---|---|---|
| **Education** | 13.02 years | 12.14 years |
| **MMSE Score** | 22.02 | 20.46 |

*Figure 4: notable features of the participant data used in the present study*

The data in the table above, illustrates that the groups considered are fairly comparable in both age and onset age categories, however the male participants average about one additional year of education than their female counterparts, and average about 1.5 points higher in their MMSE score than their female counterparts.

### 1.4 Statistical Analysis

In order to understand the ways in which linguistic features varied based on gender, a linear regression was completed in (Scikit-learn, Version 0.23.2). We identified both linguistic and demographic features that have been observed in past research, including Orimaye et al (2017). In addition to the low-level linguistic features, MMSE score from each participant was also examined, including how these features interacted with the patients' sex. For the regression analysis, the transcripts of each patient was reviewed and the count for each linguistic feature was noted, along their MMSE score and sex. For the regression analysis, MMSE score was used as the dependent variable, with sex, linguistic features, and their interactions as independent variables. The linguistic variables were centered to mitigate multicollinearity by subtracting their means.

*Figure 5: results of the ordinary least squares (OLS) regression analysis*

| | **Coefficient** | **Standard Error** | **t** | **P > \| t \|** | **[0.025** | **0.975]** |
|---|---|---|---|---|---|---|
| **Intercept** | 20.5205 | 0.606 | 33.85 | 0 | 19.316 | 21.725 |
| **Sex** | 1.439 | 0.876 | 1.642 | 0.104 | -0.303 | 3.181 |
| **Word Count** | 0.1185 | 0.061 | 1.948 | 0.055 | -0.002 | 0.239 |
| **Pronoun Count** | -0.3799 | 0.171 | -2.226 | 0.029 | -0.719 | -0.041 |
| **Noun Count** | -0.0492 | 0.134 | -0.368 | 0.714 | -0.315 | 0.217 |
| **Verb Count** | -0.2044 | 0.173 | -1.182 | 0.24 | -0.548 | 0.139 |
| **Adjective Count** | -0.1682 | 0.262 | -0.643 | 0.522 | -0.688 | 0.352 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Adverb Count** | -0.1963 | 0.161 | -1.216 | 0.227 | -0.517 | 0.125 |
| **Pauses** | -0.298 | 0.17 | -1.756 | 0.083 | -0.635 | 0.039 |
| **Retracing** | -0.5057 | 0.221 | -2.286 | 0.025 | -0.946 | -0.066 |
| **Trailoffs** | 0.2775 | 0.623 | 0.445 | 0.657 | -0.961 | 1.516 |
| **Fragments** | 0.1035 | 0.145 | 0.711 | 0.479 | -0.186 | 0.393 |
| **Shortening** | -0.1212 | 0.142 | -0.855 | 0.395 | -0.403 | 0.161 |

*Figure 5: results of the ordinary least squares (OLS) regression analysis*

According to this analysis, only pronoun count ($p$ = 0.029) and retracing ($p$ = 0.025) are significant, and an R squared of 0.320. This preliminary analysis demonstrates the concept of predicting dementia using linguistic features, which many other researchers have done. Compared to past research, the results here are low. However, in correlation with past research, pronoun count and retracing are found to be significant. This aligns with Kempler et al (2008) review of evidence for the relationship between extralinguistic cognitive and language abilities in dementia. Kempler (2008) described several extralinguistic deficits in the speech of patients with dementia, one of them being the observation that dementia patients have a difficult time constructing an informative and coherent narrative. Their narratives are often repetitive with topic changes, which can be described as retracing. Patients with dementia also tend to use more pronouns than nouns when compared to healthy participants, which our analysis also found to be significant. Again, many researchers have successfully found methods to predict dementia speech and differentiate it from the speech of healthy participants, however it is still revealing that the present study found pronoun count and retracing to be significant.

In order to address our hypothesis specifically around gender differences in patients with dementia, interactions between sex and linguistic features were included in a model. Including the interactions between features allows the model to take into account that features may behave differently for speakers of different genders.

*Figure 6: results of regression model including interactions between linguistic variables and sex*

| | Coefficient | Standard Error | t | P > \| t \| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **Intercept** | 20.4685 | 0.633 | 32.32 | 0 | 19.207 | 21.73 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Sex** | 1.2951 | 0.867 | 1.494 | 0.139 | -0.431 | 3.021 |
| **Word Count** | 0.1553 | 0.099 | 1.572 | 0.12 | -0.041 | 0.352 |
| **Word Count : Sex** | 0.0053 | 0.137 | 0.039 | 0.969 | -0.267 | 0.277 |
| **Pronoun Count** | -0.3805 | 0.239 | -1.595 | 0.115 | -0.856 | 0.095 |
| **Pronoun Count : Sex** | -0.4131 | 0.372 | -1.112 | 0.27 | -1.153 | 0.327 |
| **Noun Count** | -0.23 | 0.193 | -1.194 | 0.236 | -0.613 | 0.154 |
| **Noun Count : Sex** | 0.1545 | 0.294 | 0.526 | 0.6 | -0.431 | 0.74 |
| **Verb Count** | -0.1299 | 0.233 | -0.556 | 0.58 | -0.595 | 0.335 |
| **Verb Count : Sex** | -0.422 | 0.403 | -1.048 | 0.298 | -1.224 | 0.38 |
| **Adjective Count** | -0.5674 | 0.338 | -1.68 | 0.097 | -1.24 | 0.105 |
| **Adjective Count : Sex** | 0.9284 | 0.584 | 1.59 | 0.116 | -0.235 | 2.091 |
| **Adverb Count** | -0.0773 | 0.258 | -0.3 | 0.765 | -0.591 | 0.436 |
| **Adverb Count : Sex** | -0.0966 | 0.36 | -0.268 | 0.789 | -0.814 | 0.621 |
| **Pauses** | -0.7512 | 0.267 | -2.818 | 0.006 | -1.282 | -0.22 |
| **Pauses: Sex** | 0.7829 | 0.349 | 2.241 | 0.028 | 0.087 | 1.479 |
| **Retracing** | -0.7569 | 0.394 | -1.923 | 0.058 | -1.541 | 0.027 |
| **Retracing : Sex** | -0.0335 | 0.489 | -0.068 | 0.946 | -1.008 | 0.941 |
| **Trailoffs** | -0.6022 | 0.858 | -0.702 | 0.485 | -2.311 | 1.106 |
| **Trailoffs : Sex** | 2.0029 | 1.195 | 1.676 | 0.098 | -0.378 | 4.384 |
| **Fragments** | 0.2349 | 0.226 | 1.04 | 0.301 | -0.215 | 0.685 |
| **Fragments : Sex** | 0.0854 | 0.309 | 0.276 | 0.783 | -0.53 | 0.701 |
| **Shortening** | 0.141 | 0.265 | 0.532 | 0.596 | -0.387 | 0.669 |
| **Shortening : Sex** | -0.4135 | 0.333 | -1.243 | 0.218 | -1.076 | 0.249 |

*Figure 6: results of regression model including interactions between linguistic variables and sex*

There is only one significant interaction, which is the interaction between pauses and sex. Overall, his model has an R squared of 0.469, which explains the data better than the previous analysis. This model is able to consider how each variable interacts with sex, therefore viewing male and female patients differently. Another interesting interaction is that between trailoffs and sex. While it is not considered significant, it approaches significance ($p = 0.098$). Finally, an ANOVA model was utilized to compare the two previous regression models. This model included the interactions and yielded significant results.

*Figure 7: ANOVA model comparing both regression models*

| | Resid DF | SSR | DF Difference | SSR Difference | F | Pr( > F) |
|---|---|---|---|---|---|---|
| **OLS Regression Model** | 87 | 1526.454087 | 0 | NaN | NaN | NaN |
| **Interactions Regression Model** | 76 | 1191.940519 | 11 | 334.513568 | 1.93901 | 0.047085 |

*Figure 7: ANOVA model comparing both regression models*

The ANOVA above supports our hypothesis that gender should be considered in a dementia diagnosis. The Pr( > F) value for the interactions regression model is 0.047085, meaning that the model including interactions with gender is significantly better.

**1.5 Classifier Evaluations**

In addition to the statistical analysis, supervised machine learning algorithms were trained and tested in order to understand the differences between the two sexes in the data set. Machine learning classifiers have been utilized in a variety of human language-centric tasks, including those focused on dementia and Alzheimer's disease. The textual transcripts from the corpus's provided files (in CHAT format, which is specific to DementiaBank) were extracted, and then arrange by transcript and tag, with the tag being the participant's sex. Next, the tags and transcriptions were loaded into a dataframe, then transformed into feature vectors using various feature engineering methods. First, count vectors were created, in which every row represents a transcript from the corpus, every column represents a term from the corpus, and every cell represents the frequency count of a particular term in a particular document. Next, three types of TF-IDF vectors were created: word-level, ngram-level, and character level. TF-IDF scores represent the relative importance of a term in the document and the entire corpus. TF, *term frequency*, represents the number of times a term *t* appears in a document divided by the total number of terms in a document. TF is then divided by IDF, *inverse document frequency*. IDF is the logarithm of the total number of documents divided by the number of documents where the term *t* appears. The various TF-IDF matrix represent various features:

- *Word-level TF-IDF:* matrix represents TF-IDF scores of every term in different documents.
- *N-gram level TF-IDF:* matrix represents the TF-IDF scores of n-grams, which are the combination of *n* terms together.
- *Character-level TF-IDF:* matrix represents TF-IDF scores of character-level n-grams in the corpus.

These vectors then underwent part of speech tagging to examine the various low-level linguistic features that were identified. After performing feature engineering, the next step in the experiment was to use these features to train various classifiers. Three separate algorithms were trained on these features: Naive Bayes, Logistic Regression, and Support Vector Machine (SVM), each of which performed at roughly the same level. The results of this are presented in the table below.

*Figure 8: Machine learning classifier results*

| Model | Accuracy |
| --- | --- |
| **Naive Bayes** | |
| *Count Vectors* | 0.44 |
| *Word-Level Vectors* | 0.40 |
| *N-Gram Vectors* | 0.48 |
| *Character-Level Vectors* | 0.40 |
| **Linear Regression** | |
| *Count Vectors* | 0.56 |
| *Word-Level Vectors* | 0.40 |
| *N-Gram Vectors* | 0.40 |
| *Character-Level Vectors* | 0.44 |
| **SVM** | |
| *Count Vectors* | 0.48 |
| *Word-Level Vectors* | 0.40 |
| *N-Gram Vectors* | 0.40 |
| *Character-Level Vectors* | 0.40 |

*Figure 8: Machine learning classifier results*

The above results range from 0.40 to 0.56 accuracy. In this analysis, the classifiers were trained on both male and female utterances, and had to distinguish between the two categories. The top bigrams for each group were also collected, shown below. Bigrams were collected in attempts to identify any similarities or differences in speech between the two sexes, and the results were quite similar. For example, bigrams such as (‘cookie’, ‘jar’), (‘the’, ‘cookie’), (‘the’, ‘dishes’), (‘the’, ‘water’), (‘the’, ‘floor’) and (‘the’, ‘stool’) were all noted in both sets of bigrams.

*Figure 9: top bigrams for each sex*

| Sex | Bigram | Frequency |
|---|---|---|
| Female | 'and', 'the' | 62 |
| | 'cookie', 'jar' | 50 |
| | 'the', 'cookie' | 43 |
| | 'the', 'sink' | 39 |
| | 'on', 'the' | 39 |
| | 'in', 'the' | 34 |
| | 'the', 'water' | 33 |
| | ‘the', 'floor' | 30 |
| | 'the', 'little' | 28 |
| | 'the', 'dishes' | 25 |
| Male | 'on', 'the' | 56 |
| | 'and', 'the' | 53 |
| | 'the', 'sink' | 46 |
| | 'cookie', 'jar' | 44 |
| | 'in', 'the' | 42 |
| | 'the', 'cookie' | 36 |

| | | |
|---|---|---|
| | 'the', 'water' | 28 |
| | 'the', 'stool' | 26 |
| | 'the', 'floor' | 26 |
| | 'the', 'dishes' | 23 |

*Figure 9: top bigrams for each sex*

# CHAPTER 3: DISCUSSION AND CONCLUSION

The present study aimed to determine if there was any significant difference in the linguistic features of male and female patients with dementia, and evidence to support this hypothesis was identified. Specifically, the statistical analyses performed uniquely imagined the data in a more common regression analysis (the first OLS regression analysis), and also took into consideration interactions (the second interactions regression analysis). Both analyses yielded some significant results, however the ANOVA analysis comparing both regression models demonstrated the significance gender has on speech in dementia patients. The machine learning classifiers did not yield as strong results as the statistical analysis, however they illustrate how both sexes display subtle interactions between their linguistic features. One missing dimension to this study is time. The participants evaluated in the present study only had one transcript each from their first visit with Becker and Boller. Other studies in this space accounted for time, and this is perhaps why they had higher-scoring models. Measuring how dementia patients' speech changes over time could give more insight to how the two sexes differ not just overall, but also in how the disease progresses.

As noted in previous sections, women have historically been at higher risk for dementia due to several social and biological factors, including education (Mielke 2018 and Chapman et al, 2011). As noted in the participant summary, while both male and female groups were approximately the same age at the time of the study and even developed dementia at around the same age, the female participants averaged 1.5 points lower in their MMSE scores, meaning that their dementia was found to be, on average, more profound than their male counterparts. It was also reported that these female participants also had on average, fewer years of education than the male participants, which could have had an effect on their dementia. Education is regarded as a factor in whether or not an individual will develop dementia, with more years of education benefitting the person and potentially lowering their risk (Mielke, 2018). This observation ties back to the research performed by Chapman et al (2011), in which dementia participants had a similar education imbalance between their gender groups. The authors noted that female patients performed worse than male patients.

It is also important to reflect on the time at which Becker and Boller's study was completed, and these participants were being interviewed. The Pitt corpus' data collection began in the early 1980s, meaning that these individuals lived through and were aging in a completely different society than today. For example, it is possible that many of these participants partook in a traditional lifestyle, in which  men worked out of the home, and the women were homemakers. While having a career outside of the home is not widely discussed as a contributing factor in dementia, it could have an effect that goes unknown simply because society has not produced subjects with that differentiator quite yet. An increasing number of women are choosing to pursue higher education and work outside of the home. However, this shift may not be reflected in the participants for this study, if career has any bearing at all on the development and/or progression of dementia.

Aside from educational and career shifts between the sexes, there is also a large gap in this data concerning race and the LGBTQIA community. In the Pitt corpus, out of all of the participants, only 12 of them were of a different race other than Caucasian. This demonstrates a sizable gap in the dataset, as the CDC reported in 2014 that, of Americans 65, and older, African Americans have the highest prevalence of Alzheimer's disease and related dementias (13.8%), followed by Hispanics (12.2%), and non-Hispanic whites (10.3%), American Indian and Alaska Natives (9.1%), and Asian and Pacific Islanders (8.4%). These populations with dementia are expected to continue growing, as the CDC predicts that by 2060, 3.2 million Hispanics and 2.2 million African Americans will have Alzheimer's disease and related dementias. However, this will likely be the result of less people dying of chronic illnesses and living longer, while age will continue to be a risk factor for dementia. Since there are so few participants that identify as any race other than Caucasian in our data, it is difficult to extrapolate any meaningful conclusions from our experiment in regards to race, as it relates to gender, in linguistic markers of dementia. Similarly, we also do not know from the Pitt corpus if they involved any participants who identify as part of the LGBTQIA community. In 2019, new research was reported showing higher rates of cognitive decline among gay, lesbian, bisexual, and/or transgender Americans compared to their cisgender, heterosexual counterparts. This is a relatively new assertion, as the first data reporting the prevalence of dementia amongst sexual and gender minorities was submitted to the Alzheimer's Association International Conference (AAIC) in 2018, according to their Chief Science Officer, Dr. Maria C. Carrillo. Important social factors were uncovered in these studies as to why many LGBTQIA individuals suffer from dementia, including living alone (~60%), not being partnered or married (65%), not having children (72%), and not having a caregiver (59%) when compared to cisgender, heterosexual adults also living with dementia. It is also noted that this higher rate of dementia among LGBTQIA-identifying individuals may be due to them "not seeking treatment after a lifetime of discrimination, victimization, and bias" according to Karen Fredriksen-Goldsen, PhD, professor and director of Healthy Generations Hartford Center of Excellence at the University of Washington.

### 3.1 Conclusion

As society progresses and changes, it is vital that more emphasis is put on researching not just how people from different backgrounds develop Alzheimer's disease and associated dementias, but also how these differences can be identified and addressed. While it remains true that there are biological factors beyond old age at play in the development of dementia, there is still much to be uncovered about how social factors shape our risk for developing dementia, and how dementia is expressed by people who have differing combinations of these factors. The results of the present study illustrate the possibility of marked and significant differences between linguistic dementia markers based on the gender of the patient. How do these markers change when other layers of complexity such as race, ethnicity, sexuality, or gender expression are also taken into account? Researching these areas would not only be an advancement in the field, but would promote equity and equality within healthcare as well.

REFERENCES

Association, Alzheimer's. “Increased Risk of Subjective Cognitive Decline in the Lgbt Community.” *Increased Risk Of Subjective Cognitive Decline In The LGBT Community*, Cision PR Newswire, 14 July 2019, www.prnewswire.com/news-releases/increased-risk-of-subjective-cognitive-decline-in-the-lgbt-community-300884444.html.

Association, Alzheimer’s. “What Is Dementia?” *Alzheimer's Disease and Dementia*, www.alz.org/alzheimers-dementia/what-is-dementia.

Becker, J. T., Boller, F., Lopez, O. L., Saxton, J., & McGonigle, K. L. (1994). The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis. *Archives of Neurology*, 51(6), 585-594.

Grant support for the Pitt corpus: NIA AG03705 and AG05133.

Berube, Shauna et al. “Stealing Cookies in the Twenty-First Century: Measures of Spoken Narrative in Healthy Versus Speakers With Aphasia.” *American journal of speech-language pathology* vol. 28,1S (2019): 321-329. doi:10.1044/2018_AJSLP-17-0131

Chapman, Robert M et al. “Women have farther to fall: gender differences between normal elderly and Alzheimer's disease in verbal memory engender better detection of Alzheimer's disease in women.” *Journal of the International Neuropsychological Society : JINS* vol. 17,4 (2011): 654-62. doi:10.1017/S1355617711000452

Folstein, M F et al. “"Mini-mental state". A practical method for grading the cognitive state of patients for the clinician.” *Journal of psychiatric research* vol. 12,3 (1975): 189-98. doi:10.1016/0022-3956(75)90026-6

Gonzalez, Brian D et al. “Course and Predictors of Cognitive Function in Patients With Prostate Cancer Receiving Androgen-Deprivation Therapy: A Controlled Comparison.” *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* vol. 33,18 (2015): 2021-7. doi:10.1200/JCO.2014.60.1963

Karlekar, Sweta, et al. “Detecting Linguistic Characteristics of Alzheimer’s Dementia by Interpreting Neural Models.” *ArXiv*, 17 Apr. 2018, arxiv.org/pdf/1804.06440.pdf.

Karp, Anita et al. “Relation of education and occupation-based socioeconomic status to incident Alzheimer's disease.” *American journal of epidemiology* vol. 159,2 (2004): 175-83. doi:10.1093/aje/kwh018

Kempler, Daniel, and Mira Goral. “Language and Dementia: Neuropsychological Aspects.” *Annual review of applied linguistics* vol. 28 (2008): 73-90. doi:10.1017/S0267190508080045

MacWhinney, Brian et al. “Automated analysis of the Cinderella story.” *Aphasiology* vol. 24,6-8 (2010): 856-868. doi:10.1080/02687030903452632

Mehri, Azar et al. “Normative data for the Pyramids and Palm Trees Test in literate Persian adults.” *Iranian journal of neurology* vol. 17,1 (2018): 18-23.

Mielke, Michelle M. “Sex and Gender Differences in Alzheimer's Disease Dementia.” *The Psychiatric times* vol. 35,11 (2018): 14-17.

Rocca, Walter A et al. “Oophorectomy, menopause, estrogen treatment, and cognitive aging: clinical evidence for a window of opportunity.” *Brain research* vol. 1379 (2011): 188-98. doi:10.1016/j.brainres.2010.10.031

“U.S. Burden of Alzheimer's Disease, Related DEMENTIAS to Double by 2060.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 20 Sept. 2018, www.cdc.gov/media/releases/2018/p0920-alzheimers-burden-double-2060.html.

“Vascular Dementia.” *Harvard Health Publishing*, Harvard Medical School, 15 Dec. 2014, www.health.harvard.edu/alzheimers-and-dementia/vascular-dementia.