OBSERVING LANGUAGE CHANGES IN AGING AND ALZHEIMER'S SPEECH USING INFORMATION THEORY TECHNIQUES

A Thesis

Presented to the

Faculty of

San Diego State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Linguistics

by

Roselene M. Freeman

Fall 2015

SAN DIEGO STATE UNIVERSITY

The Undersigned Faculty Committee Approves the

Thesis of Roselene M. Freeman:

Observing Language Changes in Aging and Alzheimer's Speech Using

Information Theory Techniques

Robert Malouf, Chair
Department of Linguistics and Asian/Middle Eastern Languages

Eniko Csomay
Professor and Associate Dean, College of Arts and Letters

Paul Gilbert
Department of Psychology

June 30, 2015
Approval Date

Copyright © 2015

by

Roselene M. Freeman

All Rights Reserved

DEDICATION

To my grandfather and grandmother, who have had their lives affected by Alzheimer’s, and my parents, to whom I hope the ongoing research benefits.

ABSTRACT OF THE THESIS

Observing Language Changes in Aging and Alzheimer’s Speech
Using Information Theory Techniques
by
Roselene M. Freeman
Master of Arts in Linguistics
San Diego State University, 2015

This corpus study examines entropy via the model of perplexity in regards to the spontaneous speech of adults as they age, and compares this with the overall entropy of speech produced by those with Alzheimer’s disease. Switchboard and Buckeye corpora were used for typical adult speech, and Alzheimer’s speech came from the Carolina Conversations Collection. Two language models, one containing speech from adults under 40 and one containing speech from those over 40, were created in the SRI Language Modeling Toolkit and used to calculate perplexity from individuals randomly taken from different decade spans in the Switchboard corpus and from 33 individuals with Alzheimer’s from the CCC. Python NLTK was used to calculate various speech characteristics, such as mean length of utterance, type/token ratio, low specificity words, fillers, repetitions, and part of speech proportions. There was a general upward trend in perplexity as age increased, and this effect remained consistent in both the younger language model and the older language model. In terms of speech characteristics, mean length of utterance and usage of nouns had statistically significant effects in both the younger and the older language model.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

ACKNOWLEDGEMENTS

First of all, I'd like to thank my wonderful committee, Dr. Robert Malouf, Dr. Eniko Csomay and Dr. Paul Gilbert for their patience and support. I'm grateful as well to all of those involved with the corpora that make up this study, especially Dr. Charlene Pope and Dr. Boyd Davis for their help with the Carolina Conversations Collection.

I'd also like to thank Lambert Ninteman for his mentorship in going through the process of writing a thesis. I'm so grateful for your encouragement and advice in navigating the pitfalls.

Finally, last but not least, I thank my supportive family and friends who have all helped me through the rest of my life that surrounded this thesis. I especially thank my loving fiancée Corey Ahlgren for being my day-to-day champion throughout my graduate studies.

CHAPTER 1

INTRODUCTION TO THE STUDY

Dementias such as Alzheimer's disease have been shown to leave evidence of their cognitive effects in the language use of those afflicted, and these characteristics have been found through the computational analysis of speech and written corpora. Lexical and grammatical differences have been found between typical cognitively aged adults and those with Alzheimer's. Entropy, the numerical measure of the average amount of information contained within a message, would be affected by factors such as grammar and vocabulary. May entropy be a distinguishable factor between the speech of those with or without Alzheimer's? Since cognitive changes occur with aging, could aging affect entropy as well?

With the exception of one study that examined cross-entropy of grammar (Roark, Mitchell, Hosom, Hollingshead, & Kaye, 2011) previous research has not examined the overall entropy of Alzheimer's versus non-Alzheimer's speech, nor has it examined the entropy of speech as a person ages. Therefore, this study will add to previous corpus studies by examining the entropy of speech from three different corpora: the Buckeye Corpus and the Switchboard corpus containing speech from younger to older healthy adults, and the portion of the Carolina Conversations Collection containing Alzheimer's speakers. This entropy will be then compared with relevant lexical characteristics derived from previous research on Alzheimer's speech.

1.1 BACKGROUND AND DEFINITIONS

Central to this investigation is the concept of entropy. Developed by Claude Shannon, the entropy of information theory corresponds to the average amount of information contained in a message. Shannon considered entropy to correlate with surprisal, since more surprising messages contain more information than more predictable messages. Hence, entropy is based on the log of the probability (p) of the next message in a sequence (x) and it is defined by the following formula (Shannon, 1948):

$$H(X) = -\sum_{x} p(x) \log_2 p(x) \qquad (1.1)$$

However, the difficulty of measuring the entropy of a language is that one would need an infinitely long sample, as n → ∞ as its limit (P. F. Brown, Della Pietra, Mercer, Della Pietra, & Lai, 1992; Shannon, 1951). One also needs to assume that language is ergotic (has the same average behavior across all states) and stationary (its probabilities do not shift over time). Therefore for our research purposes we will examine entropy through using the perplexity of a language model rather than calculating the entropy of the text directly. Perplexity is a way to evaluate the effectiveness of a language model by calculating the exponent of the cross entropy of a word as produced by the model (Jurafsky & Martin, 2008). Perplexity is calculated using the following formula, with 2 as the base to the power of entropy, x representing the next message of a sequence and p is the probability of that message:

$$2^{H(p)} = -2^{-\sum_{x} p(x) \log_2 2^{H(p)}} \qquad (1.2)$$

This study will be specifically using the average perplexity per word as a means of conceptualizing the average rate of entropy.

1.2 Statement of the Problem and Purpose

As the total entropy of speech in relation to age has not been previously studied, this leaves some unanswered questions. Could overall entropy change as one ages, and could it be another distinguishing characteristic between the speech of those with and without Alzheimer’s? Is entropy correlated with the other lexical characteristics found in previous research? The prediction is that entropy will increase with typically aging participants, as vocabulary knowledge tends to increase with age (Kave & Halamish, 2015) and decrease with Alzheimer's due to the high amount of repetitions and fillers associated with this speech (Le, Lancashire, Hirst, & Jokel, 2011).

This information could potentially be very useful in building a computational model for Alzheimer's that may help supplement other examinations such as the Mini Mental State Exam (MMSE). If there is a relevant difference between the overall entropy of Alzheimer's and non-Alzheimer's speech, that could be another criterion for a computational model to use to distinguish between the two kinds of speech, in addition to other linguistic factors. If age is found to be a factor on entropy, this could also be useful information in designing such a model because that model would have to take aging entropy into account.

1.3 Theoretical Basis and Organization

The main inspiration for this research comes from previous corpus studies done with speech and written data. The Nun Study (Snowdon et al., 1996) has evidence that grammar use in early adulthood detected through written corpus analysis may be correlated with the development of Alzheimer's disease later in life. Another study revealed possible differences in grammar and vocabulary by corpus analysis of novels from an author with Alzheimer's disease and an author without Alzheimer's (Le et al., 2011). A potential computational model of Alzheimer's has shown that cross-entropy analysis of grammar could be useful (Guinn & Habash, 2012).

Entropy models have been used before in both medical and non-medical contexts, particularly to test models that are designed to recognize patterns. One of the uses of entropy is data compression, where it determines the maximum amount of compression that data can undergo (Balakrishnan & Touba, 2007). It is also involved in detecting unusual heart rate variations (Aktaruzzaman & Sassi, 2014) tracking the path that a person walks (Sinatra & Szell, 2014) among many other applications.

1.4 Limitations of the Study

One of the major limitations of this study is that the corpora used are not consistent in regards to their specific data kept. While the Switchboard corpus keeps records of the education level and dialect areas of the participants, this corpus lacks speech data for healthy older adults from 70+ onward. The Switchboard corpus used in this study begins from 20 years old all the way to 64 years old, and finding data for 60-year-old segment was difficult—only 4 individuals could be included. There is also incomplete data on the Alzheimer’s subset of the Carolina Conversations Collection, as age and education data was not available for all the subjects. While age data is available for most of the CCC speakers, over half of them lack education data, so education levels could not be controlled. The sample sizes also varied, from 1 sample for most individuals in the corpus to 13 samples for one individual.

The other main limitation of this study is that it examines a corpus of spontaneous speech from a wide selection of people, using speech data as a sample for a particular age and Alzheimer’s status. A single participant could not be followed for a large length of time, so it’s unknown if any subject in the Switchboard or Buckeye corpus developed Alzheimer’s after the data was recorded. Since they’re spontaneous conversations the topics of each dialogue are unique, and in the case of Switchboard a topic was selected for the participants to discuss. Therefore, it is impossible to separate the potential effect of topic on the speech of our subjects.

CHAPTER 2

REVIEW OF THE LITERATURE

Our current study focuses on three main areas of previous research: the speech of healthy adults, the speech characteristics of Alzheimer's, and aspects of the concept of entropy. Cognitive aging plays an important part in how speech is affected in older adults, and Alzheimer's disease accelerates and exaggerates aspects of typical aging. Here we examine this process and how it affects speech, as well as research into entropy models for speech.

2.1 Speech in Healthy Older Adults

As people age, we know that many changes are occurring in both sensory and cognitive perception. There are many aspects to speech that can all be affected by this process. Creating speech requires that one is able to follow along the context of what is being spoken, remember the words that are about to be said, put them together grammatically, and able to access the phonological representation of each word. Each of these different aspects to language production must be considered in older adults, and may each be affected differently.

One of these major changes that can affect speech is simply hearing. It is possible that speech misunderstandings in older adults could be due to changes in hearing rather than cognition errors. Older adults lose their pure tone perception and ability to hear higher frequencies, which can affect perception of aspirated stops (Park & Schwarz, 2000). Time compressed speech, where the rate of speech is increased without changing the prosody, has been shown to be more difficult for older adults to comprehend even when hearing acuity between older and younger adults is matched (Park & Schwarz, 2000). There is also evidence that older people may have more trouble filtering audio input as well, such as in one study where older adults had more problems than younger adults recalling test items while meaningful distracting speech was played in the background (Tun, O'Kane, & Wingfield, 2002).

In neurotypical adults, the process that tends to stay the most consistent throughout aging is language comprehension. Semantic priming—the idea that when presented with a concept, such as *table,* it can reduce the time it takes to identify a semantically related concept, such as *chair*—is fully present in both younger and older adults. This priming effect not only exists for individual words (Giffard, Desgranges, Kerrouche, Piolino, & Eustache, 2003) but in the context of sentence comprehension as well (Kellas, Paul, & Vu, 1995), with older adults having greater priming effects in both studies. When younger and older adults listen to a passage and are asked to stop the passage to recall what they just heard, they both stop more often at clause boundaries and have the same sensitivity to passage content (Wingfield & Tun, 1999).

Older adults are also better at recognizing more vocabulary words and in general have a greater vocabulary. They can produce greater word synonyms and solve more crossword puzzle problems than younger adults (Salthouse, 2004). In one study, researchers measured vocabulary knowledge as well as their confidence of that knowledge in older, middle-aged and younger adults by testing them in a vocabulary test involving Hebrew words. The subjects were asked to define terms in Hebrew and then give a rating from 1 to 7 as to how confident they were in these definitions. Older adults were able to not only give correct definitions for the words but also had more confidence in these definitions (Kave & Halamish, 2015).

The greater vocabulary and consistent language comprehension in older adults may be related to the finding that generally older adults perform better on tasks that rely on context and familiar information than tasks where they are exposed to new information (Burke & Mackay, 1997). This may be the cause of the "hyperpriming" effect, where older adults are more sensitive to semantic priming (Giffard et al., 2003). When learning previously unknown words in the context of a passage, older adults had more complete definitions for the new words than younger adults (Long & Shaw, 2000).

The area of speech and language processing that appears to be affected the most by aging is in language production. One well studied aspect of this is the "tip-of-the-tongue" (TOT) phenomenon, where there is a feeling of "knowing" a word, but not being able to say it. In many studies, older adults have been shown to experience more TOTs than younger adults (Burke & Mackay, 1997; Kellas et al., 1995). This phenomenon can be explained as being able to access the semantic information of a word but unable to access the phonological information. This could potentially be a natural consequence of learning more vocabulary as one ages. Over the years we encounter even more concepts and events, making the mental processing it takes to access concepts that have only been encountered a few times in one's lifetime more difficult (Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014). Indeed, when this difficulty is overcome—such as when an older adult is given cues for a word—the TOT effect reduces. For example, in a study where subjects were asked to name pictures of celebrities, producing a homophone of a name (i.e., "cherry pit") in a previous task helped older adults in producing that name (i.e., "Brad Pitt") but not as much for younger adults (Burke & Shafto, 2004).

Other aspects of language production are also affected more by aging. While the perception of spelling patterns remains constant, older adults have more trouble producing the correct spelling of a word than younger adults (Burke & Mackay, 1997). Older adults have more filled pauses than younger adults do (Burke & Shafto, 2004) and have a decreased word retrieval capacity compared to younger adults (Kave & Yafe, 2014). In a picture description study, Italian speaking adults between 20 and 84 years old were asked to produce sentence description of three different pictures—one single picture and two series of pictures. There were differences in semantic paraphrasias (where one word is substituted for a semantically similar word) and paragrammatic paraphrasias (where one morpheme is confused for another), as it was found that the very oldest adults made significantly more of them (Marini, Boewe, Caltagirone, & Carlomagno, 2005).

Another influence into this process is working memory, which relates to the amount of ideas that a person can hold in memory temporarily. Since one cannot review speech the way one might reexamine a difficult sentence in text, working memory is necessary for speech comprehension. Generally, older adults tend to perform more poorly than younger adults in tasks that rely solely on working memory, such as memorizing vocabulary lists in a lab. However, older adults may have a similar working memory capacity to younger adults, but have trouble accessing that capacity. In one study, various spatial and numerical working memory tasks were performed between older and younger adults. What they found was that older and younger adults had similar working memory capacities, but in terms of processing tasks and accuracy older adults did worse than younger adults (Salthouse, Babcock, & Shaw, 1991; Wingfield & Tun, 1999).

Given that older adults tend to have problems with retrieval and access to items in memory, the traditional model for cognitive aging has been the idea of a general cognitive slowing in older adults. This model came about due to the slower responses of older than younger adults in various tasks, and it has been an effect found in many different studies, such as in pattern recognition (Salthouse, 2004) and reporting the value of several changing variables (Salthouse et al., 1991). It can even be plotted as a linear function of the responses of younger adults, whereas younger response times increase older response times decrease faster (Fisher & Glaser, 1996). However, cognitive slowing does not complete the whole picture, for as Burke and Mackay (1997) point out, “Why is cognitive performance intact in old age for language comprehension and other semantic tasks, even though older adults perform these tasks more slowly than young adults?” (p. 1851). Several models have been proposed to explain cognition in older adults, such as inhibition deficit theory—the idea that the brain is less able to filter out irrelevant information as we age—and transmission deficit theory—that connections become weak over time, and new connections are harder to form than old ones (Burke & Mackay, 1997).

2.2 Cognitive Characteristics of Alzheimer’s Disease and Dementia

While some cognitive slowing is typical in healthily aging adults, language comprehension is still mostly preserved. However, the effects of this type of slowing become magnified in the case of Alzheimer’s disease. Individuals with Alzheimer’s disease show a much steeper cognitive slowing than healthy older adults when plotted as a linear function, with the decline disproportionately affecting lexical-decision and sentence-completion tasks (Nebes & Madden, 1988).

Alzheimer’s disease is characterized by loss of neurons and dendrites, as well as the presence of Hirano bodies and neurofibrilliary tangles in the brain (Castellani, Rolston, & Smith, 2010). The brains of many healthy older adults also have some of these aspects, especially tangles of neuron filaments (Snowdon et al., 1997). Historically it was thought that prion proteins were responsible for the effects of Alzheimer's disease, but further research has shown that β-amyloid and tau proteins are potential factors (Xiao et al., 2011) and prions may even block these proteins (Hooper & Turner, 2008). Alzheimer's is also considered to be a type of dementia, a syndrome describing a variety of neurological disorders that are all defined as "a decline in 2 or more cognitive capacities, causing impairment in function but not alertness or attention" (Rabins & Blass, 2014, p. ITC2). Of the different types of dementia—which also includes vascular dementia, dementia associated with Parkinson's disease, Lewy body dementia and Frontotemporal dementia—Alzheimer's disease is widely considered to be the most common, comprising 50-70% of all dementia cases (Blennow, De Leon, & Zetterburg, 2006; Castellani et al., 2010).

Alzheimer's disease progresses through several stages. Most cases of dementia begin as Mild Cognitive Impairment, where non-memory related cognitive functions (nonamnestic MCI) and/or memory (amnestic MCI) is impaired significantly but not so much that there is a loss of independence (Petersen, 2011). The amnestic type of MCI is considered to be the forerunner to Alzheimer's disease, while nonamnestic MCI may be linked to frontotemporal dementia (Busse, Hensel, Guhne, Angermeyer, & Riedel-Heller, 2006). During the middle stage, anywhere from 2 to 10 years after the disease has been diagnosed (Cummings, Benson, Hill, & Read, 1985) other deficits increase such as the inability to process sensory information (agnosia), the inability to produce and/or comprehend language (aphasia), and difficulty to perform motor tasks (apraxia). Eventually an Alzheimer's patient will reach a terminal stage anywhere from 8-12 years after diagnosis, where intellectual and physical ability continues to decline until the patient is bedridden.

Since Alzheimer's disease can only be definitively diagnosed after a brain examination at autopsy, clinicians have to rely on cognitive patterns to obtain a presumptive diagnosis. Usually the main symptom that presents itself is a memory problem, but AD can also present itself as a language deficit, a visuospacial disorder, or personality change (Weintraub, Wicklund, & Salmon, 2012). To find these cognitive patterns, many tests have been developed to diagnose potential Alzheimer's, namely the Mini Mental State Examination (MMSE) and the MINI-COG. The MMSE consists of a variety of tasks related to short and long term memory as well as language—examples of tasks include telling the time and date, counting serial 7s, spelling "world" backwards, and writing a sentence, among others—and performance on these tasks is scored out of 30 points (Molloy, Alemayehu, & Roberts, 1991). The MINI-COG is a 3-5 minute examination consisting of 3 items: repeating three unrelated words, drawing a clock face with the hands at 10 and 2, then remembering and repeating the three words said earlier (Doerflinger, 2007).

These cognitive difficulties present themselves not only in tests of memory, but also in the language that Alzheimer's patients use. One well documented language characteristic of Alzheimer's is anomia, an inability to name objects (Hodges, Salmon, & Butters, 1991). Echolalia, or repetition of a previously spoken turn of conversation, is also a common characteristic of dementia. An Alzheimer's adult will repeat the words that were spoken, but not necessarily in the same intonation as that of the other conversant (Cruz, 2010).

Alzheimer's produces grammatical changes as well. While grammatical function mostly remains intact, grammatical complexity does decrease sharply with Alzheimer's. One study, for example, examined two linguistic aspects in healthy aging and Alzheimer's speech—propositional density, or the amount of ideas contained in a sentence, and D-Level, the amount of complex grammatical structures in a sentence. While healthy aging subjects had a slightly declining D-level over time and relatively stable propositional density, Alzheimer's patients had much steeper declines in both (Kemper, Thompson, & Marquis, 2001). When the autobiographical writings of 35 adults with a familial genetic risk of Alzheimer's disease were analyzed, decreased propositional density was significantly correlated with the presence of the apolipoprotein E (APOE) ε4 allele (Medina et al., 2011).

Another study (Kempler, Almor, MacDonald, & Andersen, 1999) demonstrated potential difficulties with grammar and discourse continuation. Two parts of this study were done with healthy aging controls and Alzheimer's patients. First was a test of sentence comprehension, where the subjects were asked to point to the picture that corresponded to the meaning of the sentence. The sentences were simple active, passive, conjoined NP and relative clause. Alzheimer's patients did best on the Active and Conjoined NP sentences, and worst on the passive and relative clause sentences. The second test involved first listening to sentences being spoken, then reading a continuation of the sentence out loud. Half of the continuations were grammatically, semantically, or discourse topic consistent, and half of them were inconsistent. When there is an inconsistent element in a sentence it takes longer to process, and therefore the reader will spend more time reading that segment than a sentence that is consistent. Both Alzheimer's and controls spent a longer time reading out loud the inconsistent continuations. However, Alzheimer's patients had a much shorter duration reading bad discourse continuations, such as "The children loved the silly clown at the party. During the performance, the clown threw candy to *him" (Kempler et al., 1999, p. 235). This is possible evidence that Alzheimer's adults aren't noticing the inconsistency in these types of continuations as much as those without Alzheimer's, and therefore may have greater problems with following discourse topics.

2.3 Current Computational Models for Alzheimer's Disease

Distinct grammatical and syntactical characteristics are part of the course for Alzheimer's, so there is a possibility that these can be detected through natural language analysis. Most of the computational analysis research into characteristics of AD has been conducted by examining written samples of subjects with either presumptive or officially diagnosed Alzheimer's disease. These kinds of analyses, while looking at lexical characteristics such as type/token ratio and the proportion of certain parts of speech, have mostly been focused on measuring the grammatical complexity of Alzheimer's speech using several different techniques. Common measures are propositional density, or the amount of ideas packed into a sentence, and D-level, the amount of embedded clauses in a sentence. Research into Alzheimer's disease has identified two main sources of grammatical difference, namely lower proposition density and Yngve depth. Proposition density (also known as idea density) is how many propositions a sentence contains. For instance, "The old grey mare has a very large nose" contains 5 propositions: "old grey mare" has two (old, grey), "has" is one, and the last two come from "very large nose" (Covington, 2008). Yngve depth, a measure of sentence complexity created by Victor Yngve (1960), calculates the number of nodes in the left hand side of a phrase-structure tree of a sentence. Looking at "The old grey mare has a very large nose" again, the Yngve score would be 3: a noun phrase ("The old grey mare") and two adjectival phrases ("grey mare" and "old grey mare").

The first connection between propositional density and Alzheimer's began with a famous study of 93 nuns from the Milwaukee, Wisconsin area (Snowdon et al., 1996). The nuns wrote a short autobiographical piece about themselves at age 22 when they joined the convent, and were later examined in old age. Researchers found that the nuns who had the richest amount of propositional density in their written pieces were less likely to have Alzheimer's disease in later life than those with lower densities (Snowdon et al., 1996). Other studies have used computational systems such as CPIDR to calculate propositional density, which accomplishes this task by counting the parts of speech associated with higher idea density—verbs, adjectives, adverbs, prepositions, and conjunctions (C. Brown, Snodgrass, Kemper, Herman, & Covington, 2008). While other studies have looked into propositional density and found similar results (Engelman, Agree, Meoni, & Klag, 2010), another study could not find differences in propositional density between the ages (Spencer, 2012). In this study, women who responded to a health survey were examined, and showed no differences in propositional density between younger and older participants. What this study concluded was that the sample sizes were too small to show propositional density effects, which is important to take into account in further research.

Writing samples from published authors and historical figures with suspected or diagnosed Alzheimer's disease have also been computationally analyzed. One study used the letters of King James VI/I between 1604 and 1624, and while his D-level stayed consistent over time his mean length of utterance and mean verb clauses per utterance decreased (Holmes, Williams, Kemper, & Marquis, 2003). The novels of British author Iris Murdoch, who was formally diagnosed with Alzheimer's after her death in 1995, have been examined through a couple of different studies, one of them using the Computerized Language Analysis System (CLAS) based on the Stanford grammar parser (Pakhomov, Chacon, Wicklund, & Gundel, 2011). CLAS uses a combination of Yngve depth, Fraiser scoring—counting the amount of branches up from a word to the highest non-leftmost node in a tree—and sentence dependency scores, which measure "the length of grammatical dependencies between lexical items in the sentence" (Pakhomov et al., 2011, p. 138). The CLAS analysis found a significant decrease in the grammatical complexity of Murdoch's novels over the years. Also alongside this research is another study (Le et al., 2011) that made a comparison between the novels of Iris Murdoch, Agatha Christie—who was suspected of having Alzheimer's disease but never formally diagnosed—and P.D. James, an author with a similar body of work who has aged typically. Using the Natural Language Toolkit to tabulate lexical patterns, the study did find differences in vocabulary usage as well as grammar in Murdoch's and Christie's novels versus those of James. Murdoch and Christie used fewer passive sentences compared to James as they aged, fewer nouns as well as a vocabulary decrease were observed in the works of Murdoch and Christie. They also increased in the use of fillers and repetitions (Le et al., 2011). While these are essentially case studies and sample sizes are too small to make judgments, the results do fit within the general Alzheimer's pattern established earlier.

Some computational analyses have been done on spontaneous speech of Alzheimer's adults as well. Many of these focus on speech produced during a restricted situation, such as describing a story. For example, one examination was done using text samples of speech taken from the Mini Mental State Exam (Roark et al., 2011). The version of the MMSE used in this study contained Wechsler Logical Memory Tests I and II, where participants were asked to repeat a three sentence story after it was spoken (Test I) and then recall the story later after 30 minutes of unrelated activities (Test II). Subjects with MCI and healthy elderly subjects were found to have significant statistical difference in words per clause and Yngve depth. This study also examined the cross-sectional entropy between part of speech tags to find more unusual combinations, and found that this also had a significant statistical difference (Roark et al., 2011).

Other analyses have focused on spontaneous speech in an unrestricted or natural situation. One study analyzed the speech of healthy older adults alongside those with Alzheimer's after being given a prompt that changed every five years which required deep reflection—such as "Whom do you most admire and why?" or "Describe an unexpected event that happened to you" (Kemper et al., 2001, p. 603). While the healthy older adults experienced a decline in grammatical complexity measures such as D-level and propositional density over the years, Alzheimer's adults experienced a sharper decline. An analysis of the Carolina Conversations Collection (Guinn & Habash, 2012) also shows increases in repetitions, incomplete words, filler phrases and forward paraphrasing in Alzheimer's patients compared with their typical interlocutors.

While these analyses have shown lexical and grammatical differences in Alzheimer's versus non-Alzheimer's speech, so far a computational model that reliably distinguishes between the two has not been made. The aforementioned Guinn and Habash (2012) paper made the closest attempt, where they used several machine learning algorithms on the

Carolina Conversations Collection data, such as K-nearest neighbors clustering, a decision tree algorithm, and a support vector machine. The most successful of them, the decision tree, identified Alzheimer's speakers 41% of the time and non-Alzheimer's speakers 100% of the time, bringing its combined accuracy to 79.5%.

2.4 ENTROPY AND COMPUTATIONAL MODELS

While thermodynamic entropy—the systemic disorder seen in physics and chemistry—has been examined in relation to Alzheimer's disease (Drachman, 2006) this analysis focuses on entropy as a concept of information theory, and in particular the entropy rate of speech. Shannon (1951) did make an attempt to calculate the average entropy of English, and a possible upper bound of 1.75 bits has been proposed (Brown et al., 1992). However, others since then have found that entropy varies according to the given text (Debowski, 2011), including aspects such as genre (Kalimeri et al., 2012) and word length (Papadimitriou, Karamanos, Diakonos, Constantoudis, & Papageorgiou, 2010).

While this study uses entropy as a part of the model used to examine text, most applications of entropy focus on its capacity to evaluate a model's effectiveness. The concept of Maximum Entropy Estimation in particular relies on the idea that if two models are identical, then the maximum amount of entropy of each one is measured and the model that produces the least amount of entropy is chosen. Maximum Entropy models have been very important to natural language processing, where they help train classifiers by using the entropy of results in the training data to guide the model's feature selection of testing data (Ratnaparkhi, 1996). These models are often paired with a Hidden Markov Model for speech recognition (Kuo & Gao, 2006). They are also often used in part-of-speech tagging (Ratnaparkhi, 1996), dialogue act recognition (Lan, Ho, Luk, & Leong, 2008), and statistical machine translation (Xiong, Zhang, & Li, 2011).

CHAPTER 3

METHODOLOGY

As we have seen in previous research, both healthily aging adults and those with Alzheimer's go through measurable lexical and grammatical changes over time. The changes in older adults, such as a greater vocabulary and an increased level of accumulated knowledge, would be likely to produce more complex speech. However, having a larger amount of information to remember increases cognitive demands on the ability of older adults to retrieve it, as shown in their higher amounts of tip-of-the-tongue moments. This could reduce the complexity of speech, although the effects of this might be canceled out due to older adults becoming more sensitive to context. Alzheimer's individuals experience steeper cognitive declines, and this is reflected in the reduced complexity in their speech. Since entropy is a measure of the average amount of information in a message, and there is a difference in the amount of information conveyed by a person as they age and whether or not they have Alzheimer's, it's plausible that entropy would be affected by the linguistic changes occurring in aging as well as Alzheimer's.

Our hypothesis is that the perplexity—our measure of entropy—of speech will change as a person ages, and that this change will be different for those with and without Alzheimer's disease. Perplexity reflects the extra amount of bits needed to process a message, which will be dependent upon the model that provides the probabilities to measure that perplexity. Messages that are more like the language model are going to have less perplexity, while messages that are less like the model are going to have higher perplexity. Therefore, the hypothesis is that perplexity will increase over time in typical aging as their speech will increase in vocabulary, but will decrease for an Alzheimer's affected person as speech becomes less complex. The null hypothesis will be that there will be no discernable difference in perplexity between typical aging and Alzheimer's. The second part of the study involves tabulating different lexical features of the speech of each group, then correlating them with perplexity to see which aspects may be the most related.

3.1 Design of the Investigation

To test this hypothesis, there will be three main components. First, there needs to be speech samples that range across the age span of adulthood, as well as samples from Alzheimer's adults. Secondly, we need a way to measure perplexity reliably across these texts, as well as a language model that can be reliably used to generate the perplexity. Lastly, lexical features should be extracted from the texts and then correlated with perplexity values.

Samples of healthy adult speech will be taken from two main corpora: the Buckeye Corpus and the Switchboard Corpus. The Buckeye Corpus, produced in 2005 from Ohio State University, is a collection of speech samples from 40 different speakers, 20 of which are under 30 years old and 20 of which are over 40 years old (Pitt et al., 2007). The participants were recruited through newspaper ads and referrals from other speakers as well as friends and family. These speakers met with a graduate assistant who interviewed them—care was taken that the speakers met with both male and female assistants—and were told that the researchers were examining how people form opinions. The recording focuses entirely on the interviewee's responses, where the interviewer was placed outside of the range of the microphone, and transcribers were told to place any noise made by the interviewer as <IVER> in the transcript. As a result, dialogue bursts in the Buckeye Corpus are longer than in the other corpora, since the interviewee was encouraged to talk at length.

The Switchboard Corpus was created in 1992 by Texas Instruments as a corpus for testing speaker recognition systems (Godfrey, Holliman, & McDaniel, 1992). Switchboard consists of 2,500 telephone conversations from 500 participants ranging from 17 to 68 years old. These participants were randomly placed together for telephone conversations using an automated switchboard system, which then gave them a randomized prompt for discussion. While most participants have around 5 conversations in the system, 50 people who were selected as "target" speakers have about an hour and half of dialogue each. This length was determined to be the minimum amount necessary for training and testing a speaker recognition system to detect these speakers. The version of the Switchboard corpus used for this research is the NXT format Switchboard Corpus (Calhoun et al., 2010) which through the work of Christopher Potts has been collected into CSV files and has had Python scripts written to work through them (Potts, 2011).

Samples of Alzheimer's speech for this study will come from the Carolina Conversations Collection, a corpus created in 2010 by the Medical University of South Carolina (Pope & Davis, 2011). The CCC is a corpus specializing in elderly speech related to health, and contains over 400 participants with a wide variety of medical conditions. Participants were recruited through flyers placed around the community in places such as senior centers, community health centers, churches, and others (Pope & Davis, 2011). Corpus collection was meant to be as representative as possible, and includes speakers from a wider variety of races and backgrounds. This study will be examining 33 out of 34 speakers with Alzheimer’s in the CCC—one subject dropped due to having less than 50 words total in the corpus. These speakers are between the ages of 60 -90, the majority are female—26 females to 7 males—and mostly Non-Hispanic white, including two African-American speakers and one Puerto-Rican speaker. Race data was not included for the other two corpora, so unfortunately a comparison cannot be made.

Perplexity will be measured using the maximum entropy model in the SRI International Language Modeling Toolkit, or SRILM (Alumäe & Kurimo, 2010). This will calculate the average perplexity per word, hence reflecting the perplexity rate over time. Linguistic features will be measured using the Natural Language Toolkit (NLTK) in Python. This software package contains tools for word count, part-of-speech tagging, and sentence parsing (NLTK Project, 2015). Statistical calculations will be made using the programming language R (R Core Team, 2014).

3.2 Data Analysis Procedures

To calculate perplexity over time, first a training corpus will be needed to train the language model, then a testing corpus to obtain the perplexity values over time. The testing corpus must be different from the training corpus, otherwise the perplexity values will be lower than they should be due to the model having calculated the data before. The ideal situation would be to use language models for each age group, but age data isn’t entirely complete for all the corpora since the Buckeye corpus is split simply into “younger” and “older” individuals. Therefore the best approximation possible would be two separate language models: one that represents “younger” speech, consisting of samples from speakers between 20 and 40 years old, and a model that represents “older” speech with speakers between 41 and 61 years old. Since our smallest corpus in this study consists of 33 speakers, each of these language models will consist of the same number of speakers as well to normalize the data. The training corpora data will contain the Buckeye Corpus, since that is not only divided into younger and older speakers, but also has longer bursts of dialogue with which to train a model. However, the Buckeye Corpus only contains 20 younger speakers and 20 older speakers, so to complete the models 11 younger and 11 older speakers will be taken from the Switchboard corpus.

Perplexity will then be obtained from a group of younger to older individuals as well as the Alzheimer’s speakers by first using the younger language model, then using the older language model. For the first group 50 individuals from the Switchboard corpus will be selected from 4 categories, with ten randomly selected from 20-29, 30-39, and 40-49 years old categories and 20 individuals between 50 to 68 years old. This was done because there were only four speakers over 60 available that weren’t already in the training corpus, so more speakers over 50 were selected to compensate. Each category contains half male and half female speakers with five conversations taken from each speaker. The Switchboard corpus contains data on education level and the speaker’s dialect region, but since education data is incomplete for the Alzheimer’s corpus, only age and gender will be considered for grouping. Then, 84 samples from 33 individuals with Alzheimer’s will be examined, with the samples from each subject ranging from 1 to 13. Each subject will have his or her average perplexity per word plotted on a graph to determine the average rate of change over time.

Results from the two groups—Alzheimer’s adults versus non-Alzheimer’s adults—will then be tested for statistical significance using a Mixed-Effects model from the statistical package lme4 in R (Bates et al., 2015) and tested for significance using the package lmerTest (Kuznetsova, Brockhoff, & Christensen, 2015). The fixed effects are the variables that have all of their possibilities represented, and those are age, sex, Alzheimer’s status, and the lexical characteristics being examined: type/token ratio, fillers, repetitions, low specificity words, mean length of utterance, and proportions of different parts of speech. The random effects are the variables for which all the possibilities have not been exhausted and have unknown effects on the data, of which there are two main ones: the individuals themselves and the corpora they come from. The individuals are a random effect because not only do we have incomplete data—such as education, region and exact age for all the individuals—there could be other individual factors that can't be accounted for, such as a person's attitude that day. Corpus was chosen as the other random factor because of the major differences between the Switchboard and CCC in how the data was collected. While the Buckeye Corpus and the CCC contain individuals mostly within a specific region, the Switchboard Corpus is essentially a random sample of the population, containing different education levels and dialect regions. In the Carolina Conversation Collection dialogues, some of the conversation partners were taught strategies to help the Alzheimer's individual in participation, such as prompts, backchannels and use of open ended questions (Pope & Davis, 2011). The topics are meant to be relevant to the individuals with Alzheimer's, such as events in their past, their health, and what was happening to them that day. The Switchboard speakers, however, were conversing with a randomly chosen stranger on the phone without any particular direction other than the randomly chosen prompt. Since the effects of all these factors cannot be controlled for, it would be best to account for these as a random factor for the corpus. A mixed-effects model can better account for these random factors (Baayen, 2008) test to see if individual differences between the samples are significant.

After perplexity values are obtained, the next part of the experiment will be an exploratory study to find out if any linguistic factors are correlated with the perplexity values. Previous research has identified lexical and grammatical features with significant differences between Alzheimer's speakers and other adults. Alzheimer's speakers have been shown to use fewer types of words and have shorter utterances with more repetitions and fillers (Guinn & Habash, 2012). They also tend to use more open-class words like nouns and verbs, and they use less specific words more often (Le et al., 2011). Therefore, this study will look at six lexical features—type/token ratio, mean length of utterance, repetitions, fillers, lexical specificity, and word class proportions.

Type/token ratio will be counted by dividing the unique words in the sample by the total words. Fillers are counted by how many instances of a set of filler words appear. Repetitions will be counted by counting how many sets of words—from two to six words—repeat in the whole sample and will also include local repetitions, or two of the same words right next to each other (such as "And… and…"). Borrowing from Le et al. (2011), the measure of lexical specificity will look at 39 indefinite nouns and verbs considered to be low specificity:

> anything, ask, be, come, do, feel, find, forget, get, give, go, have, happen, hear, know, like, live, look, make, mean, meet, nothing, put, remember, run, say, see, seem, speak, something, suppose, take, tell, thing, think, use, walk, want, wonder. (p. 444)

To tabulate word class proportions, nine lexical categories will be calculated—nouns, pronouns, verbs, adjectives, adverbs, prepositions, lexical “to,” determiners, and conjunctions. This analysis will only include the samples that have Part of Speech (POS) already in the corpus, which are all 250 Switchboard samples and 79 of the Alzheimer’s speech. The POS features will then be tabulated using the frequency distribution function in NLTK and correlated with perplexity values in a mixed methods model in R.

CHAPTER 4

RESULTS AND DISCUSSION

In general, the Alzheimer's subjects had a higher mean perplexity than the non-Alzheimer's subjects, and perplexity generally increased upward with age. This effect held for both the younger and the older language model. However, the Alzheimer's subjects were also the elderly subjects, and the older language model produced lower perplexities overall for them, indicating that an older language model was a better fit for their speech.

4.1 PRESENTATION OF THE FINDINGS

4.1.1 Overall Composition

Blooking at the mixed model representation of the perplexity values it would be helpful to get a broad, general overview of the linguistic composition of the subjects with and without Alzheimer's disease. Since the Alzheimer's speech samples were recorded in a more naturalistic setting their samples varied more in length—ranging from 68 to 5,302 words—than those of the Switchboard corpus samples – 109 to 2217 words. Therefore, all lexical information has been normalized to show the amount per 100 words, with the exceptions of type/token ratio, unique words, and mean length of utterance since these are measured with respect to the whole text. These are the raw means and standard deviations for the normalized counts of the two groups, as well as a t-test to determine if these differences are significant. (Table 4.1)

Most of the differences between the means for Alzheimer's and non-Alzheimer's subjects were considered significant, with the exception of unique words, five word repetitions, six word repetitions and verbs. Similar to results in previous studies, the Alzheimer's subjects had an overall shorter mean length of utterance—7.57 words—as opposed to the non-Alzheimer's group—9.30 words.Type/Token ratio was also lower for the Alzheimer's subjects—0.06—than the non-Alzheimer's speech—0.31. Alzheimer's subjects

Table 4.1. Descriptive Statistics in Linguistic Composition (Normalized) Between Subjects With and Without Alzheimer's Disease

Linguistic Feature	Non-Alzheimer's Mean (SD)	Alzheimer's Mean (SD)	T-test	P-Value
Unique Words	250.60(75.49)	285.99(155.08)	-0.817	0.4158
Type/Token Ratio	0.31(0.32)	0.06(0.10)	5.9544	1.65E-08***
Mean Length of Utterance	9.30(2.50)	7.57(2.21)	-2.142	0.03476**
Low Specificity Words	6.55(1.60)	7.44(1.59)	-4.4446	1.77E-05***
Fillers	2.99(1.23)	3.61(2.56)	-2.142	0.03476**
Local Repetitions	1.09(0.95)	0.19(0.26)	13.497	2.2E-16***
Bigram Repetitions	23.94(6.05)	20.46(7.93)	3.6624	0.0003787***
Trigram Repetitions	7.70(3.32)	5.79(3.39)	4.471	1.61E-05***
Four Word Repetitions	2.59(1.61)	1.79(1.51)	4.1251	6.13E-05***
Five Word Repetitions	0.71(0.67)	0.64(0.70)	0.8516	0.396
Six Word Repetitions	0.22(0.32)	0.20(0.33)	0.6149	0.5396
Nouns	10.87(2.25)	13.94(3.26)	-7.7142	8.888E-12***
Pronouns	11.42(2.29)	15.25(3.14)	-9.987	2.2E-16***
Verbs	14.08(2.53)	14.47(3.21)	-1.0002	0.3195
Adjectives	3.74(1.07)	3.34(1.07)	2.856	0.005***
Adverbs	6.95(1.74)	8.26(2.46)	-4.3833	2.841E-05***
Lexical To	1.17(6.34)	1.77(0.94)	-5.385	5.309E-07***
Prepositions	6.81(1.67)	4.28(1.40)	13.5268	2.20E-16***
Determiners	6.42(1.32)	1.81(0.98)	33.2205	2.2E-16***
Conjunctions	3.67(1.37)	6.17(2.35)	-8.9444	3.26E-14***

Note: ** = $p < 0.05$, *** = $p < 0.005$

had a higher amount of low specificity words and fillers per 100 words (Table 4.1) than non-Alzheimer's subjects. Surprisingly, the non-Alzheimer's subjects had more repetitions per 100 words all across the spectrum, although only local repetitions to four word repetitions were statistically significant. As for the part-of-speech data, Alzheimer's subjects had more nouns, pronouns, adverbs and conjunctions, but a roughly equal amount of verbs and less prepositions, determiners and adjectives.

4.1.2 Perplexity and Intercepts

In the previous section, we've identified all but five criteria to have statistically significant differences between those with and without Alzheimer's. Therefore, in this next section, as we look at the effect on perplexity on older, younger and Alzheimer's speech, we will also look at the effects of perplexity on those statistically significant factors.

Overall, the Alzheimer's speech had a higher perplexity than the non-Alzheimer's speech, and this effect carried when both the younger and the older language model are used. The mean for the Non-Alzheimer's perplexity was 161.29 for the younger model and 157.94 for the older model, while the Alzheimer's perplexities were 207.29 for the younger model and 165.54 for the older model (Table 4.2).

Table 4.2. Mean and Standard Deviation for Perplexity

	Non-Alzheimer's Mean	Non-Alzheimer's SD	Alzheimer's Mean	Alzheimer's SD
Perplexity Younger Model	161.29	41.51	207.29	73.53
Perplexity Older Model	157.94	38.55	165.54	46.76

For both the younger and the older language model, perplexity generally increases over time. When the younger language model was used, perplexity went up at a steeper rate (Figure 4.1) than when the older language model was calculated (Figure 4.2). However, the Alzheimer's subjects have a wider range of perplexity values, as can be seen in their larger standard deviations.

Afterward, a mixed effects model was applied in R with perplexity as the dependent variable, the statistically significant factors from Table 4.1 as covariates, and the random

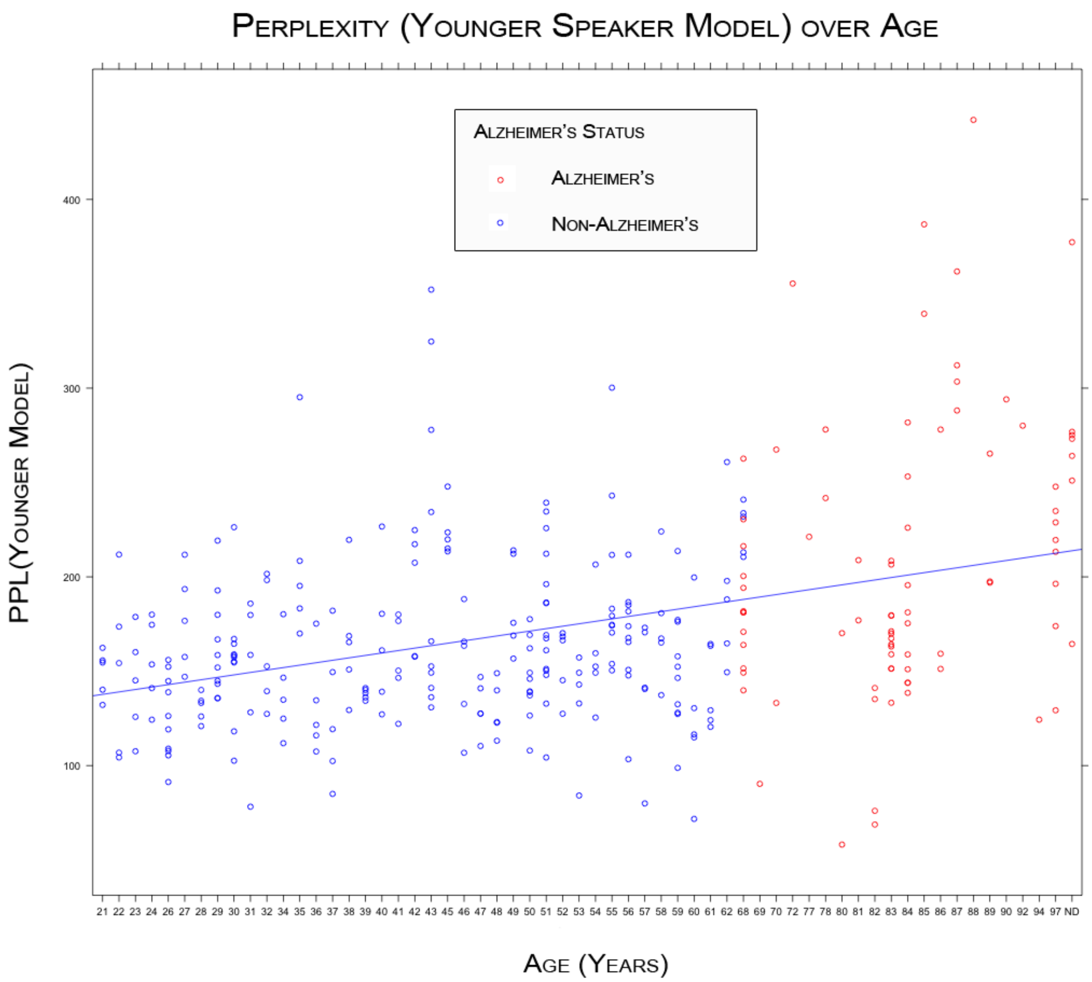

Figure 4.1. Perplexity over age plot using younger language model.

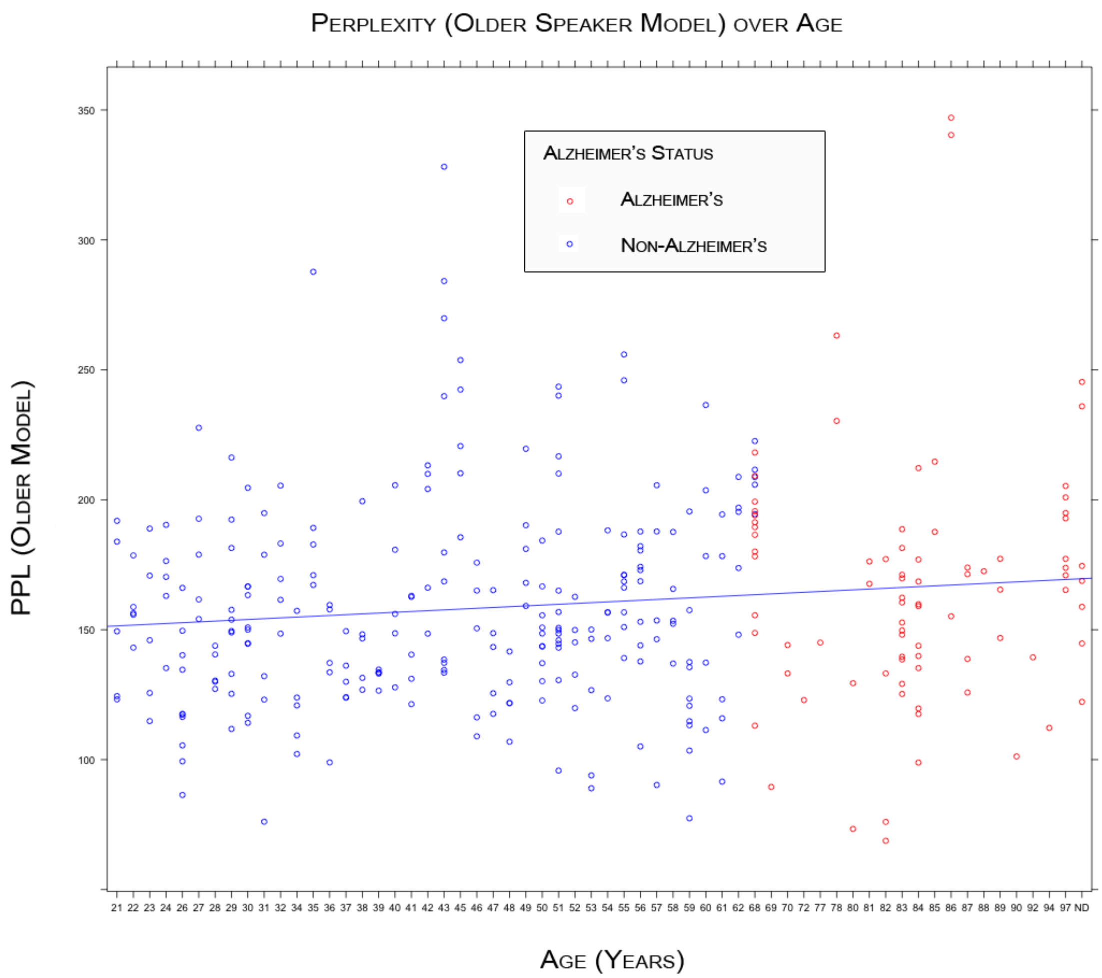

Figure 4.2. Perplexity over age plot using older language model.

effects of speaker and corpus. This was accomplished through the following command for the younger model perplexities:

```
     > library(lmeTest)
> YoungerModelPPL.lmer = lmer(PPL ~ 1 + Age + Sex +
Alz_Status + Type.Token.Ratio + Mean.Length +
LowSpecNorm + FillerNorm + LocalRepNorm + BigramNorm
+ TrigramNorm + FourNorm + NounNorm + ProNorm +
AdjNorm + AdNorm + ToNorm + PrepNorm + DetNorm +
ConjNorm + (1|Speaker) + (1|Corpus), data =
SeparatePPL)
```

The mixed methods analysis for the older model was accomplished with the following changes to the code:

```
> OlderModelPPL.lmer = lmer(PPL.Older.Model ~ 1 +
Age + Sex + Alz_Status + Type.Token.Ratio +
Mean.Length + LowSpecNorm + FillerNorm +
LocalRepNorm + BigramNorm + TrigramNorm + FourNorm +
NounNorm + ProNorm + AdjNorm + AdNorm + ToNorm +
PrepNorm + DetNorm + ConjNorm + (1|Speaker) +
(1|Corpus), data = SeparatePPL)
```

The random effects of speaker and corpus were both higher on the younger model than the older model, and they varied more widely as well. The residual, which represents the effects that couldn’t fit into the model, was also higher for the younger language model. (Table 4.3).

Table 4.3. Variances and Standard Deviation for Random Effects on Younger and Older Language Model Perplexities

Random Effects	Younger Model Variance	Younger Model SD	Older Model Variance	Older Model SD
Speaker	1513.6	38.90	412.1	20.30
Corpus	988.6	31.44	842.1	29.02
Residual	916.4	30.27	740.0	27.20

The mixed methods model then uses these variances to create intercepts for each characteristic that are adjusted for the effects of speaker and corpus. The package lmeTest generates significance for each of these values based on a T-test. While the age brackets had the highest overall intercepts and all their values were positive—since perplexity increases with age—none of the differences were statistically significant. The aspects that were statistically significant were male sex, type token ratio, mean length of utterance, fillers and local repetitions (Table 4.4).

The mixed methods intercepts for the older language model were overall lower than those for the younger language model, with increases in intercepts for local repetitions, nouns and adjectives. The statistically significant differences were in mean length of utterance, low specificity words, nouns and adverbs. Some of the intercepts went from positive to negative, such as for ages 30-39 and 50-68, as well as in type token ratio and fillers. However, these decreases were not statistically significant (Table 4.5).

The two characteristic differences with statistical significance in both models were nouns and mean length of utterance. The difference in low specificity words was significant only for the older language model, while male sex, type token ratio, fillers, and local repetitions were only statistically significant for the younger model perplexities. However, neither model had statistically significant differences in terms of age, and the older model was a better fit for the data, as indicated by the lower perplexities.

such as for ages 30-39 and 50-68, as well as in type token ratio and fillers. However, these decreases were not statistically significant (Table 4.5).

The two characteristic differences with statistical significance in both models were nouns and mean length of utterance. The difference in low specificity words was significant only for the older language model, while male sex, type token ratio, fillers, and local repetitions were only statistically significant for the younger model perplexities. However, neither model had statistically significant differences in terms of age, and the older model was a better fit for the data, as indicated by the lower perplexities.

Notably absent from both models is a direct measure of Alzheimer’s status, and that is due to the model dropping the data for being rank deficient for that particular fixed effect. An appearance of this warning is usually due to multicollinearity in the data (user974, 2012). Alzheimer’s status in our study is heavily correlated with age, so the model didn’t have sufficient data to measure directly the effect of Alzheimer’s on perplexity. Therefore, while there does seem to be a general upward trend in perplexity as one ages, it is unclear whether Alzheimer’s individual perplexities are due to upward age or advancing disease.

Table 4.4. Intercepts and Significance for Younger Language Model Perplexities

Fixed Effects	Intercept Estimates	T-Value	Significance (P-Value)
Age 20-29	131.0785	2.432	0.996166
Age 30-39	0.214	0.012	0.99085
Age 40-49	18.9272	1.001	0.322797
Age 50-68	8.5472	0.527	0.600812
Age 60 (Alzheimer's)	80.484	1.377	0.998683
Age 70 (Alzheimer's)	24.5329	0.456	0.999227
Age 80 (Alzheimer's)	44.9302	0.914	0.999348
Age 90 (Alzheimer's)	91.3398	1.651	0.998883
Age ND (Alzheimer's)	94.5254	1.802	0.999064
Sex (Male)	21.9873	2.038	0.047237 **
Type Token Ratio	128.6468	-1.864	0.063422 **
Mean Length of Utterance	7.8586	5.971	7.1e-09 ***
Low Specificity Words	-1.6259	-1.102	0.271761
Fillers	4.0552	2.261	0.024481 **
Local Repetition	-6.898	-2.869	0.004499 **
Bigram Repetition	-0.5292	-0.452	0.651663
Trigram Repetition	-4.3488	-1.34	0.18163
Four Word Repetition	3.5322	0.745	0.45701
Nouns	4.1008	3.552	0.000449 ***
Pronouns	0.9523	0.931	0.352898
Adjectives	0.8352	0.374	0.708457
Adverbs	-1.5097	-1.171	0.242723
Lexical To	-3.2799	-0.941	0.347511
Prepositions	-2.3123	-1.406	0.161204
Determiner	-0.5964	-0.291	0.771479
Conjunctives	-0.8039	-0.521	0.602568

Note: ** = $p < 0.05$, *** = $p < 0.005$

Table 4.5. Intercepts and Significance for Older Language Model Perplexities

Fixed Effects	Intercept Estimates	T-Value	Significance (P-Value)
Age 20-29	169.1097	3.609	0.99389
Age 30-39	-6.1807	-0.576	0.56695
Age 40-49	3.632	0.326	0.74575
Age 50-68	-4.0863	-0.431	0.66812
Age 60 (Alzheimer's)	19.3187	0.405	0.99882
Age 70 (Alzheimer's)	24.0369	0.523	0.9989
Age 80 (Alzheimer's)	12.6616	0.291	0.99922
Age 90 (Alzheimer's)	10.5419	0.227	0.99909
Age ND (Alzheimer's)	22.3045	0.494	0.99898
Sex (Male)	0.6536	0.1	0.92064
Type Token Ratio	-83.9596	-1.426	0.15488
Mean Length of Utterance	6.9032	6.291	1.11e-09 ***
Low Specificity Words	-2.3946	-1.875	0.06184 **
Fillers	-0.622	-0.44	0.66021
Local Repetition	1.0751	0.513	0.60865
Bigram Repetition	-0.8411	-0.825	0.41001
Trigram Repetition	-2.534	-0.907	0.36502
Four Word Repetition	4.0508	0.994	0.32116
Nouns	2.5947	2.675	0.00788 **
Pronouns	-0.3957	-0.457	0.64782
Adjectives	1.6107	0.833	0.40544
Adverbs	-3.1458	-2.881	0.00425 **
Lexical To	-1.8486	-0.613	0.54066
Prepositions	-2.0265	-1.404	0.16135
Determiner	1.5208	0.863	0.38907
Conjunctives	-1.097	-0.835	0.40412

Note: ** = $p < 0.05$, *** = $p < 0.005$

4.2 DISCUSSION OF THE FINDINGS

In terms of overall statistics, the findings that were consistent with previous research are that the Alzheimer's group had a lower type/token ratio, lower mean length of utterance, higher use of low specificity words, and higher use of fillers. However, they had less amounts of repetitions than the non-Alzheimer's subjects, and contrary to previous research, actually use more nouns and pronouns along with a slightly higher amount of verbs than the non-Alzheimer's group. They also had higher amounts of adverbs and conjunctions, but lower amounts of determiners and prepositions.

Perplexity is dependent upon the language model used to generate the probability of the combination of words within a text—in our case, the model used was dependent on bigram (two word) sequences. If a text is less similar than that of the model used to generate the perplexity probabilities, it will have a lower probability and a higher perplexity than texts that are more similar to the model. As seen before in previous research, older adults tend to have a greater vocabulary than younger adults. This property may explain why the older language model yielded overall lower perplexities than the younger language model, and why perplexity seemed to increase with age. Since our language models mainly consisted of dialogue from the Buckeye Corpus, where participants were encouraged to talk at length, the older speech may have had more content than the younger speech. Since the text from older adults had more dense content, it was a better "fit" for the text and therefore produced lower perplexities overall.

However, the results obtained from this research consistently showed a higher perplexity for the Alzheimer's speech, and this effect remained whether the younger or the older language model was used. Alzheimer's speech should be less complex than non-Alzheimer's, since it had lower mean length of utterance, lower type/token ratio, more fillers, and a greater use of low specificity words. Why would the Alzheimer's speech have higher perplexity?

One potential factor could be that, in general, the Alzheimer's speech sampled had more unique words than the non-Alzheimer's speech. While the difference in unique words between those with and without Alzheimer's was not considered statistically significant and therefore not included in the mixed methods model, there was a statistically significant difference in the use of nouns between the groups, and the intercepts produced by the mixed effects model were also statistically significant. The Alzheimer's speakers used more nouns, which are the most likely part of speech to contain unique words, lending support to this idea. In addition, many of the Alzheimer's dialogues centered around health, which may have included more unique words and bigrams than those in the Switchboard and Buckeye corpora, resulting in a higher perplexity.

What may possibly explain these results is differing grammar patterns in the Alzheimer's speech. While the words used in the dialogue may be common, differing grammar patterns can produce a higher perplexity since the combination of words will have a lower probability. As an example, compare these two segments from the Switchboard conversations and from the CCC Alzheimer's dialogues, along with the perplexity at each line. Taken from one side of the conversation, this segment is from ten lines in the Switchboard corpus spoken by a 59-year-old woman in a conversation about being raised on a farm (Table 4.6).

Table 4.6. Perplexity by Line for Subject Without Alzheimer's

Speech Line	Younger Model Perplexity	Older Model Perplexity
Um, I think life, uh, -	187.646	80.4003
now I grew up on a farm,	104.676	20.1343
I don't, -	3.23463	2.45373
what kind –	179.393	102.817
did you grow up in the farm.	289.646	333.209
Uh-huh.	16.8581	16.0464
I think though it was much slower and, uh, much more self-contained.	109.984	328.429
Uh-huh.	16.8581	16.0464
Yes.	Undefined	Undefined
Unreal, unreal by comparison.	Undefined	Undefined

The segments with the highest perplexities here are not only the ones that are more grammatically complex (such as "I think though it was much slower and, uh, much more self-contained") but also the ones that are more surprising grammatically. The contrast between "did you grow up in the farm" and "now I grew up on a farm" is that "in the farm" is statistically less likely than "on a farm," since native English speakers are much more likely to say “on a farm.” Therefore the former line has a much higher perplexity in both models (289.65 and 333.20) than the latter (104.68 and 20.13).

This result also appeared in this 10-line segment from a woman in her 70s with Alzheimer’s in the Carolina Conversations Collection. She also is in a conversation about life on the farm (Table 4.7). Both the samples contain speech fragments, and since these fragments are relatively common they have a lower perplexity. Despite having more common words, the Alzheimer’s speech sample has a higher overall perplexity because it contains some surprising grammar patterns. In particular, the last three lines have a much higher perplexity because they have an inversion (“fields, some where they uh”) and a missing verb (“but other animals”).

Table 4.7. Perplexity by Line for Subject With Alzheimer's

Speech Line	Younger Model Perplexity	Older Model Perplexity
I don't think not really.	61.7941	80.1575
Uh huh.	1.28	1.90006
Cotton	Undefined	Undefined
I Cotton?	Undefined	Undefined
Cotton and	10.4819	15.6391
horses and all kinds of stuff.	37.1172	67.3088
we, we just had to do it.	32.0282	43.4843
fields, some where they uh,	111.956	137.598
but they uh,	68.4466	73.9913
but other animals.	390.275	769.404

CHAPTER 5

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

5.1 SUMMARY

This corpus study examined the differences in entropy between Alzheimer's and non-Alzheimer's subjects using the measure of perplexity produced by the SRI Language Modeling Toolkit (SRILM). Entropy has been used in previous research to test how well a model fits a particular set of data. The language models used by the toolkit were comprised of speech from the Buckeye and Switchboard corpora, with one model consisting of speech from individuals under 40 years of age—the younger speech model—and over 40 years of age—the older speech model. The speech of non-Alzheimer's individuals tested came from the Switchboard corpus, and the Alzheimer's individuals came from the Carolina Conversations Collection. The hypothesis was that age would have an effect on entropy, where it would raise as healthy subjects grew older but lower for Alzheimer's patients.

5.2 CONCLUSIONS

Overall, perplexity—and by extension, entropy—does appear to increase over age, and this effect remained in both the younger and older language model. The perplexities were lower overall for the older language model than the younger model, indicating that the older model is a better fit for the text and had less of a difference. This effect held between the Alzheimer's and the non-Alzheimer's, indicating that it was a better fit for the Alzheimer's text as well. These results point to the conclusion that Alzheimer's speech is closer to older person speech than younger person speech, meaning that Alzheimer's status isn't as much of an effect on entropy.

However, this study is limited by the fact that data is missing—namely, non-Alzheimer's individuals that are 70 and over. This affected the mixed effects model the most, since it couldn't measure the effect of Alzheimer's directly due to insufficient data. The Switchboard corpus data on hand had the oldest individual at 64 years old, and only 4 speakers were over 60. To make the timeline results more even, more individuals over 50 were added. However, this doesn't entirely make up for the missing data, since other studies have shown possible differences in the speech of adults over 70 that may influence the results. Another aspect that could not be examined was the effect of education level, since over half the Alzheimer's data did not include this information. Education level has shown to be influential in the course of Alzheimer's disease, and could potentially affect some of the more unusual grammar patterns found.

5.3 Recommendations

In addition to adding more individuals over 70 years old and an examination of education level, it would also be useful to replicate this experiment in different environments. For instance the perplexity of written as opposed to spoken samples could be examined to see if the effect still occurs. Yngve depth could be another aspect correlated with these results, especially since mean length of utterance has been shown to have an effect. The perplexity of speech produced by the MMSE exam or a picture description would be particularly edifying since we can control for context and education level in both of these situations.

Finally, it would be interesting to incorporate these results, flawed as they are, into a computational system for distinguishing between Alzheimer's and non-Alzheimer's speech. Since the younger model perplexities produced the most difference between the two groups, it may potentially be useful to see if that language model, along with other analyses, could help a system distinguish between those with Mild Cognitive Impairment, those with Alzheimer's, and those without.

REFERENCES

Aktaruzzaman, M., & Sassi, R. (2014). Parametric estimation of sample entropy in heart rate variability analysis. *Biomedical Signal Processing and Control, 14*, 141-147.

Alumäe, T., & Kurimo, M. (2010). *Efficient estimation of maximum entropy language models with N-gram features: An SRILM extension.* Paper presented at Interspeech 2010, Chiba, Japan.

Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge, England: Cambridge University Press.

Balakrishnan, K., & Touba, N. (2007). Relationship between entropy and test data compression. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 26*(2), 386-395.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H., Singmann, H., . . . Grothendieck, G. (2015). lme4: Linear Mixed-Effects Models using "Eigen" and S4 (Version 1.1-8) [Computer software]. Retrieved from https://cran.r-project .org/web/packages/lme4/index.html

Blennow, K., De Leon, M. J., & Zetterburg, H. (2006). Alzheimer's disease. *The Lancet*, *368*(9533), 387-403.

Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., & Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods*, *40*(2), 540-545.

Brown, P. F., Della Pietra, V. J., Mercer, R. L., Della Pietra, S. A., & Lai, J. C. (1992). An estimate of an upper bound for the entropy of English. *Computational Linguistics*, *18*(1), 31-40.

Burke, D. M., & Mackay, D. (1997). Memory, language, and ageing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *352*(1363), 1845-1856.

Burke, D. M., & Shafto, M. (2004). Aging and language production. *Current Directions in Psychological Science*, *13*(1), 21-24.

Busse, A., Hensel, A., Guhne, U., Angermeyer, M., & Riedel-Heller, S. (2006). Mild cognitive impairment—Long-term course of four clinical subtypes. *Neurology*, *67*(12), 2176-2185.

Calhoun, S., Carletta, J., Brenier, J., Mayo, N., Jurafsky, D., Steedman, M., & Beaver, D. (2010). The NXT-format switchboard corpus: A rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language Resources and Evaluation*, *44*(4), 387-419.

Castellani, R. J., Rolston, R., & Smith, M. (2010). Alzheimer's disease. *Disease-a-Month*, *56*(9), 484-546.

Covington, M. (2008). *Idea density—A potentially informative characteristic of retrieved documents* [PowerPoint slides]. Retrieved from http://ai1.ai.uga.edu/caspr/Covington-2009-Idea-Density-slides.pdf

Cruz, F. (2010). Verbal repetitions and echolalia in Alzheimer's discourse. *Clinical Linguistics & Phonetics*, *24*(11), 848-858.

Cummings, J. L., Benson, F., Hill, M. A., & Read, S. (1985). Aphasia in dementia of the Alzheimer type. *Neurology*, *35*(3), 394-397.

Debowski, L. (2011). Excess entropy in natural language: Present state and perspectives. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *21*(3), 037105-037105-11.

Doerflinger, D. (2007). The mini-cog. *The American Journal of Nursing*, *107*(12), 62-71.

Drachman, D. (2006). Aging of the brain, entropy, and Alzheimer disease. *Neurology*, *35*(3), 394-397.

Engelman, M., Agree, E., Meoni, L., & Klag, M. (2010). Propositional density and cognitive function in later life: Findings from the precursors study. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *65B*(6), 706-711.

Fisher, D., & Glaser, R. (1996). Molar and latent models of cognitive slowing: Implications for aging, dementia, depression, development, and intelligence. *Psychonomic Bulletin & Review*, *3*(4), 458-480.

Giffard, B., Desgranges, B., Kerrouche, N., Piolino, P., & Eustache, F. (2003). The hyperpriming phenomenon in normal aging: A consequence of cognitive slowing? *Neuropsychology*, *17*(4), 594-601.

Godfrey, J., Holliman, E., & McDaniel, J. (1992). Switchboard: Telephone speech corpus for research and development. *ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, *1*, 517-520.

Guinn, C., & Habash, A. (2012). *Language analysis of speakers with dementia of Alzheimer's type*. (AAAI Technical Report FS-12-01). Palo Alto, CA: Artificial Intelligence for Gerontechnology.

Hodges, J., Salmon, D., & Butters, N. (1991). The nature of the naming deficit in Alzheimers and Huntingtons-disease. *Brain*, *114*(4), 1547-1558.

Holmes, F., Williams, K., Kemper, S., & Marquis, J. (2003). Written language clues to cognitive changes of aging. *Journals of Gerontology: Series B*, *58*(1), 42-44.

Hooper, N., & Turner, A. (2008). A new take on prions: Preventing Alzheimer's disease. *Trends in Biochemical Sciences*, *33*(4), 151-155.

Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition.* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Kalimeri, M., Constantoudis, V., Papadimitriou, C., Karamanos, K., Diakonos, F., & Papageorgiou, H. (2012). Entropy analysis of word-length series of natural language texts: Effects of text language and genre. *International Journal of Bifurcation and Chaos*, *22*(9), 1250223-1-1250223-8.

Kave, G., & Halamish, V. (2015). Doubly blessed: Older adults know more vocabulary and know better what they know. *Psychology and Aging*, *30*(1), 1-6.

Kave, G., & Yafe, R. (2014). Performance of younger and older adults on tests of word knowledge and word retrieval: Independence or interdependence of skills? *American Journal of Speech-language Pathology*, *23*, 36-45.

Kellas, G., Paul, S., & Vu, H. (1995). Aging and language performance: From isolated words to multiple sentence contexts. In P. A. Allen & T. R. Bashore (Eds.), *Age differences in word and language processing* (pp. 87-109). Amsterdam, Holland: Elsevier.

Kemper, S. J., Thompson, M., & Marquis, J. (2001). Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content. *Psychology and Aging*, *14*(3), 873-881.

Kempler, D., Almor, A., MacDonald, M., & Andersen, E. (1999). Working with limited memory: Sentence comprehension in Alzheimer's disease. In S. Kemper & R. Kliegl (Eds.), *Constraints on language: Aging, grammar and memory* (pp. 227-247). New York, NY: Kluwer Academic Publishers.

Kuo, H., & Gao, Y. (2006). Maximum entropy direct models for speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, *14*(3), 873-881.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2015). lmerTest: Tests in linear mixed effects models (Version 2.0-25) [Computer software]. Retrieved from http://CRAN.R-project.org/package=lmerTest

Lan, K., Ho, K., Luk, R. W., & Leong, H. (2008). Dialogue act recognition using maximum entropy. *Journal of the American Society for Information Science and Technology*, *59*(6), 859-874.

Le, X., Lancashire, I., Hirst, G., & Jokel, R. (2011). Longitudinal detection of dementia through lexical and syntactic changes in writing: A case study of three British novelists. *Literary and Linguistic Computing*, *26*(4), 435-461.

Long, L. L., & Shaw, R. J. (2000). Adult age differences in vocabulary acquisition. *Educational Gerontology*, *26*(7), 651-664.

Marini, A., Boewe, A., Caltagirone, C., & Carlomagno, S. (2005). Age-related differences in the production of textual descriptions. *Journal of Psycholinguistic Research*, *34*(5), 439-463.

Medina, L., Rodriguez-Agudelo, Y., Greschwind, D., Gilbert, P., Liang, L., Cummings, J. L., & Ringman, J. M. (2011). Propositional density and apolipoprotein e genotype among persons at risk for familial Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, *148*(1), 102-105.

Molloy, D., Alemayehu, E., & Roberts, R. (1991). Reliability of a standardized mini-mental-state-examination compared with the traditional mini-mental-state-examination. *American Journal of Psychiatry*, *148*(1), 102-104.

Nebes, R., & Madden, D. (1988). Different patterns of cognitive slowing produced by Alzheimer's disease and normal aging. *Psychology and Aging*, *3*(1), 102-104.

Natural Language Toolkit [NLTK] Project. (2015). Natural Language Toolkit (Version 3.0) [Computer software]. Retrieved from http://www.nltk.org/

Pakhomov, S., Chacon, D., Wicklund, M., & Gundel, J. (2011). Computerized assessment of syntactic complexity in Alzheimer's disease: A case study of Iris Murdoch's writing. *Behavior Research Methods*, *43*(1), 136-144.

Papadimitriou, C., Karamanos, K., Diakonos, F., Constantoudis, V., & Papageorgiou, H. (2010). Entropy analysis of natural language written texts. *Physica A: Statistical Mechanics and Its Applications*, *389*(16), 3260-3266.

Park, D. C., & Schwarz, N. (2000). *Cognitive aging: A primer*. Philedelphia, PA: Psychology Press.

Petersen, R. (2011). Mild cognitive impairment. *New England Journal of Medicine*, *364*(23), 2227-2234.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye corpus of conversational Speech* [Data file]. Retrieved from http://www.buckeyecorpus.osu.edu

Pope, C., & Davis, B. (2011). Finding a balance: The Carolinas Conversation Collection. *Corpus Linguistics and Linguistic Theory*, *7*(1), 143-161.

Potts, C. (2011). *The Switchboard Dialog Act corpus*. Retrieved from http://compprag .christopherpotts.net/swda.html

Rabins, P. V., & Blass, D. M. (2014). In the clinic dementia. *Annals of Internal Medicine*, *161*(3), ITC1-ITC16.

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of life-long learning. *Topics in Cognitive Science*, *6*(1), 5-42.

Ratnaparkhi, A. (1996). *A maximum entropy model for part-of-speech tagging*. Paper presented at the Conference on Empirical Methods in Natural Language Processing, Philadelphia, PA.

R Core Team. (2014). R: A language and environment for statistical computing (Version 3.1.3) [Computer software]. Retrieved from https: www.r-project.org/

Roark, B., Mitchell, M., Hosom, J., Hollingshead, K., & Kaye, J. (2011). Spoken language derived measures for detecting mild cognitive impairment. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(7), 2081-2090.

Salthouse, T. A. (2004). What and when of cognitive aging. *Current Directions in Psychological Science*, *13*(4), 140-144.

Salthouse, T. A., Babcock, R. L., & Shaw, R. J. (1991). Effects of adult age on structural and operational capacities in working memory. *Psychology and Aging, 6*(1), 118-127.

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379-423.

Shannon, C. (1951). Prediction and entropy of printed English. *The Bell System Technical Journal*, *30*, 50-64.

Sinatra, R., & Szell, M. (2014). Entropy and the predictability of online life. *Entropy*, *16*(1), 543-556.

Snowdon, D., Greiner, L., Wekstein, D., Kemper, S., Mortimer, J., & Markesbery, W. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings from the Nun Study. *Journal of the American Medical Association*, *275*(7), 528-532.

Snowdon, D., Riley, K., Greiner, L., Griener, P., Mortimer, J., & Markesbury, W. (1997). Brain infarction and the clinical expression of Alzheimer disease: The Nun Study. *Journal of the American Medical Association*, *277*(10), 813-817.

Spencer, E. (2012). Language and aging—Exploring propositional density in written language—Stability over time. *Clinical Linguistics and Phonetics*, *26*(9), 743-754.

Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging*, *17*(3), 453-467.

user974. (2012, August 25). Re: What is rank deficiency, and how to deal with it? [Online forum comment]. Retrieved from http://stats.stackexchange.com/questions /35071/what-is-rank-deficiency-and-how-to-deal-with-it

Weintraub, S., Wicklund, A., & Salmon, D. (2012). The neuropsychological profile of Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*, *2*(4), a006171-a006117.

Wingfield, A., & Tun, P. A. (1999). Working memory and spoken language comprehension: The case for age stability in conceptual short-term memory. In S. Kemper & R. Kliegl (Eds.), *Constraints on language: Aging, grammar and memory* (pp. 29-52). New York, NY: Kluwer Academic Publishers.

Xiao, A., He, J., Wang, Q., Luo, Y., Sun, Y., Zhou, Y. P., . . . Dai, J. P. (2011). The origin and development of plaques and phosphorylated tau are associated with axonopathy in Alzheimer's disease. *Neuroscience Bulletin*, *27*(5), 287-299.

Xiong, D., Zhang, M., & Li, H. (2011). A maximum-entropy segmentation model for statistical machine translation. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(8), 287-299.

Yngve, V. (1960). A model and hypothesis for language structure. *Proceedings of the American Philosphical Society, 104*(5), 444-466.