Longitudinal Language Analysis in the Diagnosis of Mild Cognitive Impairment: A Case Study

A Preprint

Jomar C. Alcantara
Department of Computer Science
School of Engineering and Applied Sciences
Aston University
Birmingham, UK
alcantaj@aston.ac.uk

Peter Sawyer
Department of Computer Science
School of Engineering and Applied Sciences
Aston University
Birmingham, UK
p.sawyer@aston.ac.uk

Cristina Romani
Department of Psychology
School of Life and Health Science
Aston University
Birmingham, UK
c.romani@aston.ac.uk

George Vogiatzis
Department of Computer Science
School of Engineering and Applied Sciences
Aston University
Birmingham, UK
g.vogiatzis@aston.ac.uk

February 8, 2019

ABSTRACT

Dementia is a neurological disorder that has been consistently growing in numbers over time. The analysis of language has been highlighted as a potential method for diagnosing Alzheimer's disease at its early or pre-clinical stages. This paper builds on the work of Berisha and Liss who looked at extracting language features from the non-scripted portions of press conferences from two US presidents, President Ronald Reagan, who would later go on to receive a diagnosis of Alzheimer's Disease and President George Herbert Walker Bush who received no such diagnosis. In this paper, we add the transcripts of President Donald John Trump who in terms of age is a much more appropriate match and we extract a larger variety of language features which have been shown to be associated with language decline due to Dementia. Once the features were extracted we completed a regression analyses. We found that President Reagan showed a significant decrease in unique words over time and his usage of nouns and adjectives. We also found that there was an increase in his usage of non-specific nouns, adverbs and pronouns. We found no significant trends in the transcripts of Presidents Bush and Trump. This study shows that language deterioration can be tracked over time in those with Mild Cognitive Impairment and is generally differentiable from healthy controls. The results also potentially demonstrate the utility of longitudinal language analysis as a method of diganosing Mild Cognitive Impairment in the future.

*K*eywords Mild Cognitive Impairment · Alzheimer's Disease · Natural Language Processing · Longitudinal · Early Diagnosis · Language

1 Background

Dementia has been identified as one of the fastest growing difficulties facing the world today. A recent report suggests that in 2015 there were 46 million people with a diagnosis of dementia and that number is expected to hit 131.5 million by 2050 [1]. The report also states that the worldwide cost of dementia in 2018 is estimated to be in the region of one trillion US dollars.

In 2009, the UK's Department of Health designed it's National Dementia strategy and as part of this made early diagnosis and support one of it's key priorities [2]. A lot of work has gone into this with research focused on two areas - identifying biological markers and analyzing the cognitive decline of those who are suspected to have the disease [3]. Recently, researchers have highlighted that technology, more specifically the domains of Natural Language Processing (NLP) and Machine Learning (ML), is ideally placed to aid in this endeavour [4]. It is theorized that we may be able to use these technologies to enable an earlier diagnosis of those with MCI and AD using samples of spontaneous speech, natural language processing (NLP) and machine learning (ML) [5, 4].

There is a large body of research that looks at the decline in language of those with MCI and AD [3, 4, 6]. However there is conflicting evidence in these studies about which declining language factors are associated with MCI and AD [3, 4]. Some work therefore should look at these features in more detail and a clarification of this currently disorganised picture should go some way to helping researchers further understand the disease and it's progression. There is some evidence to support this, Berisha et al [7], has shown through a longitudinal language analysis of spontaneous speech that there are marked differences in different parts of speech in this process between those who would go on to have a diagnosis of AD and a healthy control.

The potential impact of this research in this area is immense. Research has shown that early diagnosis of people with AD or MCI improves sufferers quality of life and can, in some cases, slow the progress of the disease. Early diagnosis can increase the number of research opportunities for understanding the early stages of dementia and how the disease progresses so that more research can be conducted which may, in the future, lead to new treatments and other interventions.

There has been a significant research in the area of language deterioration as a means of detecting Alzheimer's Disease. This usually takes the form analysis of speech recorded as part of a cognitive assessment such as the Picture Description Task [5, 8]. Given that language samples are relatively easy to collect, research has moved towards analysis of spontaneous speech. An good example of this type of research is the study conducted by Berisha and Liss which looked at the differences in language use between two US presidents, Ronald Reagan (who would go on to receive a diagnosis of Dementia) and George H.W. Bush

who acted as a matched control based on Age [7]. They found several differences in language use which they felt acted as indicators of Reagan's difficulties with language due to dementia. These significant differences were in the number of unique words used per speech, the use of non-specific nouns and fillers and low-imageability verbs [7].

This study replicates work done by Berisha and Liss and extends this by adding Donald Trump as an alternative, more appropriate comparison to Ronald Reagan as he is much closer in age than George H.W. Bush. This experiment will look at the features originally identified by Berisha and Liss, as well as any others that have potential as discussed in the literature review above.

2 Methodology

I took 46 transcripts of Ronald Reagan's (RR) press conferences from 1981 to 1988 and compared them with 134 press conferences by George H. W. Bush (GHWB) and 29 press conferences conducted by Donald J. Trump (DJT). I analyzed transcripts for lexical features shown to change longitudinally with dementia (for a comprehensive review of these, see the literature review above). For this collection of documents, I generated a number of features which looked at a number of different aspects of each document. These features encompassed, word level, sentence level and document level features and included a number of features contained in the study by Berisha and Liss with the aim of replicating and extending on their findings. These findings were. number of unique words, non-specific nouns and fillers and low imageability (LI) verbs. Imageability is characterized, according to Berisha and Liss, as the ease with which a term gives rise to a sensory .mental image. I compared the trends described in the transcripts of RR and GHWB, but also included DJT. Berisha and Liss originally made the comparison as it GHWB (GHWB - age at the start of presidency - 64 years, 222 days) was the closest match to RR in terms of age (RR - age at the start of presidency, 69 yeaars and 349 days). However, with the inauguration of Trump, he now is the closest comparible president in terms of age (DJT - age at the start of presidency - 70 years, 220 days). It would be interesting to look at a comparison of RR and DJT to see whether the comparisons made by Berisha and Liss hold true with this more appropriate match (in terms of age). DJT as with GHWB has no known diagnosis of AD. I used the press conference transcripts in the American Presidency Project (APP) archive as a data source for this project. The APP is a comprehensive and organized searchable database of presidential documents, including transcripts of speeches, transcripts of news conferences, and other public documents.

2.1 Pre-processing

To generate the files necessary for analysis, I downloaded each transcript and performed the following changes. I

omitted the prepared statement by the president and any speech by other individuals. I started each transcript at the beginning of the first answer to a question. I filtered any annotations that were added to the transcript, including any references or clarifications, and any laughter. It's worth noting that there appears to be a difference in how 'hesitations' were marked down between each president, for RR hesitations were marked by a single hyphen whereas for GHWB hesitations are marked by a double hyphen. In order to maintain consistency when parsing through the documents, I have changed both types of hesitation to be marked by a single hyphen. I also omitted one word sentences as this data would, from a theoretical perspective, not be relevant for language analysis. I did not control for the length of the document, but generated features which would normalise by the length of the document. I therefore was able to include all press conferences by both Ronald Reagan and George H.W. Bush where there was a question and answer session conducted at least in part by the sitting president (2 press conferences of GHWB were ommited due to a lack of a question and answer session).

2.2 Feature Selection

We calculated the following features for each transcript in turn using the NLTK (see section 2 for a description) [9] and Python.

2.2.1 Measures of lexical variation

We constructed two features of lexical variation. Firstly we looked at the number of unique words. To do this we were able to split each transcript into individual words and changed them to lowercase using NLTK and were then count the number of unique words that appeared in each transcript. We also used the TTR formula (see section 2 for a description) for a feature that measures lexical diversity independent of sample size [10].

2.2.2 Fillers, Non-Specific Nouns and LI Verbs

For these features, we counted the number of occurrences for different categories (see table for list of categories tracked and the words counted). The features were used by Berisha and Liss in their research [7] and were taken from work done by Bird et al [11].

Category	Words
Fillers	*"well", "so", "basically", "actually", "literally", "um", "ah"*
Non Specific Nouns	*"something", "anything", "thing", "everything"*
LI Verbs	*"be", "come", "do", "get", "give", go", "know", "look", "make", "see", "tell", "think", "want"*

Table 1: Categories and Words Counted

2.2.3 Usage of parts of speech

This section involves using a Part of Speech tagger (PoS) which analyses a sentence and assigns a 'tag' to each word based on the function the word has in a sentence. At a basic level this can be divided into the eight defined parts of speech: 'nouns', 'pronouns', verbs', 'adjectives', 'adverbs', 'conjunctions', 'prepositions' and 'interjections' but can be further subcategorised. We used the PoS tagger built into NLTK to tag each transcript in turn and used these the counts from each of these eight categories in our analysis. In addition to frequency counts we also normalised these features by dividing the frequency count by the number of words in the document to take into account transcript length.

3 Results

One of the most important thing to note is the wide variety of samples between the three presidents and also the varying timescales. RR participated in 46 press conferences over eight years (an average of 5.75 a year) which is the fewest number of press conferences given by an American president during their term of office. GHWB participated in 136 press conferences over four years (an average of 34 a year) and DJT participated in 29 press conferences to date (an average of 19.3 per year). Equally, there are variances in the average number of words. RR produced an average of 3424 words per conference compared to 2608 by GHWB (unpaired t = 4.434, p<0.001) and DJT at 1849 words (unpaired t = 6.524, p<0.001).

	RR	GHWB	DJT
Total Words	3423.91 (416.42)	2607.72 (1210.38)	1848.65 (1549.38)
Unique Words	894.13 (85.15)	667.76 (218.67)	481.82 (221.29)
Non Specific Nouns	12.72 (4.63)	6.78 (4.32)	7.41 (8.75)
LI Verbs	124.22 (17.89)	103.45 (51.87)	84.48 (75.78)

Table 2: Means and Standard Deviations of important features

In terms of unique words, we found that RR used significantly more unique words, non-specific nouns and low imageability verbs than GHWB and DJT (see Table 3.3). Some of these differences are due to the length of the sample, particularly in the case of DJT where his average sample is almost half the sample of RR. It could also be said that this could be down to differences in linguistic abilities or speaking style [7, 12]. However, we can certainly see that as controls GWHB and DJT are comparative in relation to non-specific nouns and LI verbs.

	RR v GHWB	RR v DJT	GHWB v DJT
Total Words	**4.434***	**6.524***	**2.899**
Unique Words	**6.832***	**11.403***	**4.137***
Non Specific Nouns	**7.877***	**3.426**	-0.574
LI Verbs	**2.656**	**3.420***	1.628

* denotes p<0.05

** denotes p<0.01

*** denotes p<0.001

Table 3: RR T-tests vs GWB and DJT

	RR	GHWB	DJT
Word Count	**-0.31***	**-0.21***	0.08
Unique Words	**-0.56***	**-0.25**	0.16
Non Specific Nouns	**0.30***	-0.08	-0.03
LI Verbs	-0.19	**-0.20**	0.02
Nouns Normalised	**-0.70***	-0.03	0.14
Verbs Normalised	**0.36**	**0.24***	-0.03
Adjectives Normalised	**-0.40**	0.08	-0.34
Adverbs Normalised	**0.41***	0.02	**-0.36***
Pronouns Normalised	**0.65***	0.13	0.07

* denotes p<0.05

** denotes p<0.01

*** denotes p<0.001

Table 4: Pearson Correlations for Features

4 Discussion

President Reagan received his diagnosis of AD in August 1994 but using transcripts of speeches he made in his two terms as President (January 1981 - January 1989) we have be able to identify certain changes in his use of language that we might ascribe to the onset of MCI and early AD. Despite differences in our methodology, our research supports the findings of Berisha and Liss in that we both find a significant decrease in unique words over time and an increase in non-specifc noun usage. Compared to our controls (GWHB and DJT), we find some slight trends with GWHB but no such trends with DJT in his speech albeit his samples of speech span a shorter amount of time.

We then looked at the data from a longitudinal perspective as we are interested seeing whether we can track various language variables and their progress over time. We ran a number of Pearsons correlations with transcript index number as a time reference and the dependant variables (Table 3.4). For our controls, we found them to be stable for the most part with the main highlights being a decrease in Adverb usage for DJT (R=-0.36, p=0.049) and a steady but not severe decline in a number of variables for GWHB, namely total word count, unique words, low imageability words and verb usage.

For RR, his decline is more marked and more widespread through his language use. We noticed an significant increase in adverb (R=0.41, p=0.004) and pronoun usage (R=0.65, p<0.001), as well as a slight usage increase in Non-specific nouns(R=0.30, p=0.03). There was a highly significant decrease in number of unique words (R=-0.56, p<0.001) and noun usage (R=-0.70, p<0.001). Also very significant decrease in adjective usage (R=-0.40, p=0.005) and a significant decrease in total word count (R=-0.31, p=0.03).
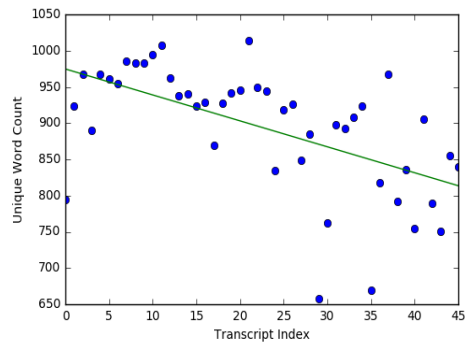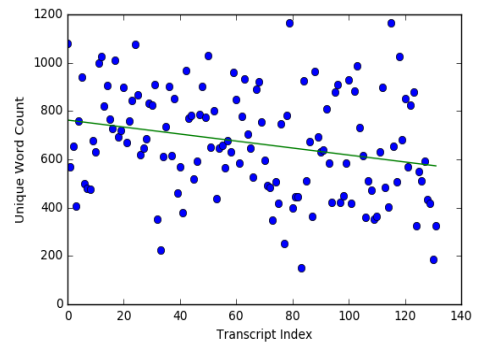
Figure 1: Ronald Reagan - Unique Words over time

Figure 2: George H.W. Bush - Unique Words over time

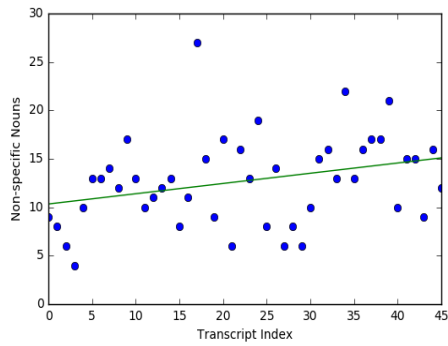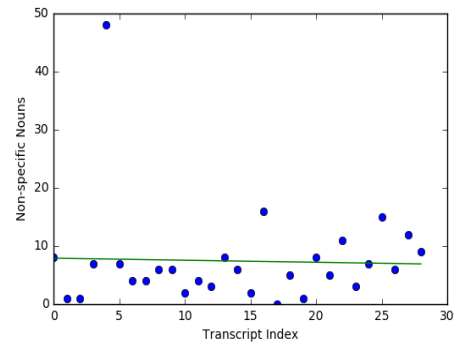

Figure 3: Ronald Reagan - Non-specifc Nouns over time

Figure 4: Donald J. Trump - Non-specific Nouns over time

A criticism of Berisha and Liss's work is the problems they had with normalising the transcripts in terms of length. This was also a problem in the work of Garrard et al [13, 12]. Whilst it is important to control for outliers, there are other ways in which we can control for length of sample.

Interestingly, when we normalised the various types of words used by the presidents we found some interesting patterns that further differentiated RR from the controls. Whilst Non-specific nouns increased over time, we found that noun usage in general significantly decreased and pronouns increased similarly significantly. The increase in pronoun for those with early AD has been identified in literature, although there are only a few studies that explore this [14, 15]. Wendlestein et al propose that the increased used of pronouns is an expression of an impaired ability to adapt language to the listener's needs [15]. Almor et al attributed this reliance on pronouns due to a impaired working memory [14].

The decrease in overall noun usage has also been identified as a feature. Jarrold et al found that AD patients would use more pronouns, verbs and fewer nouns than controls [16]. Wendlestein in their investigations into noun usage found that decreased later on in AD progression and was unaffected in the pre-clinical stages of AD [17]. Our results are supported by existing literature and this potentially means that language analysis in the way we have structured it may have diagnostic or prognostic properties.
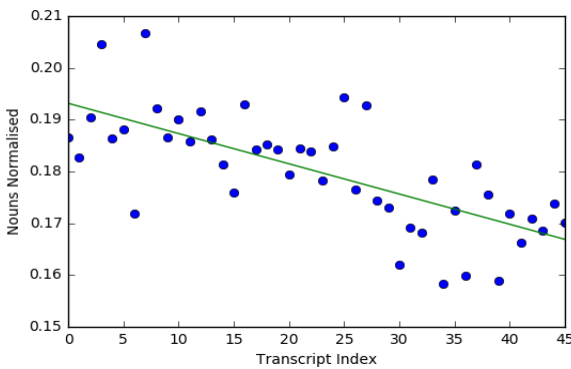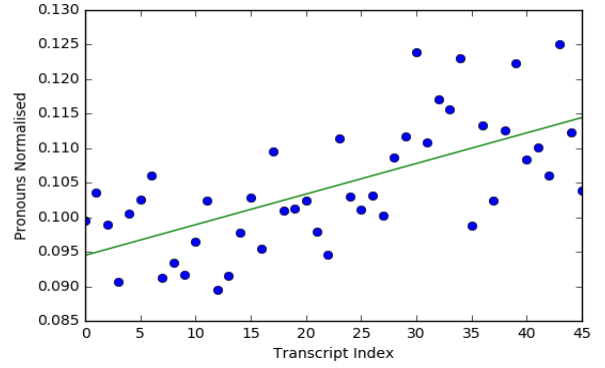

Figure 6: Ronald Reagan - Pronouns Normalised over time

There are limitations of this research. Whilst in terms of age, DJT is certainly more suitable as a control to match with RR, in some ways they held very different styles of press conferences in that RR preferred to do solo press conferences and DJT has shown a preference for doing joint press conferences which have an impact on the amount of language produced. This artifact of the data is in itself notable as it illustrates the problems we may have with smaller amounts of speech. Also, the problem of finding an appropriate control is a common one in this domain. Given that those with MCI and early dementia have such variable presentations, it might prove of limited value in matched pairs design.

With further work, it is not feasible to the vast array of samples over a timeframe, as we have had with the president corpus and so it would be worth exploring how the quality of these predictions might lessen when faced with considerably fewer samples and over a smaller time period. It would also be worth extending this research further to encompass more of the linguistic features Fraser used in her work [8] to see if there are any further insights to be gained. In addition, this replication and extension has demonstrated the potential utility of using longitudinal data as a means of comparing language use of a person at two or more time periods and using this information as a diagnostic aid for MCI and therefore more work would be helpful from a longitudinal perspective to see if this approach may be valid in moving towards a solution for this particular problem.

The results of this work show that we can track a person's use of language through time in a number of ways, and it is possible for an individual to be his or her own control. This is important as it means the heterogenous nature of the MCI population does not impact results as much as if we were comparing those with MCI to controls. Equally, it would be helpful to have controls to ascertain what would be usual to expect in the decline of language in a healthy older adult.

Figure 5: Ronald Reagan - Nouns Normalised over time

References

[1] Martin Prince, Anders Wimo, Guerchet M, Ali GC, Wu YT, and Prina M. World Alzheimer Report 2015 The Global Impact of Dementia An analysis of prevalence, incidence, cost and trends. *Alzheimer's Disease International*, 2015.

[2] Department of Health. Living well with dementia : Dementia Strategy (England). Technical report, Department of Health, 2009.

[3] Vanessa Taler and Natalie A. Phillips. Language performance in Alzheimer's disease and mild cognitive impairment: A comparative review. *Journal of Clinical and Experimental Neuropsychology*, 30(5):501–556, 2008.

[4] Veronica Boschi, Eleonora Catricalà, Monica Consonni, Cristiano Chesi, Andrea Moro, and Stefano F. Cappa. Connected speech in neurodegenerative language disorders: A review. *Frontiers in Psychology*, 8(MAR), 2017.

[5] Sylvester Olubolu Orimaye, Jojo Sze-Meng Wong, and Karen Jennifer Golden. Learning Predictive Linguistic Features for Alzheimer's Disease and related Dementias using Verbal Utterances. *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 78–87, 2014.

[6] Olga B. Emery. Language Impairment in Dementia of the Alzheimer's Type: A Hierarchical Decline? *International Journal of Psychiatry in Medicine*, 30(2):145–164, jul 2000.

[7] Visar Berisha, Shuai Wang, Amy LaCross, and Julie Liss. Tracking Discourse Complexity Preceding Alzheimer's Disease Diagnosis: A Case Study Comparing the Press Conferences of Presidents Ronald Reagan and George Herbert Walker Bush. *Journal of Alzheimer's Disease*, 45(3):959–963, 2015.

[8] Kathleen C. Fraser, Jed A. Meltzer, and Frank Rudzicz. Linguistic Features Identify Alzheimer's Disease in Narrative Speech. *Journal of Alzheimer's Disease*, 49(2):407–422, oct 2015.

[9] Steven Bird, Ewan Klein, and Edward Loper. *Language Processing and Python*. O'Reilly Media Inc, 2009.

[10] Brian Richards. Type/Token Ratios: what do they really tell us? *Journal of Child Language*, 14(02):201, jun 1987.

[11] Helen Bird, Matthew A. Lambon Ralph, Karalyn Patterson, and John R. Hodges. The rise and fall of frequency and imageability: Noun and verb production in semantic dementia. *Brain and Language*, 73(1):17–49, 2000.

[12] Xuan Le, Ian Lancashire, Graeme Hirst, and Regina Jokel. 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, 2011.

[13] Peter Garrard, Lisa M. Maloney, John R. Hodges, and Karalyn Patterson. The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. *Brain*, 2005.

[14] Amit Almor, Daniel Kempler, Maryellen C MacDonald, Elaine S Andersen, Lorraine K Tyler, Lee Willis, Lori Altmann, Mariela Gil, Laila Lalami, Karen Marblestone, Sarah Schuster, and Karen Stevens. Why Do Alzheimer Patients Have Difficulty with Pronouns? Working Memory, Semantics, and Reference in Comprehension and Production in Alzheimer's Disease. *Brain and Language*, 67:202–227, 1999.

[15] Britta Wendelstein, Jörn Stegmeier, Claudia Frankenberg, Ekkehard Felder, and Johannes Schröder. Changes in the use of pronouns in spoken language in the course of preclinical to early stages of Alzheimer's disease. *Alzheimer's & Dementia*, 11(7):P784–P785, jul 2015.

[16] William Jarrold, Bart Peintner, David Wilkins, Dimitra Vergryi, Colleen Richey, Maria Luisa Gorno-Tempini, and Jennifer Ogar. Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 27–37, 2014.

[17] Britta Wendelstein, Ekkehard Felder, Jörn Stegmeier, and Johannes Schröder. NOUN OR PRONOUN?: USING PARTS OF SPEECH IN ALZHEIMER'S DISEASE. *Alzheimer's & Dementia*, 10(4):P434, jul 2014.