Changes in the style and content of Australian election campaign speeches from 1901 to 2016: A computational linguistic analysis

Michael Dalvean, School of Politics and International Relations, Australian National University

Abstract

There have been significant social and political changes in Australian society since federation in 1901. The issues that are considered politically salient have also changed significantly. The purpose of this article is to examine changes in the style and content of election campaign speeches over the period 1901–2016. The corpus consists of 88 election campaign speeches delivered by the Prime Minister and Opposition leader for the 45 elections from 1901 to 2016. I use natural language processing to extract from the speeches a number of linguistic variables which serve as independent variables and use the year of delivery as the dependent variable. I then use machine learning to develop a regression model which explains 77 per cent of the variance in the dependent variable. Examination of the salient independent variables shows that there have been significant linguistic changes in the style and content of election speeches over the study period. In particular speeches have become less linguistically complex, less introspective, more focused on work and the home, and contain more visual and social references. I discuss these changes in the context of changes in Australian society over the study period.

1 Introduction

In this paper I track linguistic changes in the major election campaign speeches of the leaders of the incumbent party and leaders of the opposition for 45 elections from 1901 to 2016. The idea here is that tracking such changes can provide an insight into the social, political and economic phenomena which party leaders consider are most likely to influence voters at different points in time. As will be demonstrated, changes in the substantive content that over time are reasonably transparent in that the linguistic changes that take place are congruent

with the changes in political, economic and social changes. For example, the most significant change is in the use of terms relating to the home, and this is consistent with what we understand to have been a significant change in the expectations of the Australian people over the last 115 years. That is, it is accepted that a large proportion of the population now considers home owner-ship and the concept of a settled home environment to be a given, whereas in periods of war and economic depression, such issues were not as salient. Simi-larly, the issues associated with work have become more important over time. This makes sense in that the increasing wealth and prosperity over the last 115 years are associated with discussions of such things as work, income and taxes.

However, what is less transparent about the changes in the speeches is the style of the speeches. There are a number of linguistic changes in the corpus which show a number of social changes which might not have been noted by mere inspection of the speeches. The linguistic and machine learning methods used in this paper enable us to get an insight into a number of social changes that are quite surprising. One of the most striking is that, although there is a rising nominal level of education over the study period, party leaders have chosen to reduce the level of analytical content in their election campaign speeches over time. Party leaders clearly have an interest in appealing to voters and they seem to have considered voters in the later segments of the period to be less inclined to be influenced by analytical argument than they were in earlier segments of the period. Interestingly, the most 'analytical' speeches occurred in the year 1937.

Similarly, the style of speeches over time has become more visual and less introspective. Given these changes we might be tempted to say that whatever gravitas was associated with political activity in days gone by has effectively diminished and leaders in more recent years have realised that appealing to vot-ers involves emulating the techniques used in popular entertainment.

There are two stages involved in extracting the linguistic characteristics from the speeches. The first stage involves creating a number of linguistic vari-ables from the texts. To do this I use the content analysis program Linguistic Inquiry and Word Count (LIWC). This program reduces raw text to 81 linguistic variables. As well as these 81 variables I include two additional variables. I include a variable measuring the concreteness of the speeches based on con-creteness ratings for 40,000 English words (Baesbart *et al.* 2014). I include this because there have been a number of interesting political phenomena associated with concrete/abstract language. I also include a measure of text complexity, the Flesch-Kincaid readability index (Kincaid *et al.* 1975). This was included to take into account changes in speeches that may be associated with changes in educational levels over time, the idea being that party leaders would need to tai-

lor their speeches to audiences with differing education levels at different stages in the study period. Thus, with the 81 variables from LIWC, the measure of concreteness and the Flesch-Kincaid readability index, there are 83 independent variables. The second stage involved in extracting the salient linguistic variables from the corpus consists of creating a classifier using machine learning methods. The dependent variable is the year in which speeches were delivered. With the 83 independent variables and the year of speech delivery as the dependent variable, I use machine learning to determine which of the independent variables are most important in explaining the variance in the dependent variable. The result is a selection of 23 variables which explain 72 per cent of the variance in the dependent variable.

I then use a selection procedure to further narrow down the number of independent variables to six. These consist of two substantive variables (home references and work references) and four style variables (analytical thought, cognitive processes, visual references, and a composite variable that tracks changes in positive emotion and self/other references). From these six variables we can see that the campaign speeches provide an interesting insight into social changes that have occurred over the period under consideration.

2 Previous work in content analysis

There are various types of content analysis. The most common by far is qualitative content analysis which focuses on extracting the salient content based on what the researcher considers to be salient. The method used in this paper is quantitative rather than qualitative. The approach is to use a content coding dictionary to compare texts on the basis of predetermined linguistic categories. The idea here is that if the predetermined linguistic categories are well defined and cover a variety of social, psychological and human interest categories, content differences between texts will be discernible in changes in the linguistic categories. As we will see, there is a very significant change in the content of election speeches over the years 1901–2016 as measured by such categories over time.

The specific content analysis coding dictionary used in this paper is the 2015 version of Linguistic Inquiry and Word Count (Pennebaker *et al.* 2015). The 2001 version of LIWC was developed for analysing differences in texts on the basis of social and psychological criteria. The dictionary was initially developed to determine how those undergoing psychological intervention were responding to various forms of intervention. The changes in subjects were monitored by observing how their written expression changed over time as the interventions progressed. The idea behind the subsequent development of LIWC was to cre-

ate a content coding dictionary to track the salient elements of modern human thought and emotion. Since its initial development, LIWC has been undergone two major upgrades in 2007 and 2015 and has been used in hundreds of applications in such diverse fields as psychology, political science and literary analysis.

LIWC 2015 has 81 linguistic categories, which include such categories as Negative Emotion, Affect, Leisure, Work, Family, Social Activities and Psychological Processes. Each category consists of a list of words pertaining to the category. For example, the category Family consists of 65 terms pertaining to the family such as 'father', 'sister' and 'spouse'. The value for this category for a given text is the number of times any of the 65 words is used divided by the total number of words in the text as a whole. Thus the text "I am thy father's spirit" with one occurrence of a word from the Family list and five words in total, would be scored as 1/5 or .2 on the Family category.

The efficacy of the LIWC approach has been documented in over 100 studies Tausczik and Pennebaker (2010). In the political sphere, Hirst, Riabinin and Graham (2010) used LIWC to analyse speeches in the Canadian Parliament in two periods over which the government had changed. As might be expected, the main finding here was that the government party uses more positive emotion words while the opposition party uses more negative emotion words. Niederhoffer and Pennebaker (2002) used LIWC in an analysis of the Watergate tapes. The specific linguistic characteristic they were looking for was linguistic synchrony, which is the idea that in dyadic conversation the words or types of words of one speaker will co-vary with the words or types of words of the other. The idea here is that, in conversation, the first speaker's words will prime the second speaker to use particular words and/or word categories. At both the turn-by-turn and the total conversation level, they found support for language synchrony. Dalvean (2015a) used LIWC to analyse the difference between poems written by 'professional' poets as opposed to poems written by 'amateurs'. It was found that the professional poems used fewer affect terms, more articles and fewer present tense terms than amateur poems.

Clearly, LIWC seems to be a good way of extracting useful information from texts. In the current study we will see that there are several categories that show significant changes over the study period. This indicates that, in respect of the categories that change, there have been substantial changes in the way that candidates appeal to the electorate.

3   Data and analysis

The data used consist of linguistic variables derived from the election campaign speeches of the leaders of the two major parties in contention in the elections over the period 1901–2016. The text of the speeches for 1901–2013 come from the Museum of Australian Democracy Old Parliament House website (http://moadoph.gov.au/), which provided 86 of the 88 speeches delivered by party leaders over the period. The two speeches that are missing are the 1901 George Reid speech, which is not available, and the 1910 Andrew Fisher text, which is not a verbatim transcription but a newspaper report. These speeches were therefore not included in the corpus. The texts of the 2016 campaign speeches by Liberal party leader Malcolm Turnbull and ALP leader Bill Shorten were added to the corpus. The result is a corpus of 88 speeches from 36 party leaders over the period 1901–2016.

3.1   Variables

The dependent variable is the year in which the speech was delivered. No specific coding other than the year was used. The reason for this is that the intention is to track the changes in language over time.

The independent variables were derived from several sources. The majority of variables were derived from the 2015 version of the content coding scheme Linguistic Inquiry and Word Count (LIWC). As discussed above, this scheme breaks texts down into 81 linguistic variables. Some of these variables are structural, such as Words per Sentence, and Number of words of Six Letters or More. Punctuation variables are also available but were not used because punctuation conventions are likely to have changed over the study period. The vast majority of variables used in this study are related to linguistic content such as Adverbs, Pronouns, Function words, Affect (in total as well as positive and negative) and Social Processes (such as Family, Leisure and Religion). Thus, the coding gives a highly transparent insight into the nature of the text under study.

The Flesch-Kincaid measure of reading complexity (Kincaid *et al.* 1975) was a further variable included in the study. This is a measure of the ease of reading for a given text. Although it is designed to measure reading complexity, it was included as a proxy for the overall complexity of the text. It is useful to consider this because there is a possibility that the sophistication of speakers/ speech writers has changed over the study period, possibly in response to changes in the sophistication of the electorate. This measure has the advantage of being widely understood and is relatively easy to calculate, with most standard word processors including it. One complication with its use, however, is that such things as syllable counts and hyphenated words can be counted differ-

ently by different programs. For this reason, the calculation was done by a professional reading analysis program, Readability Studio (2012).

A measure of word concreteness was included in order to track this quality in speeches over time. Texts were scored on this measure using the data from Brysbaert *et al.* (2014). This data provides concreteness ratings for 40,000 words/phrases. These ratings were then used to score the individual texts on concreteness using the method in Dalvean (2015a). Linguistic Concreteness has been associated with a number of social, political and psychological phenomena. For example, higher levels of concreteness in text have been associated with localised, contextualised thinking while lower levels of concreteness (higher levels of abstraction) have been associated with 'big picture' thinking (Trope and Lieberman 2010). Interestingly, in the current study, this measure is positively associated with Year, indicating that election speeches have become more concrete over time ($r = .41$; $p < .01$; $n = 88$).

Thus, there are 83 different variables which are in the pool of potential variables which the machine learning procedures draw from. The vast majority of these variables, we will see, are not important in the generated models. They will therefore not be described in detail. For the purposes of exposition, it is worthwhile to briefly describe the important variables before we review the results of the machine learning experiments. The variables that are important in the modelling below are described in Table 1:

Table 1: Salient linguistic variables

Variable	Definition	Examples/References	No. of Words in Category
Analytic	Analytical thinking	Pennebaker *et al.* 2014	n/a
Authentic	Speaker's belief in the text	Newman *et al.* 2003	n/a
auxverb	Auxiliary verb use	am, will, have	141
cause	Causation words	because, effect	135
Clout	Perceived social status	Kacewicz *et al.* 2013	n/a
cogproc	Cognitive process words	thought, know, ought	797
home	Words associated with the home	kitchen, landlord	100
male	Male reference terms	boy, his, dad	116
see	Words associated with sight	view, saw, seen	126
shehe	3rd person singular	she, her, him	17

work	Words associated with work	job, majors, salary	444
Flesch-Kincaid	Flesh-Kincaid reading grade level	Kincaid *et al.* 1975	n/a
they	3rd person plural	they, their, they'd	11
WPS	Words per sentence	n/a	n/a
anger	Words associated with anger	hate, kill, annoyed	130
drives	Affiliation, achievement, power, reward and risk words	allied, win, superior, danger	1103
social	Words associated with social processes	mate, talk, they	756
WC	Word count	n/a	n/a
focusfuture	Words denoting the future	may, will soon	97
achieve	Words associated with achievement	win, success, better	213
conj	Conjunction words	and, but, whereas	43
Dic	LIWC dictionary word use Per cent	n/a	n/a
friend	Words associated with friendship	buddy, neighbour	95

3.2 Machine learning

With 83 linguistic variables from 88 speeches delivered over 115 years we have a problem: the number of independent variables is far greater than we would normally attempt to analyse using traditional hypothesis testing. The approach in this paper is therefore to use machine learning to extract the salient variables from the data. With this approach we try to maximise the predictive accuracy of the model and, once we have achieved a certain level of predictive accuracy, we look at what variables are driving the model and whether the important variables are positively or negatively associated with the dependent variable. Thus, the first stage is to maximise the predictive accuracy of the model.

The actual measure of accuracy is the leave-one-out cross validated (LOOCV) correlation between the predicted and actual year of the speeches. This method of cross validation involves the creation of 88 models, one for each observation (speech). In a single iteration, a given speech is excluded and a model is built using the remaining 87 speeches. The 'held out' speech is then run through the model and the difference between actual and predicted year is recorded. The same procedure is repeated 88 times, each time with a different speech. For each of the 88 iterations there is a predicted year for the 'held-out'

speech. The correlation between the predicted and actual year for each speech is the measure of accuracy of the model. The method ensures that there has the model has not 'overfitted' the data, the situation in which the model learns the noise in the data and is unable to generalise on data from outside the sample used to build the model.

However, there is a problem with focusing on accuracy alone: there are certain methods that have a high level of predictive accuracy but which are effectively 'black boxes'. That is, with some modeling methods it is not always possible to determine how the independent variables are associated with the dependent variable. Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) are of this type. What this means is that we need to focus on more transparent induction methods such as traditional linear regression as well as more recent developments such as regression trees. Both these induction methods can provide high levels of accuracy and provide a transparent means of determining how the model achieves its result and as such we can make confident statements about what variables have changed over the 115 year study period. However, as a parallel exercise I will include the results of experiments using other induction methods such as ANNs and SVMs.

Before moving on to the modelling in earnest, it is worthwhile to review briefly the induction methods used in this paper. Note that this modeling was undertaken using Weka v3.6.13 (Witten *et al.* 2016).

Linear Regression – This is the standard traditional method for linear modeling with several independent variables. We will see that, using the variables discussed above, logistic regression achieves a Leave-One-Out cross validation accuracy of .75 . In other words, the correlation between actual and predicted year is .75 with the variance (r^2) = .56. The particular instantiation of linear regression uses the Aikaike Information Criterion (AIC) to create an optimal model. Variables are added to the training models until the minimum AIC is achieved.

Decision Stump – This is a basic form of decision tree 'regression'. Here, the term 'regression' is used to denote an induction method that relates a given dependent variable to a continuous dependent variable. The method works by trawling through the independent variables to find one variable that can be used to predict the dependent variable by 'splitting' the independent variable at some point. The algorithm searches through each variable for the split that minimises the mean squared error in the dependent variable. Thus, the decision stump in the modelling below shows that Home terms are the best single variable for predicting the year of the speech. In particular, the model assigns a value (year) of

1928.64 to speeches that have a use of Home terms =< .295 per cent and a value (year) of 1980.12 if the speech has a percentage of Home terms > .295 per cent. The split point here is achieved by varying the split and optimising the resulting mean squared error between the predicted and actual year.

M5P – M5P is a hybrid of a regression tree and standard linear regression. This algorithm works by building a model similar to that of a standard regression tree, using a method similar to that explained in the description of the Decision Stump above. In the case of a Regression Tree, variable splitting is used to create optimum cut points for variables but instead of stopping with just one variable, as occurs in a decision stump discussed above, the process continues recursively with searches through all variables. Thus, the initial split is achieved by searching through the data to find the variable with the greatest ability to minimise the error of the dependent variable. The optimal outpoint for this variable is determined and is then used to create two branches. Each branch leads to a node. At each node, either a value is assigned to the data (as in the decision stump above) which is calculated by taking the average value of the cases that reach the node, or another variable is used to create another split. The tree terminates when no further error reduction is possible. The M5P algorithm differs from the standard regression tree in two respects 1) the splitting criterion is based on the reduction of the standard deviation of the dependent variable in the subset of cases that reaches a given node and 2) the terminating nodes do not assign a value but create a linear regression model. The model therefore consists of a tree that maximises the 'standard deviation reduction' in the data in the tree stage and then when the data is 'tamed', the terminating nodes (leaves) consist of linear regression models that are used to make a final set of predictions. Thus, subsets of the data are assigned to different linear models according to 'decisions' about the subsets made at the tree stage. The benefit of this approach is that problems associated with non-normality, noise or extreme outliers are attenuated at the tree stage thereby making the data more amenable to linear modelling. The procedure also involves pruning to remove variables that do not aid error reduction and smoothing of results across the different linear models.

Multilayer Perceptron – This is a species of artificial neural network which uses back propagation to adjust the weights between the input layer, one or more hidden layers and the output layer. The learning phase uses the training sample to find relationships between the inputs (variables) and the output (dependent variable) and the internal weights are adjusted in order to approximate this relationship. The method provides good predictive accuracy in general due to its ability to model highly non-linear variables. However, interpreting the causal relation-

ships between variables is problematic. The results are included in this paper for comparative benchmarking against the other more 'transparent' methods.

The results for these induction methods are presented in Table 2. Note that the Weka defaults are used for all but the MLP with hidden layers = 1:

Table 2: Accuracy metrics for induction methods

			Model		
Accuracy	Linear Regression	Decision Stump	M5P Model Tree	MLP – Hidden Layers: 82	MLP – Hidden Layers: 1
Correlation	0.75	0.65	**0.85**	0.84	0.73
Variance	0.56	0.43	**0.72**	0.70	0.53
Mean Absolute Error	19.11	19.13	**12.24**	13.65	19.59
Root Mean Squared Error	25.83	25.10	**17.43**	19.00	26.29
Relative Absolute Error	67.07	67.14	**42.97**	47.89	68.75
Root relative Squared Error	78.09	75.90	**52.69**	57.44	79.80
Number of Instances	88	88	88	88	88

Given that the highest level of accuracy is provided by the M5P, with the correlation between predicted and actual year = .85 (variance = .72), the remainder of this paper will focus on the insights provided by the M5P model.

Figure 1 provides the decision tree and the 11 linear regression models created by the modeling process. The tree section demonstrates the basis upon which a given case (speech) is directed to one of the 11 constituent linear regression models used for prediction. If, for example, the use of home terms in a given speech is less than or equal to .295 per cent, then a second test is applied: if the use of cause terms is less than or equal to 1.755 then linear model 1 is applied. If the use of cause terms is greater than 1.755 a test relating to Words per Sentence (WPS) is applied and the results of this test result in the direction of the case to linear model 5 if WPS <= 21.345, or linear model 6 if WPS > 21.345. The model can be followed iteratively to determine what the decision points are and which linear models are applied.

```
home <= 0.295 :
|   cause <= 1.755 : LM1 (10/12.63%)
|   cause >  1.755 :
|   |   drives <= 11.515 :
|   |   |   work <= 7.62 :
|   |   |   |   they <= 0.945 : LM2 (8/6.718%)
|   |   |   |   they >  0.945 : LM3 (3/8.027%)
|   |   |   work >  7.62 : LM4 (7/6.602%)
|   |   drives >  11.515 :
|   |   |   WPS <= 21.345 : LM5 (4/10.839%)
|   |   |   WPS >  21.345 : LM6 (4/4.588%)
home >  0.295 :
|   social <= 9.135 : LM7 (29/21.79%)
|   social >  9.135 :
|   |   male <= 0.435 :
|   |   |   Clout <= 82.69 : LM8 (4/12.305%)
|   |   |   Clout >  82.69 : LM9 (5/14.604%)
|   |   male >  0.435 :
|   |   |   focusfuture <= 2.175 : LM10 (6/4.493%)
|   |   |   focusfuture >  2.175 : LM11 (8/14.325%)
```

Figure 1: M5 Pruned model tree (smoothed linear models)

The coefficients of the linear regression models are given in Table 3. The final column of the table is the zero order correlation between the variable and the year. For example, the 'Analytic' measure of the speeches is associated with the year of the speeches with a correlation of -.48 (p<.01; n = 88), indicating that the level of analytical thinking in the speeches has fallen over time.

Thus, the modeling procedure has isolated 23 variables across 11 linear models. The validation method is such that we can be sure that the variables

have not been selected due to over fitting. Furthermore, the high r^2 of .72 between predicted and actual year indicates that the modeling procedure has effectively captured the linguistic changes over time.

The next step in the analysis is to determine which of the variables can be relied on to provide an insight into what the actual changes are that have occurred over the study period. To put this question into context, consider that the variable with the greatest correlation with year is Drives ($r = .72$, $p<.01$, $n = 88$). However, this variable only occurs in five of the 11 models, indicating that it is not highly important in classifying speeches. In contrast, the index variable Clout occurs in all 11 models. On first inspection this indicates that the variable is a good indicator of changes in linguistic behaviour over time. The problem here is:

Table 3: Coefficients for 11 linear nodels in the MP5 model and correlations between variables and year

Variable	1	2	3	4	5	6	7	8	9	10	11	Variable – Year R
Analytic	-1.27	-1.62	-1.62	-1.62	-1.74	-1.86	-1.19	-1.60	-1.60	-1.51	-1.55	-0.48
Authentic	0.27	0.27	0.27	0.27	0.27	0.27	0.35	0.37	0.37	0.37	0.37	0.62
auxverb	-2.01	-2.01	-2.01	-2.01	-2.01	-2.01	-3.68	-6.11	-6.11	-6.12	-6.09	-0.17
cause	3.73	3.73	3.73	3.73	3.73	3.73	2.84	2.84	2.84	2.84	2.84	-0.21
Clout	0.24	0.24	0.24	0.24	0.24	0.24	0.03	-0.21	-0.21	-0.17	-0.17	0.73
cogproc	-0.92	-0.92	-0.92	-0.92	-0.92	-0.92	-1.83	-2.01	-2.01	-2.01	-2.01	-0.23
home	6.61	6.61	6.61	6.61	6.61	6.61	5.03	5.03	5.03	5.03	5.03	0.67
male	3.39	3.39	3.39	3.39	3.39	3.39	2.58	4.74	4.74	4.36	4.36	-0.07
see	-8.79	-8.79	-8.79	-8.79	-8.79	-8.79	-60.90	-25.17	-25.17	-25.17	-25.17	-0.22
shehe	-6.45	-6.45	-6.45	-6.45	-6.45	-6.45	-4.91	-4.91	-4.91	-4.91	-4.91	-0.18
work	5.25	5.28	5.28	5.37	5.13	5.13	1.27	1.27	1.27	1.27	1.27	0.38

Flesch-Kincaid	-2.25	-0.44	-0.44	-0.44	-0.03	-0.03						-0.33
they	-11.63	-13.14	-13.42	-12.46	-11.53	-11.53						0.23
WPS	0.81	0.02	0.02	0.02	-0.67	-0.67						-0.29
anger		-3.01	-2.18	-2.18	-3.13	-3.13						-0.37
drives		1.07	1.07	1.07	1.54	1.54						0.75
social							1.50	1.74	1.74	1.74	1.74	0.62
WC		0.00	0.00	0.00	0.00	0.00						-0.3
focusfuture										1.07	1.07	0.62
achieve							4.01					0.58
conj							-3.20					0.18
Dic				-0.26								0.47
friend	-30.63											0.37
Constant	2030.97	2049.26	2048.85	2069.78	2069.53	2080.03	2080.81	2165.54	2165.37	2163.42	2167.55	n/a

that the coefficients for this variable are positive in the first seven models and are negative for the final four. This indicates that this variable cannot be seen as a reliable indicator of linguistic change over the study period. Thus, what we need to consider is those variables which 1) occur in all models and 2) do not change their coefficients over the 11 models. A third criterion is also required: we must ensure that 3) the sign of the coefficient is the same as the sign of the zero order correlation. The reason for this is that if they differ, the variable is likely to be a suppressor variable and can therefore not be said to be reliably related to the dependent variable. The variable male, for example, occurs in all models, has the same sign for the coefficients in each model but has a non-significant negative correlation with the dependent variable (r = -.07; ns; n = 88). The difference in sign between the zero order correlation and the coefficients indicates that the variable achieves its positive coefficients by 'suppressing' irrelevant variance in the other variables in the models. Thus, it is not a 'predic-

tor' in the classical sense. The final criterion is that 4) the zero order correlation must be significantly correlated with the independent variable. The reason for this is that we are looking for variables that are associated with linguistic change over time. The above modeling process is a way of winnowing the unimportant variables away from the important variables. However, we still have a high number of independent variables, some of which may not shed significant light on the real linguistic changes over time. By specifying a requirement for significant zero order correlation, this problem is attenuated. With a sample size of 88 the cutoff for significance of the zero order correlation at the $p<.05$ level is .210 while that for the $p<.01$ level is .273. I shall use the $p<.05$ level.

Applying these criteria we are left with the following variables: home ($r = .67$); Analytic ($r = -.62$); Authentic ($r = .48$); work ($r = .38$); cogproc ($r = - .23$) see ($r = .22$).

The above procedure was designed to identify the salient variables in the large pool of variables. We have identified 6 out of the potential 83 variables. In order to provide a final test for the efficacy of these six variables it is worthwhile to create a standard linear regression model using these variables. This provides a transparent means of seeing how the independent variables are related to the dependent variable. Table 4 shows the parameter estimates for the model:

Table 4: Parameter estimates for the six variable model

Variable	Coef.	Std. Err.	t	P>|t|	[95% Conf.Interval]	
analytic	-2.52	0.33	-7.61	0.00	-3.17	-1.86
authentic	1.22	0.31	3.96	0.00	0.61	1.83
cogproc	-4.15	1.74	-2.39	0.02	-7.61	-0.70
see	-45.80	16.74	-2.74	0.01	-79.11	-12.50
work	6.55	1.36	4.82	0.00	3.85	9.25
home	31.87	7.15	4.46	0.00	17.63	46.10
_cons	2154.62	47.67	45.20	0.00	2059.76	2249.47

The model is efficient in explaining variance in the dependent variable: $F_{(6, 81)}$ = 57.51; Prob > F = 0.00; R-squared = .8099; Adj R-squared = 79.88; Root MSE = 14.859. Importantly, the LOOCV accuracy is .8789 (r^2 = .772). Interestingly, this is significantly higher than the accuracy of the linear regression model using all variables reported in Table 2 of .75 (r^2 = .56). It is also higher than the accuracy of the M5P model (r = .85; r^2 = .72).

Thus, the foregoing modelling procedures have identified 6 highly predictive variables out of a pool of 83 variables and determined that the relationship between these 6 variables and the dependent variable is not the result of overfitting. In the next section we shall begin our examination of these variables and what the variation in them over time tells us about changes in the social circumstances of the Australian body politic.

4   Discussion

4.1   Changes in substantive content: Home and work

The home has played an important role in the evolution of election strategy over the last 115 years. Figure 3 shows the yearly rate of home term usage for each election year. The average is .40 (sd = .24, n = 45). It should be noted that for the purposes of creating the graphs in this section, the average was taken for the two speeches in each year for all years except 1901 and 1910, for each of which only one speech was available. Thus, the unit of analysis is the speech-year:
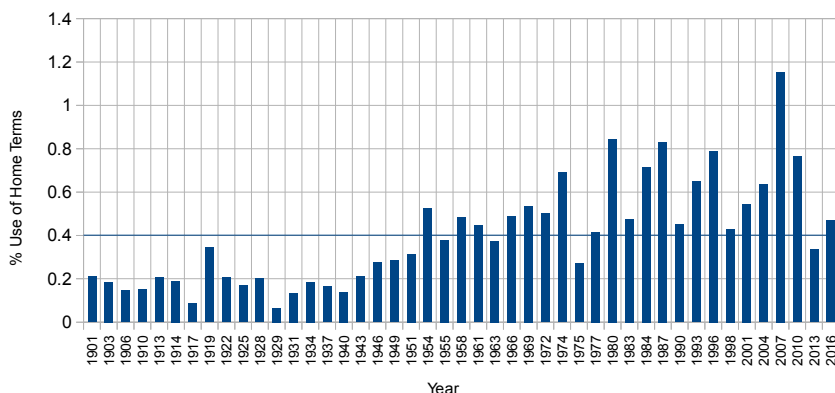
Figure 2: Per cent use of home terms by year

There is a spike in the discussion of homes after WW1 and as the government took pains to re-settle returning soldiers. There was also a rise in 1946 to .275 which, excluding 1919, was the highest since Federation. However, instead of falling in subsequent years, as it did in 1919, it rose to .285 in 1949, fell slightly to .315 in 1951 and then jumped to .525 in 1954. From here there next major change in the trend was in 1974 when the rate jumped to .69. Interestingly, it

19

was a the conservative candidate Billy Snedden who used more home terms in that year at .77 compared to the liberal Whitlam who used home terms at a rate of .61. In 1975 both the liberal Whitlam and the conservative Fraser reduced their respective parties' previous high levels of home references to .33 and .21 respectively for an average rate of .27. Clearly, the home had fallen as a topic of election priorities as the 1975 Australian 'constitutional crisis' was the major focus of the parties and the electorate. After a high of .835 in 1980 and there was a slight downward trend until 2004. In 2007, John Howard's (conservative) rate of 1.28 and Kevin Rudd's (liberal) rate of 1.030 boosted the rate to 1.155, the highest since federation. The subsequent two elections in 2010 and 2013 saw two consecutive falls to .765 and .335 respectively. By 2016 there was a slight increase to .470

The discussion of the home seems to map essentially what we might expect. After wars, there is a rise due to the attempt to re-settle returning soldiers. However, after WWII the rise did not subside, as it did after WWI, due to the long boom and the rise of the welfare state, an important element of which was the idea of home ownership. The interest in 1975 was on the constitutional crisis. The rise in 1980 and subsequent slow declining trend until 2007 is difficult to interpret. The falls in 2010, 2013 and 2016 may reflect a fall in the expectations of government in a post-Global Financial Crisis world that it might be best to emphasise other matters given that the goal of home ownership may be becoming increasingly difficult to achieve.

Work terms have played an increasingly important role in election speeches in the study period. Figure 3 shows the use of work terms in election speeches over the period:
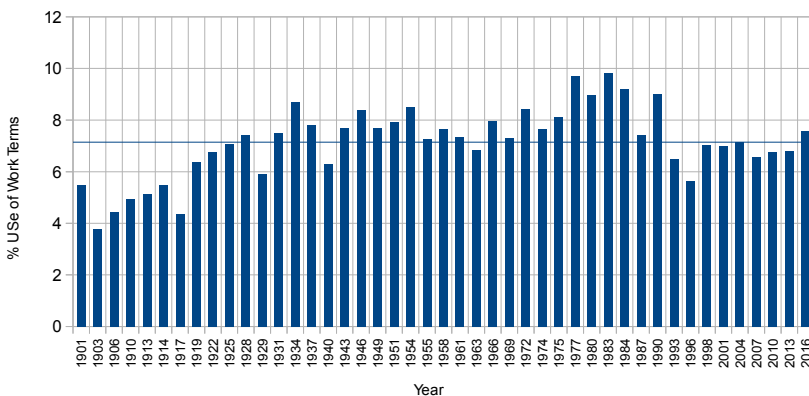

Figure 3: Per cent use of work terms by year

The average rate of work term use was 7.14 per cent (sd = 1.38, n = 45). The most rapid rate of increase in the use of work terms was from 1903 to 1934. From 1937 to 1975 the trend plateaued and then rose again in 1977. With the exception of 1987, work term use was well above average until 1993, after which work term use was at or below average until 2016 when the rate rose to levels slightly above average.

The rise after Federation in 1901 may have been due to the rhetorical strategy associated with the transformation of the Australian economy from an essentially agricultural to an industrial economy. This was a period in which there were rising international commodity prices, which favoured Australian wool and wheat exports, but this was accompanied by rising government debt associated with the development of industrial infrastructure (Copland 1934: 15). The fall in 1929 occurred in the election campaign of 12 October that year, which was prior to the October 1929 stock market crash but after a series of economic and industrial problems in the 1928–9 period. Thus, the rhetorical strategy may have been to de-emphasise these politically sensitive issues.

By 1934 the unemployment rate in Australia had risen from its high in 1932 and the rise in work term use effectively signaled the beginning of the recovery period which, with the exception of 1940, when discussion turned to war, continued until the 'stagflationary' period of the early 1970s. The rise in work terms in 1977 occurred in an election the context of which was the economic crisis of 1973-5 associated with the Whitlam (liberal) government which had lost the 1975 election. In 1977 Whitlam ran against the incumbent Fraser (conservative) and much of the rhetoric was associated with economic vilification of Fraser, on the part of Whitlam and economic justification, on the part of Fraser.

The 1980s were characterised by a recession in 1980–81 followed by a rise in economic growth until the stock market crash of 1987. The election occurred at the height of the boom in July 1987 but, interestingly, the use of work terms fell in that election in comparison to the previous election.

In 1990 there was a recession which led to the failure of a number of financial institutions. However, these events occurred well after the election in March 1990. The high level of work terms is due to the high profile given in the election campaign to economic issues including the discussion surrounding the Coalition opposition's policy of abolishing the capital gains tax.

For the 20-year period 1993–2013 the use of work terms fell in comparison to the highs of the 1977–1990 period. This is so even for the election of 2010 which followed the Global Financial Crisis. One possible explanation for this is that social issues replaced work issues due to the high level of prosperity being experienced by the Australian people. Although social term use is not one of the

6 variables under consideration it is worthwhile considering this variable for the insight it gives us into the use of work terms. If we consider Table 3 we can see that the use of social terms is positively associated with year. Figure 4 demonstrates that for all elections from 1993 onwards, the use of social terms is above the average of 7.80 (sd = 1.97, n = 45). For all years after 1993 other than 2001 and 2007, the rate was more than one standard deviation above the average:
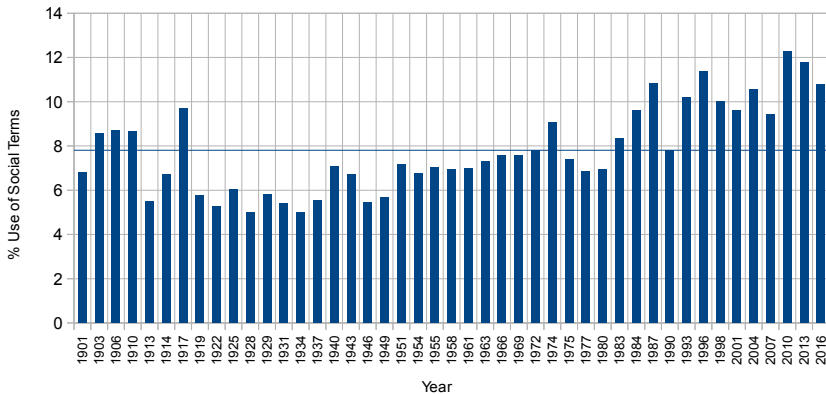

Figure 4: Per cent use of social terms by year

4.2   Changes in style: Analytic, authentic, cogproc and see

The Analytic variable is an index of the level of analytical thought (Pennebaker *et al.* 2014). The measure was derived by analysing the linguistic differences between 50,000 US college entrance essays of students who subsequently achieved high college grades and those who subsequently achieved low college grades. In particular, high GPA students tended to use high levels of articles and prepositions, indicating a focus on concepts and categories while the low GPA students tended to use more 'dynamic' language associated with personal experience. The Analytic index is based on the 'Categorical – Dynamic Index', the calculation of which is as follows: CDI = 30 + article + preposition – personal pronoun – impersonal pronoun – auxiliary verb – conjunction – adverb – negation.

The average Analytic score over the study period is 90.27 (sd = 6.30, n = 45). Figure 5 shows that the analytical style of speeches has fallen over the study period. However, the fall has not been uniform. There was a rapid rise over the 6 elections from 1901 to 1914. After a fall in 1917 there was a plateau

which undulated around the mean. The high point for the series occurred in 1937 at 96.905. Interestingly, this was at a time when, in contrast to today, most students did not finish high school and only a small fraction of the population attended university. Nevertheless, leaders felt it appropriate to aim their speeches at a higher level of analytical thinking than has obtained since that time. The possibility is that the intensity of the school education at the time was such that the population was able to handle relatively high levels of complex thinking. There is some evidence for this in surveys of the reading material presented to school students over the last century (Stotsky 1990), although this observation should be considered in the light of the realisation that far fewer students attended high school in the past:
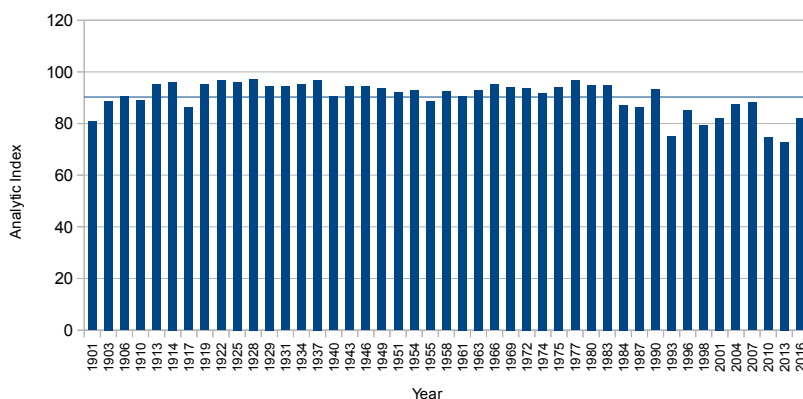

Figure 5: Analytic index by year

The speeches of the election campaign of 1984 signaled the beginning of a generally decreasing trend over which, with the exception of 1990, all scores were below the long-term average. It seems that the modern Australian leader considers the voting population as being more interested in personality than conceptual content in election speeches.

This perception is reinforced if we consider another measure of cognitive content – cognitive processes. Figure 6 shows the percentage use of Cogproc terms over the study period. This is an index of terms associated with cognition such as 'consider', 'contemplate' and 'know'. It maps the use of the internal workings of the speaker/writer's mind. There is some evidence that it maps the kinds of introspective, reflexive thinking that occurred with the emergence of

the self-awareness during the Axial Age (Dalvean 2015b). Being of a cognitive nature, it is related to the Analytic index but it must be kept in mind that both these variables occur in each of the constituent linear regression models of the MP5 model and their joint inclusion in the models can only occur if they independently contribute to a lowering of the AIC. Thus, despite both mapping cognitive phenomena they each map different cognitive phenomena, with the Analytic index mapping conceptual, object oriented cognition and cogproc mapping the internal workings of the mind.

The general trend is that cogproc terms have fallen in election speeches. Interestingly, the highest score was in 1993 at 11.37 (average = 9.06; sd = .971, n = 45), which represents a deviation in the generally declining trend in the later segment of the study period.

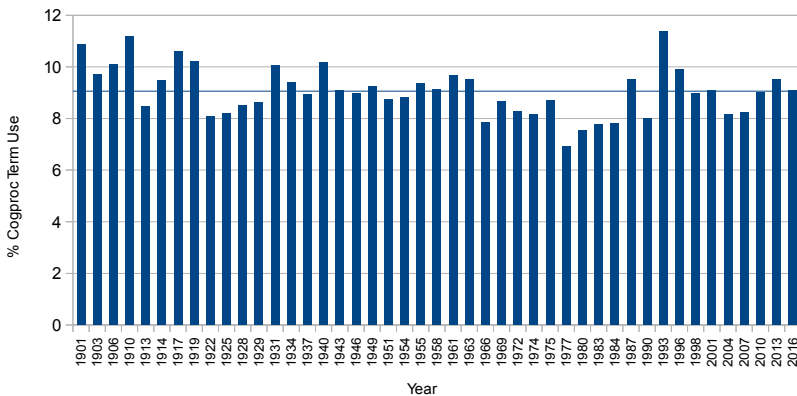However, the second highest score was 11.19 in 1910:

Figure 6: Per cent cognitive process terms by year

The period with the highest cognitive process term use was 1901 to 1919. This was followed by a significant drop in 1922 and a steady rise over the next four elections. From 1934 to 1963 there was a fairly steady rate of cogproc use that remained close to the mean. From 1963 to 1984 the rate remained below average and then jumped above average in 1987 and fell back in 1990. After the high of 1993 there was a general decline and the last three elections have seen the rate hover close to the average.

The overall pattern is difficult to interpret. Clearly the speeches of the period 1901–1919 had higher levels of internal thought than those of 1966–1984,

which is not the case with the Analytic Index. Similarly, the last three elections demonstrate a reversion to the average for cognitive processes whereas the Analytic index recedes from the average for these elections. But how these observations shed light on the social conditions accompanying the speeches is difficult to determine. One possibility for the generally falling trend is that reflective thinking has become 'unfashionable' and leaders have increasingly realised that an appeal to the external is preferable to a more 'contemplative' approach to campaigning. One indication that this may be the case is from the 1910 speech from Alfred Deakin, who was Australia's second Prime Minister and was the incumbent in 1910. His 1910 speech has the second highest cogproc score of 11.19. Referring to the absence of party discord, he states that "There is no present rift within the lute...". This is a reference to Tennyson's *Merlin and Vivien*:

> Unfaith in aught is want of faith in all.
> It is the little rift within the lute
> That by-and-by will make the music mute,
> And ever widening, slowly silence all.

Interestingly, he feels no need to point out the origin of the phrase, which suggests that his audience was well aware of the source. Poetry in general is considered an introspective muse and the lute is possibly the most introspective instrument of all. That Deakin felt this reference would resonate with his audience suggests that the audience was open to introspective pursuits.

Let us now consider the speech with the lowest cogproc score which was delivered by the incumbent conservative Prime Minister Malcolm Fraser in 1980. This speech scores 5.85 which is more than three standard deviations lower than average. If we look for a similar kind of reference to that in Deakin's speech, such as a reference to the arts, we find the following statement: "The world of film, science, literature, theatre, opera, dance and sport has produced great Australians." Although ostensibly referring to several traditionally introspective activities (along with some notoriously non-introspective activities), Fraser's actual focus is people – "great Australians" – rather than the elements of the cited muses themselves. It is well established that a focus on people is a characteristic of extraversion (Lakhal *et al.* 2012).

These observations should be interpreted with one caveat – the speech with the highest cogproc score in the corpus is conservative opposition leader John Hewson's 1993 speech, which scores 11.38. This speech is characterised by even lower references to introspective activities than Fraser's, the high cogproc score being due to the high use of terms such as 'believe' and 'interests'. It is

likely that this use of introspective terms was a deliberate strategy on the part of the party to present Hewson as a nuanced personality given that he was perceived as being uncharismatic. His opponent in the election, Paul Keating, used this perception to his advantage saying that debating with him "...was like being flogged with a warm lettuce" (Carroll 2010: 83).

Despite the anomalies in the trend, the general trend of cogproc is downward over the 115 year period and this is consistent with a fall in the prevalence of introspective proclivities and a rise in the prevalence of high impact personality-based campaigning.

The use of See terms is negatively associated with time, as indicated by Figure 7. The average rate of use is 0.325 per cent (sd = 0.114). There is a generally falling trend until 1983 after which the rate of use increases, although the lowest rate of use occurs in 1990:
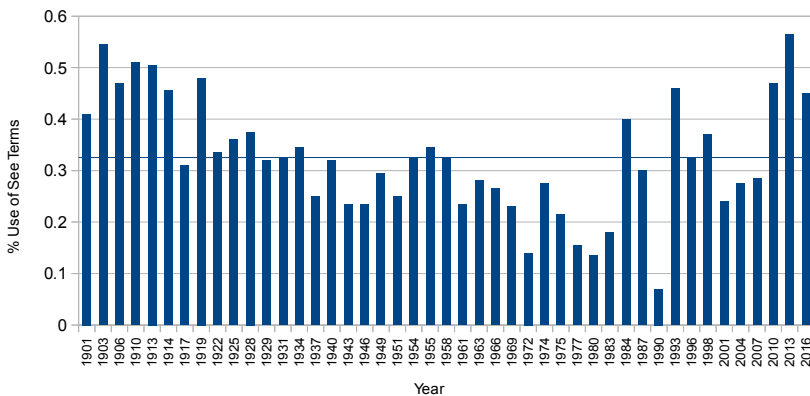

Figure 7: Per cent See terms by year

A plausible approach to explaining this trend is to consider the efficacy of using a visual as opposed to a linguistic framework to express ideas as society changed over the period 1901–2016. We have seen above that the most analytical speeches occurred in 1937 and that this level was maintained until the 1980s after which there was a general decline. If we also consider that the first several elections from 1901 showed low levels of analytical thinking we can be fairly confident in saying that the analytical thinking variable tracks the ability of the public to entertain complex ideas. If we consider the contours of the See variable in this light we could conceivably maintain that the See variable is picking

up the extent to which the population is visually as opposed to a conceptually oriented. At the time of federation there were generally low levels of substantive education and literacy and the use of visual images in speeches was therefore apposite. As education and literacy levels increased, the use of visual imagery decreased until some time between 1972 and 1980, when, with the exception of 1990, the trend reversed. The explanation for this reversal may be that television was used to its full effect at approximately this time. Television had been in Australia since the 1950s but it had not been used to its full effect for political advertising until the 1970s. We could possibly mark the turning point as the liberal ALP's 'It's Time' campaign which used television and media personalities in high-impact advertisements on television. Thus, the visual began to supersede the conceptual. The increasing extent of visual images in popular culture, education and daily life has provided a basis for political leaders eschewing the conceptual for the visual at an increasing rate to the extent that the speeches of the 2013 election had the highest levels of See terms in the series and the levels at the 2016 election were on par with those of 1914 when low education and literacy levels precluded the used of conceptual ideas in political speeches.

The Authentic index measures the extent to which the speaker/speech writer believes the content of the text. The index is based on Newman *et al.* (2003) which tracked the linguistic differences between true and false accounts of attitudes about abortion, feelings towards friends and a mock crime. The sample consisted of 188 transcriptions of videoed statements, 44 typed statements and 55 handwritten statements. It was found that veridical accounts had higher cognitive complexity, had a higher number of self and other-references and used fewer negative emotion words. The LIWC Authentic measure is based on these factors and varies from 0–100. Figure 8 shows the changes in this index by year:
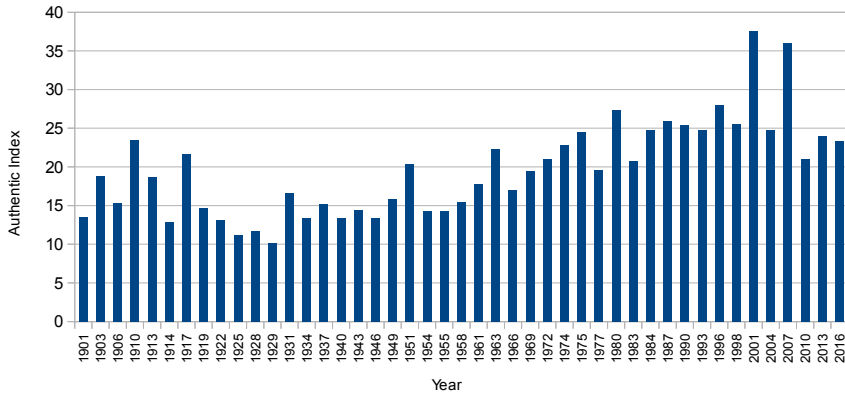
Figure 8: Authentic index scores by year

In the speech corpus, the average rate of authenticity is 19.67 (sd = 6.16). The level of authenticity rises over time indicating that the extent to which the speech writers/speakers believe the content of their speeches increases over time. Interestingly, the two highest scores on this index were 37.59 in 2001 (conservative incumbent PM John Howard= 41.81; liberal opposition leader Kim Beasley = 33.36) and 35.98 in 2007 (conservative incumbent PM John Howard = 29.29; liberal opposition leader Kevin Rudd = 42.58).

It is difficult to maintain that there should be a significant difference in the extent to which speech writers believe the content of their texts in the manner in which the index varies over time. It is more plausible that the index is picking up language change. The constituents of the index include self and other-references and negative emotion words. If we consider that social terms follow a similar contour it is plausible that a person-oriented approach to speech writing might be what the Authentic index is tracking, given that such an approach would nec-essarily involve referring to the self and others. Similarly, with increasing levels of economic prosperity over the last century it seems plausible that negative emotion words are less likely to be employed as time goes by. This is, in fact, what we find in the data: the correlation between year and negative emotion words is -.358 (p<.01; n = 88). Thus, we can conclude that the Authentic index is picking up language change rather than self-credulousness on the part of speech writers/speakers.

5   Conclusion

In this paper I have tracked the linguistic changes in election campaign speeches over the period 1901–2016. Machine learning methods were used to extract 23 important linguistic variables associated with diachronic variation. This set of variables was subject to further sifting to ensure that only those variables which can be said to have genuinely changed were subjected to close analysis.

The 6 variables analysed reveal an interesting picture of the way party leaders have chosen to represent themselves to voters over the period under consideration. In terms of the substantive content of speeches, discussions of the home have become more important, as have discussions associated with work. In terms of the style of the speeches, there has been a decline in the analytical content of speeches and speeches have been designed to appeal to voters' sense of social and personal sensibilities and have become less introspective. Furthermore, in keeping with changes in education and literacy levels, and the influence of visual culture, there has been a decrease in the use of visual imagery over most of the 20th Century with a rise in the last 40 years associated with television and visual campaigning. It will be interesting to see whether these trends continue in the years to come.

References

Brysbaert, Marc, Amy Beth Warriner and Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods* 46(3): 904–911.

Carroll, Brian. 2010. *Vintage Keating*. Rosenberg Publishing: Kenthurst, NSW.

Copland, Douglas. 1934. *Australia in the world crisis, 1929–33.* London: Cambridge University Press.

Dalvean, Michael. 2015a. Ranking contemporary American poems. *Digital Scholarship in the Humanities* 30(1): 6–19.

Dalvean, Michael. 2015b. *The axial age: A linguistic examination of cognitive evolution*. Working Paper. School of Politics and International Relations. Canberra: Australian National University. https://www.academia.edu/25063310/The_Axial_Age_A_Linguistic_Examination_of_Cognitive_Evolution. Accessed12/12/2016.

Kacewicz, Ewa, James W. Pennebaker, Matthew Davis, Moongee Jeon and Arthur C. Graesser. 2014. Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology* 33(2): 125–143.

Kincaid, J. Peter, Robert P. Fishburne Jr, Richard L. Rogers and Brad S. Chissom. 1975. *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*. No. RBR-8-75. Naval Technical Training Command Millington TN Research Branch.

Lakhal, Sawsen, Éric Frenette, Serge Sévigny and Hager Khechine. 2012. Relationship between choice of a business major type (thing-oriented versus person-oriented) and Big Five personality traits. *The International Journal of Management Education* 10(2): 88–100.

Newman, Matthew L., James W. Pennebaker, Diane S. Berry and Jane M. Richards. 2003. Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin* 29(5): 665–675.

Pennebaker, James W., Ryan L. Boyd, Kayla Jordan and Kate Blackburn. 2015. *The development and psychometric properties of LIWC2015*. University of Texas: Austin. https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf?sequence=3. Accessed 18/12/2016.

Pennebaker, James W., Cindy K. Chung, Joey Frazee, Gary M. Lavergne and David I. Beaver. 2014. When small words foretell academic success: The case of college admissions essays. *PloS one* 9(12): e115844.

Readability Studio [Computer Software]. 2012. Oleander Software Ltd, http://www.oleandersolutions.com/readabilitystudio.html Accessed 03/05/2016.

Stotsky, Sandra. 1990. Do we have, or have we had, a literary canon in our secondary schools? What literature surveys reveal. Paper presented at the Annual Meeting of the National Council of Teachers of English (80th, Atlanta, GA, November 16–21, 1990), http://files.eric.ed.gov/fulltext/ED326877.pdf. Accessed 12/12/2016.

Trope, Yaacov and Nira Liberman. 2010. Construal-level theory of psychological distance. *Psychological review* 117(2): 440–463.

Witten, Ian H., Eibe Frank, Mark A. Hall and Christopher J. Pal. 2016. *Data mining: Practical machine learning tools and techniques*. Cambridge, MA, USA: Morgan Kaufmann.