Analyzing Lexical Simplification in Interviews of Donald Trump and Joe Biden: Tracking the Trend of Simplification in Presidential Discourse Using Lexical Sophistication Indices

Woonhyung Chung†
Yonsei University

ABSTRACT

This research aimed to explore whether the observed lexical simplification in Donald Trump's interviews is an idiosyncratic trait of his linguistic style or if it transcends to his successor, Joe Biden. To achieve a nuanced analysis, the research introduced novel indices of lexical sophistication, moving beyond traditional surface-level lexical measures such as lexical density and lexical diversity. These new indices gauged the use of difficult words, taking into account academic language and psycholinguistic properties such as concreteness, familiarity, imageability, and meaningfulness. While results based on surface-level lexical indices did not reveal any discernible trend, findings derived from the lexical sophistication indices indicated that lexical simplification was not exclusive to Trump. Instead, it emerged as a trend that persisted in Biden's interviews notably in aspects such as the use of academic words, concreteness, and, to some extent, meaningfulness.

Keywords: lexical simplification, lexical sophistication, political discourses, Donald Trump, Joe Biden

1. Introduction

This study was conducted with the purpose of exploring whether the simple language of Donald Trump is limited solely to his personal linguistic traits, or it is part of an ongoing trend of lower language complexity in US presidential discourse. Trump's simple and direct rhetoric has been proven in a good deal of previous research (Ahmadian et al., 2017; Chen et al., 2019; Conway & Zubrod, 2022; Degani, 2016; Kayam, 2018; Savoy, 2018a, b; Wang & Liu, 2018), and the

* I would like to express my sincere gratitude to the three anonymous reviewers of this journal for providing valuable comments in improving on this paper.
† Corresponding author: woonka82@hanmail.net

evaluations on his language use are mixed. His simple language is widely criticized as a reflection of his single-mindedness, poverty of thought, and erraticism (Conway & Zubrod, 2022; Pullum, 2015; Savoy & Wehren, 2022). On the contrary, his conversational style accompanied by his eloquence is evaluated as one of the factors attracting the audience, thereby leading to his win in the Republican Party primaries and the 2016 presidential election (Ahmadian et al., 2017; Chen et al., 2019; Savoy, 2018a, b). Meanwhile, Trump's straightforwardness and brevity have been explained to be in line with the growing trend of anti-intellectualism that is contemporarily widespread in culture, society, and politics (Chen et al., 2019; Degani, 2016; Jordan & Pennebaker, 2017; Jordan et al., 2019; Kayam, 2018). Anti-intellectual voters tend to make their decisions based on politicians' simple messages and portrayed images rather than pondering over the policies that politicians present, often criticized as lacking reasoning and empathy; in this wave of anti-intellectualism, Trump's simple and colloquial language has rapidly gained great support from the audience (Degani, 2016; Kayam, 2018).

Trump's presidential successor, Joe Biden, is also not free from such criticism. Contrary to a general perception that Biden's political rhetoric fits into the tradition of the US President's discourse, some linguistic studies (AlAfnan, 2022; Anggoro et al., 2022; Conway & Zubrod, 2022) revealed that Biden's language deviates from the norm of political rhetoric. Indeed, Biden had much lower complexity than US Presidents for the last 50 years in presidential debates and the State of the Union Speeches; more surprisingly, in presidential debates his language is simpler than Trump's (Conway & Zubrod, 2022). Biden's use of simple language has been explained as an extension of anti-intellectualism, a reflection of America's polarized political climate that has been accompanied by divisive and simple language. It has also been seen by some as an indication of his cognitive decline (Conway & Zubrod, 2022).

Despite numerous analyses and evaluations on Trump's and Biden's simple languages, it remains inconclusive whether the simplified lexicon observed in Trump's language is his unique trait or part of a growing trend in American political rhetoric extending to his successor, Biden. As an attempt to address this issue, this study compares Donald Trump with his three predecessors (Bill Clinton, George Bush, and Barack Obama) and his immediate successor (Joe Biden) in a more sophisticated manner. In addition to lexical density and lexical diversity which are surface-level lexical measures employed in previous research (e.g., Savoy, 2018a, b; Wang & Liu, 2018), this study introduces new finer-grained lexical

sophistication indices that gauge a text's lexical difficulty: the use of academic languages, word abstractness (concreteness), word exposure (familiarity), the evocation of mental and sensory images (imageability), and the relation to other words (meaningfulness) (Crossley & McNamara, 2012). These deeper-level linguistic measures enable us to explore how advanced and sophisticated words are used in each President's discourses and to evaluate what aspects of the presidential discourse are being simplified - a dimension not adequately addressed in previous studies. Therefore, this approach offers a more conclusive answer regarding whether lexical simplicity is found in Trump and/or Biden, providing an opportunity to track a trend of simplification in American political rhetoric in a more sophisticated manner.

2. Related Work

A popular perception on Trump's simple and conversational style has been empirically confirmed by a significant body of research (Ahmadian et al., 2017; Chen et al., 2019; Chung, 2021; Conway & Zubrod, 2022; Degani, 2016; Kayam, 2018; Savoy, 2018a, b; Wang & Liu, 2018). His less use of complex words (i.e., words with more than three syllables) or big words (i.e., words with six letters or more) stood out especially in interviews and debates compared to other candidates (Kayam, 2018; Savoy, 2018a, b) or predecessors (Chung, 2021). It is partly due to his preference of one-syllable words such as *great, say, big, think*, and *take* (Degani, 2016) with a limited use of big words to the words such as *tremendous, beautiful*, and *incredible* (Chung, 2021). Trump's frequent use of informal language was also attributed to his unconventional communication style. Ahmadian et al.' analysis (2017) on informality in 2016 presidential campaign speeches found that Trump displayed informal language more frequently such as swear words, accents (e.g., *agree*, *ok*, or *yes*), non-fluencies (e.g., *er*, *hm*, or *umm*), or filler (e.g., *I mean*, or *you know*) compared to other nine Republican candidates. Furthermore, Trump's less diverse vocabulary was indicated by his low values of the type-token ratio (Savoy, 2018a, b), the moving-average type-token ratio (Wang & Liu, 2018), and the lexical density (Savoy, 2018a, b) during the 2016 presidential election. His low values indicated that he repeatedly used a narrow range of vocabulary, delivering less information and topics. Trump's low-complexity language has also been verified in his low readability. Several studies (Degani, 2016; Kayam, 2018; Wang & Liu,

2018) conducted commonly used readability tests (e.g., Flesch Kincaid Reading Ease, Flesch Kincaid Grade Level, Gunning Fog Score, Simple Measure of Gobbledygook Index, Coleman Liau Index, and Automated Readability Index), which are gauged based mainly on the numbers of characters, syllables and complex words, and mean sentence length. The results showed that Trump employed significantly low readability in interviews, TV debates, and campaign speeches compared to his opponents during the 2016 US presidential elections. Regarding his low complexity and non-standard language, the common view is that Trump is an outlier who noticeably deviated from long-standing political norms, characterized as his egocentricity, low intelligence, and lack of analytic thinking (Conway & Zubrod, 2022; Pullum, 2015; Savoy, 2018b). Meanwhile, Ahmadian et al. (2017) argued that successful politicians and leaders have tended to reduce their linguistic complexity while seeking power; in this regard, Trump's getting elected could be predictable, presenting a positive correlation between language informality and the number of states won by nine Republican candidates in the 2016 primaries. Jordan et al. (2019) demonstrated that the trend toward informal language in politics is prevalent within the United States and other English-speaking countries, and Trump's language is consistent with this trend. Degani (2016) and Kayam (2018) considered this trend of voters' shunning rhetorical complexity and inclining to simple messages as anti-intellectualism.

Contrary to general intuition, Joe Biden, who is Trump's successor, also employs simple rhetoric. Savoy & Wehren (2022) compared Trump's and Biden's styles during the 2020 US presidential election, revealing that their lexical density of TV debates and interviews was identical (43.1%) while the type-token ratio of Biden (0.360) was higher than Trump (0.331). In a comparison of the inaugural speeches of Trump and Biden, AlAfnan (2022) demonstrated that Biden's speech (30.2%) had a lower lexical density than Trump's (37.2%). AlAfnan argued that Biden's lower value may be attributed to his intention to make his speech more understandable to the American public and international observers. Conway and Zubrod (2022) showed that Trump's integrative complexity (1.51) was lower than Biden's (1.83) in inauguration speeches, while Biden's integrative complexity (1.70) was lower than Trump's (1.80) in pre-election debates. They suggested that, combined with the overall trend of simplification in the US presidential discourse, simplification has been intensified in both Trump and Biden, possibly due to extreme division in recent American politics as well as individual traits (Trump's power motive and Biden's cognitive decline).

Based on previous studies on Trump's and Biden's linguistic complexity, this research evaluates their lexical complexity through comparison with three former Presidents (Bill Clinton, George Bush, and Barack Obama). The primary goal is to ascertain if the trend toward lexical simplification has intensified in the two recent Presidents. This evaluation employs new lexical sophistication indices that have not been previously explored in political discourse. The motivation for introducing these new indices is the recognition of limitations in solely relying on surface-level measures. Prior research heavily relied on conventional surface-level measures such as lexical density and lexical diversity, which gauge the proportion of lexical words or different word types. While informative, these measures might not fully capture the nuances of the lexical simplification trend in US presidential discourse, particularly in specific aspects of word use. Thus, this study introduces five lexical sophistication indices: academic language, familiarity, concreteness, imageability, and meaningfulness. These indices offer a more nuanced examination on what aspects of their vocabulary have been simplified from more diverse angles, thereby effectively capturing the trend of simplification. This study explores two key research questions:

RQ1. Do conventional surface-level measures, such as lexical density and lexical diversity, reveal pronounced lexical simplification in Trump and/or Biden when compared to Clinton, Bush, and Obama?

RQ2. Do new lexical sophistication indices, namely academic language, familiarity, concreteness, imageability, and meaningfulness, expose the aspects of lexical simplification in Trump and/or Biden compared to Clinton, Bush, and Obama? What specific aspects of their word use are undergoing simplification?

The first research question investigates whether traditional lexical measures detect significant changes in the lexical complexity of Trump and Biden compared to their predecessors. Additionally, it explores the effectiveness of commonly used indices of lexical density and diversity in discerning a trend of lexical simplification.

The second research question delves into the newly introduced lexical sophistication indices in this study. It aims to uncover specific dimensions of lexical simplification by examining academic language, familiarity, concreteness, imageability, and meaningfulness. The goal is to provide a more detailed and

nuanced understanding of how the vocabulary of Trump and Biden differs from that of their predecessors.

3. Measures

The study focuses on the lexical complexity of presidential interviews, investigating three dimensions of lexical complexity: lexical density, lexical diversity, and lexical sophistication. In addition to lexical density and lexical diversity, which have been used in previous studies to measure the breadth of a President's word use (i.e., how many different words a President uses), this study selects and analyzes features related to lexical sophistication, measuring the depth of word use (i.e., how difficult words a President uses), which has not been explored before. In measuring lexical sophistication of US presidential political discourse, this study adopts academic languages and psycholinguistic word information. Despite their potential to capture the trend of simplification of political discourse, these indices have never been used in the analysis of political discourse. This paper attempts to measure the lexical complexity of political discourse through these new indices and seeks to track the trend of simplification in US presidential discourse in a more sophisticated manner. Details on the indices used in the analysis are as follows.

3.1. Lexical density

Lexical density measures the proportion of lexical words in a total of words compared to function words in a given text (Ure, 1971). The index for lexical density ranges from 0 to 1, with a higher value indicating a larger vocabulary size. Function words consist of closed-class grammatical words, which include determiners (e.g., *a*, *the*), pronouns (e.g., *he*, *she*, *we*), prepositions (e.g., *at*, *in*, *on*), and conjunctions (e.g., *and*, *but*, *or*). In contrast, lexical words are generally considered open-class words such as nouns, verbs, adjectives, and adverbs. When a text has a greater use of lexical words, it means that a wider range of open-class words is employed in the text. This suggests that the information and content within the text are richer and more complex.

3.2. Lexical diversity

Lexical diversity refers to the proportion of unique words (i.e., word types) in

a text compared to the total number of running words (i.e., word tokens) (Templin, 1957). It is measured on a scale ranging from 0 to 1, where a higher value indicates a broader vocabulary. Traditional measures of lexical diversity (e.g., type/token ratio (TTR), the number of different words (NDW)) are dependent on text length, raising the issue of distortion that a longer text tends to yield a lower lexical diversity value than a shorter one. To address this issue, more developed measurements for lexical diversity have been devised, and one well-known index is the moving-average type-token ratio (MATTR; Covington & McFall, 2010), which remains stable from text length.[1] MATTR calculates the average TTR of segments that are cut by a smoothly moving window. The size of the moving window is arbitrarily set (e.g., MATTR50 has a 50-word moving window) and the window moves through a given text one word at a time, cutting the text in an overlapping manner (e.g., MATTR50 cuts a given text into words 1-50, words 2-51, words 3-52, etc.) This method ensures the MATTR values are not sensitive to text length or the accidental determination of segment boundaries.

3.3. Lexical sophistication

Lexical sophistication is distinct from lexical density and lexical diversity. Lexical density and lexical diversity primarily measure the breadth of word use in a text and they are relatively simple in that (a) they use simple algorithms (e.g., the proportion of lexical words, the percentage of the number of word types divided by the words tokens), (b) data are derived from the texts themselves, and (c) its high correlation with texts' qualities are firmly proved (Salsbury et al., 2011). While these measures are useful, they do not present important qualitative information on depth of word use, specifically, the word difficulty of a text (Vermeer, 2000; Daller et al., 2003; Salsbury et al., 2011). On the other hand, lexical sophistication focuses on assessing the difficulty of words used in a text to delve deeper into the quality of the text. Unlike lexical density and diversity which are calculated from the text itself, features related to lexical sophistication in this study[2] are measured using

1) Besides MATTR, there are several variations of TTR including Mean segmental TTR (MSTTR; Johnson, 1944), Corrected TTR (CTTR; Carroll, 1964), Root TTR (RTTR; Guiraud, 1960) Bilogarithmic TTR (LogTTR; Herdan, 1964), the Uber Index (Dugast, 1979), the D measure (Malvern et al., 2004; Mckee et al., 2000).

2) While some lexical sophistication measures, such as word frequency and word range, are considered relatively "shallow" (Vajjala & Meurers, 2012), this study opts for more in-depth measures in accessing lexical difficulties: academic language and psycholinguistic word information.

external sources (e.g., Academic Word List (Coxhead, 2000) for investigating academic language, Medical Research Council Psycholinguistic Database (Wilson, 1988) for exploring psycholinguistic word information). The premises and calculations of each measure are outlined below.

3.3.1. Proportion of academic words

The use of academic words contributes to measuring the lexical difficulty and quality of a text. The premise behind academic word measurement is that a greater number of academic words indicate a higher level of a text. Computational tools provide an automatic measure of the proportions of academic words in a text by finding the words that correspond to words in academic lists. The Academic Word List (AWL; Coxhead, 2000), created from academic corpora of journal articles and textbook chapters with 28 subject areas in four broad disciplines, is one of the representative academic lists with high coverage that have been widely used in measuring texts' academic language.

3.3.2. Psycholinguistic scores

The psycholinguistic scores are determined based on psycholinguistic properties of words that affect lexical processing (Salsbury et al., 2011). The premise behind psycholinguistic scores is that higher cognitive demands for word retrieval or process indicate higher word difficulties. There are several properties related to word processing, such as word form and word frequency, but properties related to word meanings is of concern in the current study, so four commonly studied attributes related to meaning-based word difficulties have been chosen: concreteness, meaningfulness, familiarity, and imageability. These properties are evaluated by the difficulty of words in terms of how easy it is to access core items and how salient a word is. Medical Research Council (MRC) Psycholinguistic Database (Wilson, 1988)[3] provides psycholinguistic scores of each attribute determined by human judgements; a word that requires lower processing demand is scored higher. Computational tools automatically measure psycholinguistic scores by identifying all words that correspond to words in MRC Psycholinguistic Database and by dividing the combined scores of the words by the total number of the words that are given the scores.

3) Besides MRC psycholinguistic database are several widely used eternal sources such as Brysbaert et al. (2014) and Kuperman et al. (2012).

3.3.2.1. Concreteness scores

Concreteness of a word refers to how concrete or abstract a word is. Concrete words can be described simply by pointing to an object, material, or person (e.g., *apple, milk, doctor*) (Kyle & Crossley, 2015), requiring less cognitive burden in processing than abstract words. Being located on the right side of a scale that presents the degree of concreteness, concrete words are scored higher. In contrast, abstract words need other words to explain themselves (e.g., *impossible, aspect*), receiving a lower score, being posited on the opposite side (Kyle & Crossley, 2015).

3.3.2.2. Imageability scores

Imageability of a word refers to how quickly and easily a word can evoke a mental and sensory image. Highly imageable words such as *buffalo* and *beautiful* readily conjure the images, receiving higher imageability scores. In contrast, less imageable words such as *relevant* and *philology* require greater cognitive effort for the construction of mental pictures, having lower scores (Salsbury et al., 2011).

Some words can have high scores in both concreteness and imageability, while others do not (Salsbury et al., 2011). For example, the word *apple* has high concreteness and imageability scores because it refers to a tangible object and allows a quick and easy access to its mental representation. However, the word *marriage* does not point to a specific object but can be easily visualized as a mental image, making its concreteness score lower than its imageability score (Salsbury et al., 2011).

3.3.2.3. Familiarity scores

Familiarity of a word is defined as how commonly a word is exposed (Eguchi & Kyle, 2020). Words that are frequently experienced, such as *breakfast*, *girl*, and *paper*, are scored high, whereas words that occur in a limited context such as *sultan*, *buffoon*, and *puck*, are scored low (Eguchi & Kyle, 2020). This attribute has a strong correlation with word frequency (Kim at al., 2018).

3.3.2.4. Meaningfulness scores

Meaningfulness of a word is based on how associated a word is to other words (Kyle & Crossley, 2015). Words that have a high association with other words,

such as *food*, *music*, and *people*, require less cognitive demand because the words have more semantic links with other words that facilitate the organization and storage of the words (Crossley et al., 2011), leading to high meaningfulness scores. In contrast, words that have a weak association with other words such as *adze*, *brisket*, and *amorphous*, have fewer semantic links with other words, requiring more cognitive demands and having lower meaningfulness scores.

4. Method

4.1. Corpus collection

This study applies the new language analysis indices to the interview genre, which is characterized as spontaneous and less affected by speechwriters or prepared scripts, thereby revealing the characteristics of a political figure more clearly. Media interviews of five Presidents during their presidency were collected. The corpus was constructed by collecting transcripts either from the websites of broadcasting companies such as CNN, NPR, and MSNBC, or from the American Presidency Project website (www.presidency.ucsb.edu) that provides various types of US Presidents' and presidential candidates' discourse including media interviews. The number of interview texts per each President were 16 except for 14 of Biden, for a total of 78 texts, comprising 290,490 words (See Table 1).

Table 1. Composition of the corpus

	Interview years	# of texts	# of tokens	Mean length of each text	Standard deviation
Clinton	2000-2001	16	73,084	4,567.750	1,464.994
Bush	2007-2008	16	58,315	3,644.688	1,444.968
Obama	2016-2017	16	59,154	3,697.125	1,612.367
Trump	2017-2018	16	48,035	3,002.188	1,968.629
Biden	2021-2022	14	51,902	3,707.286	3,220.732
Total number		78	290,490		

Table 1 illustrates interview years, the number of interview texts, and the total number of tokens per each President, as well as the mean length of text per President and their standard deviations. The reason why the number of Biden's

interview texts (14) is fewer than the other Presidents (16 respectively) is that Biden has not completed his term so there are fewer complete transcripts available than the others. As complete interview texts were used, the difference in the total number of tokens per each President was unavoidable; however, the one-way ANOVA did not reveal any significant differences in the numbers of tokens of texts per President ($F(4, 73) = 1.226$, $p = .307$, $\eta^2 = 0.063$). Parts from interviewers and other interviewees, and descriptions of audience's reactions such as "applause" and "laughter" were removed.

4.2. Data analysis

This study used two computational automatic analysis tools to measure three dimensions of lexical features - lexical density, lexical diversity, and sophistication - among five former and current US Presidents: the Tool for the Automatic Analysis of LExical Diversity (TAALED) version 1.3.1 (Kyle et al., 2021) and the Tool for the Automatic Analysis of LExical Sophistication (TAALES) version 2.2 (Kyle et al., 2018). These tools exhibit several advanced features, including (a) fast processing speeds, (b) user-friendly interfaces, (c) compatibility with most operating systems (Windows, Mac, and Linux), (d) batch processing capabilities for .txt files, and (e) incorporation of newly developed indices (Kyle & Crossley, 2015). In this study, TAALED 1.3.1 measured lexical density values and lexical diversity values. Lexical sophistication values were achieved using TAALES 2.2. These computational tools automatically calculated the proportions of lexical words in each text (for achieving texts' lexical density), the values of MATTR50 (the number 50 after MATTR means the length of each segment is set to 50 words) (for gauging texts' lexical diversity), and the scores of academic languages, word concreteness, imageability, meaningfulness, and familiarity (for measuring texts' lexical sophistication). To identify statistical significance among each value, a one-way analysis of variance (ANOVA) test was conducted using JASP (JASP Team, 2022) version 0.16.4.

5. Results

This study aimed to evaluate whether Trump's and Biden's language tended to be lexically simplified compared to the predecessors (Clinton, Bush, and Obama)

in order to determine if the trend of lexical simplification has intensified over the past decade. In addition to traditional indices used in political discourses, such as lexical density and lexical diversity, this study adopted indices of lexical sophistication, including the proportion of academic words and psycholinguistic scores. The objective was to discern whether Trump and Biden differ from their predecessors in terms of their use of simpler lexicon.

5.1. Lexical density

The descriptive statistics showed that Obama had the highest value (M = 0.457, SD = 0.013) and Clinton had the lowest value (M = 0.421, SD = 0.017) in lexical density (See Figure 1 and Appendix 1). The one-way ANOVA revealed a significant difference among the five Presidents, $F(4, 73) = 13.011$, $p < .001$, $\eta^2 = 0.416$. Post-hoc tests (Tukey) showed that Obama was significantly higher than the other four Presidents, with the rest four Presidents showing no significant difference from each other (See Table 2). The results for lexical density indicated that Obama had a higher lexical density than the other Presidents, but neither Trump nor Biden were found to be distinct from predecessors other than Obama in terms of lexical density.

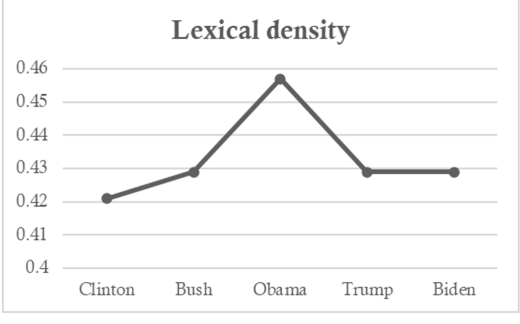

Figure 1. Lexical density of Presidents

Table 2. Post-hoc comparisons of lexical density

		Mean difference	SE	*t*	p_{tukey}
Trump	Biden	-1.749×10^{-4}	0.006	-0.031	1.000
	Obama	-0.028	0.005	-5.163	< .001
	Bush	1.142×10^{-4}	0.005	0.021	1.000
	Clinton	0.008	0.005	1.472	0.584
Biden	Obama	-0.028	0.006	-4.957	< .001
	Bush	2.890×10^{-4}	0.006	0.051	1.000
	Clinton	0.008	0.006	1.453	0.596
Obama	Bush	0.028	0.005	5.184	< .001
	Clinton	0.036	0.005	6.635	< .001
Bush	Clinton	0.008	0.005	1.451	0.597

5.2. Lexical diversity

The descriptive statistics showed that Obama had the highest value (M = 0.783, SD = 0.015) and Trump had the lowest value (M = 0.696, SD = 0.014) in the MATTR50 (See Figure 2 and Appendix 1). The one-way ANOVA revealed a significant difference among the five Presidents $F(4, 73) = 84.467$, $p < .001$, $\eta^2 = 0.822$. Post-hoc tests (Tukey) showed that Obama was significantly higher than the other four Presidents, and Trump was significantly lower than the other four Presidents, with the rest three Presidents showing no significant difference from each other (See Table 3). The results indicated that Obama displayed the greatest variety of words and Trump used the least variety of words. The diminished lexical diversity observed in Trump did not manifest in Biden.

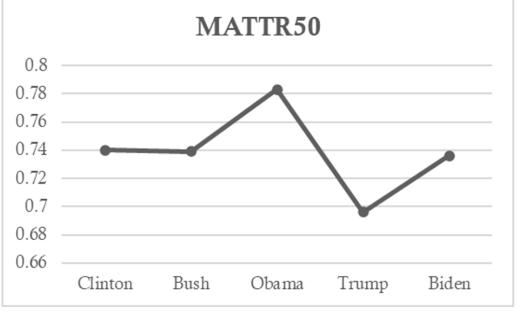

Figure 2. MATTR50 of Presidents

Table 3. Post-hoc comparisons of MATTR50

		Mean difference	SE	*t*	p_{tukey}
Trump	Biden	-0.039	0.005	-8.047	< .001
	Obama	-0.087	0.005	-18.352	< .001
	Bush	-0.043	0.005	-9.081	< .001
	Clinton	-0.044	0.005	-9.282	< .001
Biden	Obama	-0.047	0.005	-9.683	< .001
	Bush	-0.004	0.005	-0.726	0.950
	Clinton	-0.005	0.005	-0.921	0.888
Obama	Bush	0.044	0.005	9.271	< .001
	Clinton	0.043	0.005	9.070	< .001
Bush	Clinton	-9.508×10^{-4}	0.005	-0.201	1.000

5.3. Lexical sophistication

Besides gauging lexical density and lexical diversity, this study adopted new lexical sophistication indices to assess the trend of lexical sophistication at a deeper level. The proportion of academic words and four kinds of psycholinguistic scores were measured.

5.3.1. Proportion of academic words

The descriptive statistics showed that Obama used the highest proportion of academic words (M = 0.044, SD = 0.009) and Trump used the lowest proportion (M = 0.019, SD = 0.005) (See Figure 3 and Appendix 1). The one-way ANOVA revealed a significant difference among the five Presidents in the proportions of academic words, $F(4, 73) = 35.278$, $p < .001$, $\eta^2 = 0.659$. Post-hoc tests (Tukey) showed that proportions of Obama, Bush, and Clinton were not significantly different from each other, and the three Presidents had higher proportions of academic words than Trump and Biden, while Biden used significantly greater academic words than Trump (See Table 4). These results indicated that although Biden used more academic words than Trump, both Trump and Biden were significantly distinguished from the three predecessors in the use of academic words, which raised the possibility of lexical simplification trend found in Trump and Biden.

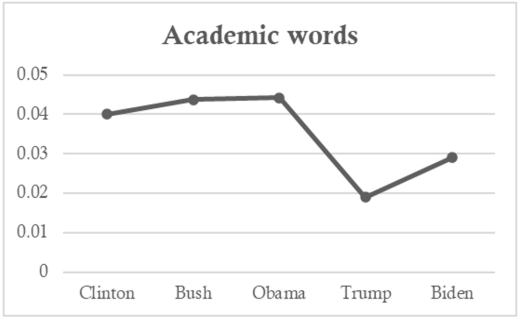

Figure 3. MATTR50 of Presidents

Table 4. Post-hoc comparisons of proportion of academic words

		Mean difference	SE	*t*	p_{tukey}
Trump	Biden	-0.009	0.003	-3.456	0.008
	Obama	-0.025	0.003	-9.676	< .001
	Bush	-0.024	0.003	-9.483	< .001
	Clinton	-0.021	0.003	-7.999	< .001
Biden	Obama	-0.016	0.003	-5.892	< .001
	Bush	-0.015	0.003	-5.705	< .001
	Clinton	-0.011	0.003	-4.272	< .001
Obama	Bush	4.984×10^{-4}	0.003	0.194	1.000
	Clinton	0.004	0.003	1.677	0.454
Bush	Clinton	0.004	0.003	1.483	0.577

5.3.2. Psycholinguistic scores

Along with the proportions of academic words, four types of psycholinguistic scores - concreteness, imageability, familiarity, and meaningfulness scores - were measured to examine the depth of vocabulary use. The higher psycholinguistic scores, the less sophisticated the language.

5.3.2.1. Concreteness scores

The descriptive statistics showed that Obama achieved the lowest concreteness score (M = 290.965, SD = 4.767) and Trump gained the highest score (M = 302.222, SD = 3.752) (See Figure 4 and Appendix 1). The one-way ANOVA revealed a significant difference among the five Presidents in their concreteness scores, $F(4, 73) = 18.868$, $p < .001$, $\eta^2 = 0.508$. Post-hoc tests (Tukey) showed that the concreteness

scores of Obama, Bush, and Clinton were not significantly different from each other, and the three Presidents had lower scores than Trump and Biden, while Trump and Biden did not show the significant difference (See Table 5). These results indicated that Trump and Biden were significantly distinguished from the three predecessors in that the two used more concreteness words, raising the possibility of lexical simplification trend found in Trump and Biden.

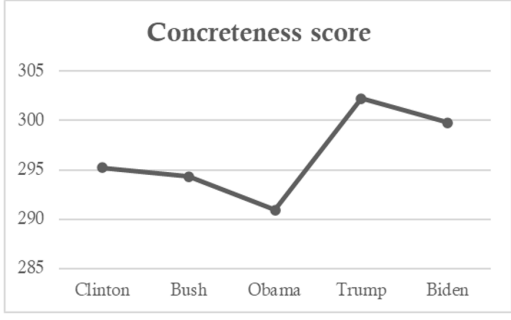

Figure 4. Concreteness scores of Presidents

Table 5. Post-hoc comparisons of concreteness score

		Mean difference	SE	*t*	p_{tukey}
Trump	Biden	2.430	1.498	1.622	0.488
	Obama	11.257	1.447	7.777	< .001
	Bush	7.869	1.447	5.436	< .001
	Clinton	6.995	1.447	4.833	< .001
Biden	Obama	8.827	1.498	5.891	< .001
	Bush	5.439	1.498	3.630	0.005
	Clinton	4.565	1.498	3.047	0.026
Obama	Bush	-3.388	1.447	-2.341	0.144
	Clinton	-4.262	1.447	-2.945	0.034
Bush	Clinton	-0.874	1.447	-0.604	0.974

5.3.2.2. Imageability scores

The descriptive statistics showed that Obama achieved the lowest imageability score (M = 314.012, SD = 4.307) and Trump gained the highest score (M = 325.310, SD = 3.440) (See Figure 5 and Appendix 1). The one-way ANOVA revealed a significant difference among the five Presidents in their imageability scores, $F(4, 73) = 16.740$, $p < .001$, $\eta^2 = 0.478$. Post-hoc tests (Tukey) showed that the imageability

scores of Obama, Bush, and Clinton were not significantly different from each other, Trump was significantly higher than those of the other four Presidents, and Biden was not significantly different from Bush and Clinton, having a significantly higher score than Obama (See Table 6). These results indicated that Trump displayed more words easily accessed to their mental pictures compared to the other four predecessors, but it was difficult to ascertain the presence of a lexical simplification trend in Trump and Biden based on the use of imageable words.

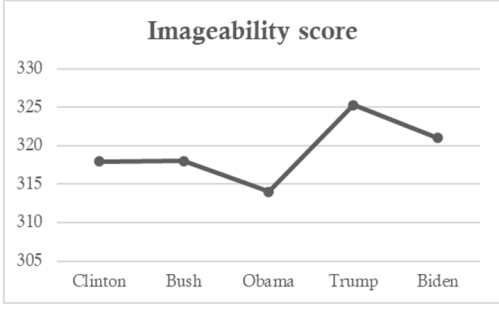

Figure 5. Imageability scores of Presidents

Table 6. Post-hoc comparisons of imageability score

		Mean difference	SE	*t*	p_{tukey}
Trump	Biden	4.297	1.497	2.871	0.041
	Obama	11.297	1.446	7.812	< .001
	Bush	7.279	1.446	5.034	< .001
	Clinton	7.348	1.446	5.081	< .001
Biden	Obama	7.000	1.497	4.677	< .001
	Bush	2.982	1.497	1.992	0.280
	Clinton	3.051	1.497	2.038	0.258
Obama	Bush	-4.018	1.446	-2.779	0.052
	Clinton	-3.949	1.446	-2.731	0.059
Bush	Clinton	0.069	1.446	0.048	1.000

5.3.2.3. Familiarity scores

The descriptive statistics showed that Obama achieved the lowest familiarity score (M = 595.826, SD = 0.972) and Trump gained the highest score (M = 597.931, SD = 1.266) (See Figure 6 and Appendix 1). The one-way ANOVA revealed a significant difference among the five Presidents in their familiarity scores, $F(4, 73) = 17.657$, $p < .001$, $\eta^2 = 0.445$. Post-hoc tests (Tukey) showed that

Trump had a significantly lower score than the other four Presidents, with the four showing no significant difference from each other (See Table 7). The results indicated that Trump used more familiar words than his predecessors, but no lexical simplification trend was found from the aspect of the use of familiar words.

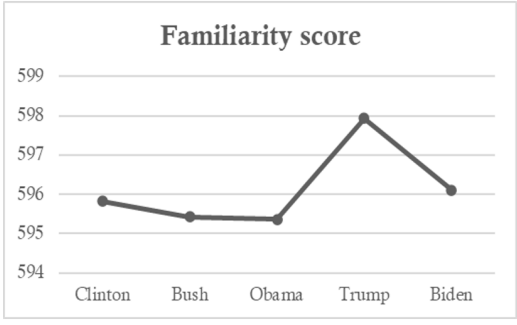

Figure 6. Familiarity scores of Presidents

Table 7. Post-hoc comparisons of familiarity score

		Mean difference	SE	*t*	p_{tukey}
Trump	Biden	1.824	0.402	4.533	< .001
	Obama	2.565	0.389	6.599	< .001
	Bush	2.504	0.389	6.443	< .001
	Clinton	2.105	0.389	5.416	< .001
Biden	Obama	0.741	0.402	1.842	0.358
	Bush	0.680	0.402	1.691	0.446
	Clinton	0.282	0.402	0.700	0.956
Obama	Bush	-0.061	0.389	-0.156	1.000
	Clinton	-0.460	0.389	-1.182	0.761
Bush	Clinton	-0.399	0.389	-1.026	0.843

5.3.2.4. Meaningfulness scores

The descriptive statistics showed that Obama achieved the lowest meaning score (M = 347.697, SD = 4.628) and Trump gained the highest score (M = 363.122, SD = 4.057) (See Figure 7 and Appendix 1). The one-way ANOVA revealed a significant difference among the five Presidents in their familiarity scores, $F(4, 73) = 28.224$, $p < .001$, $\eta^2 = 0.607$. Post-hoc tests (Tukey) showed that Obama had a significantly lower score than Trump, Biden, and Clinton; Trump had a significantly higher score than the other four Presidents; Biden had a significantly higher score than Obama and Bush (See Table 8). The results indicated that Trump employed a higher

frequency of words which are related to other words than his predecessors. Additionally, Biden demonstrated a tendency to utilize words related to other words than Obama and Bush. This observation may be indicative of lexical simplification in Trump, and to a lesser extent, in Biden as well.

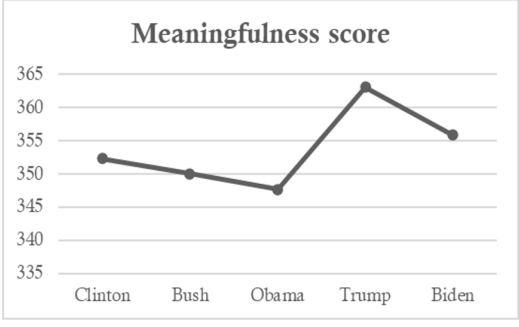

Figure 7. Meaningfulness scores of Presidents

Table 8. Post-hoc comparisons of meaningfulness score

		Mean difference	SE	*t*	p_{tukey}
Trump	Biden	7.245	1.655	4.377	< .001
	Obama	15.425	1.599	9.646	< .001
	Bush	13.096	1.599	8.190	< .001
	Clinton	10.800	1.599	6.754	< .001
Biden	Obama	8.180	1.655	4.942	< .001
	Bush	5.851	1.655	3.535	0.006
	Clinton	3.555	1.655	2.148	0.212
Obama	Bush	-2.330	1.599	-1.457	0.593
	Clinton	-4.626	1.599	-2.893	0.039
Bush	Clinton	-2.296	1.599	-1.436	0.607

6. Discussion and Conclusion

This research was conducted to investigate whether the simple lexicon observed in Trump's discourse was limited to Trump himself or it exhibits a broader trend as the lexical simplification continues with Biden. Furthermore, if such a trend is identified, the research aimed to determine whether the lexical trait shared by the two Presidents set them apart from their three predecessors (Clinton, Bush, and Obama). To make a precise assessment, this research introduced lexical sophistication indices in addition to the surface-level features of vocabulary, such

as lexical density (i.e., proportion of content words) and lexical diversity (i.e., number of word types). The new indices measure the use of difficult words based on academic language and a word's psycholinguistic properties (i.e., concreteness, familiarity, imageability, and meaningfulness). These new indices enable a more detailed examination of which aspects of words have been simplified, going beyond the evaluation of overall simplification at a surface level.

The exploration of the first research question, employing two conventional lexical measures, did not reveal a clear trend. When examining lexical density, Obama had a higher value than the other four Presidents, while the remaining four did not significantly differ from each other. In terms of lexical diversity, the results showed that Obama had a higher value than the other Presidents, and Trump had a lower value compared to the others, with no significant difference among the remaining three. These findings may suggest that lexical simplification was not pronounced in Biden (and Trump, in terms of lexical density) compared to the three predecessors, but it could imply that these conventional indices may not accurately reflect the occurrence of lexical simplification.

With this awareness of the issue, the study sought to explore the second research question. Analyses utilizing lexical sophistication indices captured a partial trend toward lexical simplification. Notably, both Trump and Biden employed fewer academic words and a greater number of concrete words when compared to the other three Presidents. In terms of academic language, Biden used fewer academic words compared to his predecessors, showing a greater use of them compared to Trump. The concreteness scores did not significantly differ between Trump and Biden. On the other hand, regarding imageability, familiarity, and meaningfulness scores, Trump scored significantly higher than the other Presidents. Biden was higher than Obama and Bush solely in the meaningfulness score.

These results can be interpreted in connection with previous studies. To begin with, the findings regarding lexical density, one of the simple lexical measures, indicated that there was no statistically significant difference among the four Presidents, except for Obama. This aligns with Jordan et al. (2019), who concluded that Trump's language use, rather than being an outlier, did not significantly deviate from the overall simplification trend in English-speaking countries. On the other hand, when considering another simple lexical measure, lexical diversity, Trump was found to use a less diverse vocabulary than the other Presidents, indicating that he repeatedly used a narrow range of words. This finding supports the general perception and prior research on Trump's notably simple rhetoric

(Ahmadian et al., 2017; Chen et al., 2019; Degani, 2016; Kayam, 2018; Savoy, 2018a, b; Wang & Liu, 2018).

To offer a more precise explanation of these mixed interpretations, this study incorporated finer-grained lexical sophistication indices, including academic language, concreteness, familiarity, imageability, and meaningfulness. Across all of these five lexical sophistication indices, Trump's interviews exhibited a simplified lexicon. Notably, Biden also used academic words less frequently, employed more concrete words, and, to some extent, used more words closely related to other words. These results indicated that lexical simplification is not solely limited to Trump; rather, the trend toward lexical simplification also appears in Biden in some aspects of word use, such as academic language, concreteness, and partially, meaningfulness.

Several factors may have contributed to the trend of lexical simplification observed in both Trump and Biden. As noted by Degani (2016) and Kayam (2018), it is possible that while anti-intellectualism has been on the rise in politics over the past decade, it could be more directly reflected in the use of non-academic language and concrete words. Furthermore, there is a possibility that the severe polarization in American politics over the past decade may have influenced the lexical simplification of both Presidents (Conway & Zubrod, 2022), leading to their preference for non-academic and concrete words. On the other hand, the lexical simplification of the two Presidents can be attributed to their personal traits. As frequently mentioned, it is conceivable that Trump's intellectual ignorance and business-oriented persona have influenced his less use of difficult words. Conversely, Trump's word choice could be a deliberate strategy to gain power, as power-oriented leaders often employ simpler language for this purpose (Ahmadian et al., 2017; Conway & Zubrod, 2022). In contrast, Biden's word choices may be influenced by his cognitive decline (Conway & Zubrod, 2022), which could be preventing him from using more academic and abstract words.

While the data cannot definitively answer the question of why lexical simplification is observed in both Trump and Biden, this study holds significance in introducing new indices for analyzing political discourse to offer a more precise evaluation of the word use changes in US Presidents, which past research has not suggested before. To arrive at a more definitive conclusion, further research is necessary to investigate whether this trend of lexical simplification continues with future US Presidents and other politicians. The new indices that this study adopted should be used so that research can provide an accurate assessment of this linguistic

trend in US political discourses.

References

Ahmadian, S., Azarshahi, S., & Paulhus, D. L. (2017). Explaining Donald Trump via communication style: Grandiosity, informality, and dynamism. *Personality and Individual Differences*, *107*, 49-53.

AlAfnan, M. A. (2022). Public discourse: Systemic functional analysis of Trump's and Biden's inaugural speeches. *Journal of Language and Linguistic Studies*, *18*(1), 1-14.

Anggoro, A. R., Mukhrizal, M., & Sufiyandi, S. (2022). A syntactical analysis on sentence structures spoken by Joe Biden and Donald Trump in the election night speeches. *Journal of English Education and Teaching*, *6*(2), 188-206.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904-911.

Carroll, J. B. (1964). *Language and thought*. Englewood Cliffs, NJ: Prentice-Hall.

Chen, X., Yan, Y., & Hu, J. (2019). A corpus-based study of Hillary Clinton's and Donald Trump's linguistic styles. *International Journal of English Linguistics*, *9*(3), 13-22.

Chung, W. (2021). Linguistic features of Donald Trump's political discourse: Focusing on genre differences and change over time. *Language Facts and Perspectives*, *52*, 61-102.

Conway III, L. G., & Zubrod, A. (2022). Are US Presidents becoming less rhetorically complex? Evaluating the integrative complexity of Joe Biden and Donald Trump in historical context. *Journal of Language and Social Psychology*, *41*(5), 613-625.

Covington, M. A., & McFall, J. D. (2010). Cutting the gordian knot: The moving-average type-token ratio (MATTR). *Journal of Quantitative Linguistics*, *17*(2), 94-100.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238.

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, *35*(2), 115-135.

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). Predicting

lexical proficiency in language learner texts using computational indices. *Language Testing*, *28*(4), 561-580.

Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, *24*(2), 197-222.

Degani, M. (2016). Endangered intellect: A case study of Clinton vs. Trump campaign discourse. *Iperstoria - Testi Letterature Linguaggi*, *8*, 131-145.

Dugast, D. (1979). *Vocabulaire et stylistique. I Théâtre et dialogue [Vocabulary and style. Vol. 1 Theatre and dialogue]*. Geneva, Switzerland: Slatkine-Champion.

Eguchi, M., & Kyle, K. (2020). Continuing to explore the multidimensional nature of lexical sophistication: The case of oral proficiency interviews. *The Modern Language Journal*, *104*(2), 381-400.

Guiraud, P. (1960). *Problèmes et méthodes de la statis-tique linguistique [Problems and methods of statistical linguistics]*. Dordrecht, The Netherlands: D. Reidel.

Herdan, G. (1964). *Quantitative linguistics*. London: Butterworths.

JASP Team (2022). JASP (Version 0.16.4) [Computer software].

Johnson, W. (1944). Studies in language behavior: I. A program of research. *Psychological Monographs*, *56*, 1-15.

Jordan, K. N., & Pennebaker, J. W. (2017). The exception or the rule: Using words to assess analytic thinking, Donald Trump, and the American presidency. *Translational Issues in Psychological Science*, *3*(3), 312.

Jordan, K. N., Sterling, J., Pennebaker, J. W., & Boyd, R. L. (2019). Examining long-term trends in politics and culture through language of political leaders and cultural institutions. *Proceedings of the National Academy of Sciences*, *116*(9), 3476-3481.

Kayam, O. (2018). The readability and simplicity of Donald Trump's language. *Political Studies Review*, *16*(1), 73-88.

Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development, and writing quality. *The Modern Language Journal*, *102*(1), 120-141.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978-990.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, *49*(4), 757-786.

Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity using direct judgements. *Language Assessment Quarterly*, *18*(2), 154-170. https://doi.org/10.1080/15434303.2020.1844205

Kyle, K., Crossley, S. A., & Berger, C. (2018). The tool for the analysis of lexical sophistication (TAALES): Version 2.0. *Behavior Research Methods*, *50*(3), 1030-1046. https://doi.org/10.3758/s13428-017-0924-4

McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, *15*, 323-337.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Houndmills, England: Palgrave MacMillan.

Pullum, G. (2015). Trump's aphasia. Language Log. Available at: http://languagelog.ldc.upenn.edu/nll/?p=20490 (accessed 3 September 2015).

Salsbury, T., Crossley, S. A., & McNamara, D. S. (2011). Psycholinguistic word information in second language oral discourse. *Second Language Research*, *27*(3), 343-360.

Savoy, J. (2018a). Analysis of the style and the rhetoric of the 2016 US presidential primaries. *Digital Scholarship in the Humanities*, *33*(1), 143-159.

Savoy, J. (2018b). Trump's and Clinton's style and rhetoric during the 2016 presidential election. *Journal of Quantitative Linguistics*, *25*(2), 168-189.

Savoy, J., & Wehren, M. (2022). Trump's and Biden's styles during the 2020 US presidential election. *Digital Scholarship in the Humanities*, *37*(1), 229-241.

Templin, M. (1957). *Certain language skills in children: Their development and interrelationships*. Minneapolis: The University of Minnesota Press.

Ure, J. (1971). *Lexical density: A computational technique and some findings.* In M. Coultard (Ed.), *Talking about text* (pp. 27-48). Birmingham, England: English Language Research, University of Birmingham.

Vajjala, S., & Meurers, D. (2012). On improving the accuracy of readability classification using insights from second language acquisition. *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP* (pp. 163-173).

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, *17*(1), 65-83.

Wang, Y., & Liu, H. (2018). Is Trump always rambling like a fourth-grade student? An analysis of stylistic features of Donald Trump's political discourse during the 2016 election. *Discourse & Society*, *29*(3), 299-323.

Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, *20*(1), 6-10.

Woonhyung Chung
Doctoral Student
Department of English Linguistics and Language
Yonsei University
50, Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea
E-mail: woonka82@hanmail.net

Received: October 30, 2023
Revised version received: December 10, 2023
Accepted: December 14, 2023

Appendix 1. Descriptive statistics

Presidents	Number of texts	Lexical density		MATTR50		Academic words	
		Mean	Std. deviation	Mean	Std. deviation	Mean	Std. deviation
Clinton	16	0.421	0.017	0.740	0.011	0.040	0.008
Bush	16	0.429	0.015	0.739	0.014	0.044	0.008
Obama	16	0.457	0.013	0.783	0.015	0.044	0.009
Trump	16	0.429	0.012	0.696	0.014	0.019	0.005
Biden	14	0.429	0.019	0.736	0.012	0.029	0.004

Presidents	Concreteness		Familiarity		Imageability		Meaningfulness	
	Mean	Std. deviation	Mean	Std. deviation	Mean	Std. deviation	Mean	Std. deviation
Clinton	295.227	4.200	595.826	1.119	317.961	4.349	352.322	4.725
Bush	294.353	3.998	595.427	1.097	318.030	4.508	350.026	5.472
Obama	290.965	4.767	595.366	0.972	314.012	4.307	347.697	4.628
Trump	302.222	3.752	597.931	1.266	325.310	3.440	363.122	4.057
Biden	299.792	3.586	596.107	1.007	321.013	3.688	355.877	3.295