Adjective Ordering in Presidential Speeches and Debates: Testing Subjectivity-Based Preferences

LING 8803 Spring 2024 Final Project

Grace Driskill

Introduction

When considering the ordering of multiple adjectives in a noun phrases, speakers have typical preferences for what sounds correct. For example, "the small white car" sounds more natural that "the white small car." This project explores if adjective ordering preferences are followed to the same extent in US presidential speeches and debates as they are in general language. The subjectivity-based preference suggested by Scontras et. al. (2017) is tested in a corpus of US presidential speeches and US presidential candidate debates. These adjective ordering preferences are due to the speaker wanting the listener to understand the referent object (Scontras et. al. 2019), but the persuasive nature of political contexts brings to question if the speaker is still trying to help the listener. Only a weak correlation was found between adjective subjectivity and distance from the noun. However, in instances where subjectivity scores for both the adjectives in the phrases were available, examples of the ordering not following the subjectivity were rare.

Related Works

The adjective ordering preferences of having the less subjective adjectives appear closer to the noun was experimentally validated by Scontras et. al. (2017). This has been explained by the hierarchical structure of modified nouns, closer adjectives modify the noun earlier. Placing the less subjective, more meaningful adjectives closer to the noun allows the speaker to narrow down the search space and help the listener understand the referent object (Scontras et. al. 2019). This is related to the rational speech acts (RSA) framework, the process of recursive reasoning between a speaker and listener (Goodman & Frank 2016).

Adjective ordering preferences are thought to be a cross-linguistic feature, and the subjectivity-based preference has been found in other languages (Samonte & Scontras 2019, Kachakeche & Scontras 2020). Other explanations of adjectivity ordering also exist, such as definiteness of denotation (Martin 1969) or multifactorial explanations (Wulff 2003). The subjectivity-based theory was chosen as the basis for this project because it was an assigned reading and related back to course material.

The extent that adjective ordering preferences are followed has been questioned. Kotowski and Hartl (2019) argue these are preferences, not strict grammar rules. Their study on untypical

adjective ordering examples in German suggest that this may be due to discourse linking, when certain adjectives relate more to the subject of a text.

The language used in politics has many interesting features given the context and traditions. Characteristics that differ from average language been previously studied, such as the indirectness in face-threatening situations (Obeng 1997) and the violation of Grice's Maxims (Buddharat et. al. 2017).

Methods

Data Collection

Corpus

The corpus for this project is comprised of speeches from US presidents and US presidential candidate debates. These were selected because there was readily available data, and speeches are likely written before hand, while debates are more spur of the moment language, so both types of language are included. The speech text from the Miller Center of Public Affairs, University of Virginia contains 152 speeches by US presidents from 1789 to 2024. Presidential candidate debates from 2000 to 2016 is from The American Presidency Project, UC Santa Barabara

Subjectivity Scores

The adjective subjectivity scores came from Scontras et. al. 2017. The dataset of individual's responses rating the subjectivity of a set of 78 adjectives (called the "target adjectives" for the rest of this report) was found on Github. The individual scores were then averaged for each adjective (detailed data in appendix).

Extracting Adjective-Adjective-Noun Phrases

The corpus was searched for instances of the 78 target adjectives. For each instance of a target adjective, it was checked if it was a part of an adjective-adjective-noun phrase (AAN). SpaCY, a natural language processing library, was used to tag the part of speech of words to find AAN phrases. Next, the identified ANN phrases were manually reviewed, resulting in 33 removals. These included typos ("individual potential andachievement") or words with multiple parts of speech being tagged incorrectly ("But now **new light breaks** upon us"). Compiled from the speeches and debates, 3153 AAN phrases were found.

Results

Subjectivity-Distance relationship

Results

With this list of AAN phrases, the average distance of each target adjective from the noun was computed, where 0 indicates being directly next to the noun. Of the 78 target adjectives, 57 of them

appeared in AAN phrases in the corpus. A linear trend line was fitted to this data, revealing a weak correlation between subjectivity and distance (r^2 = 0.21, 95% CI [-0.07, 0.49]) (Figure 1).
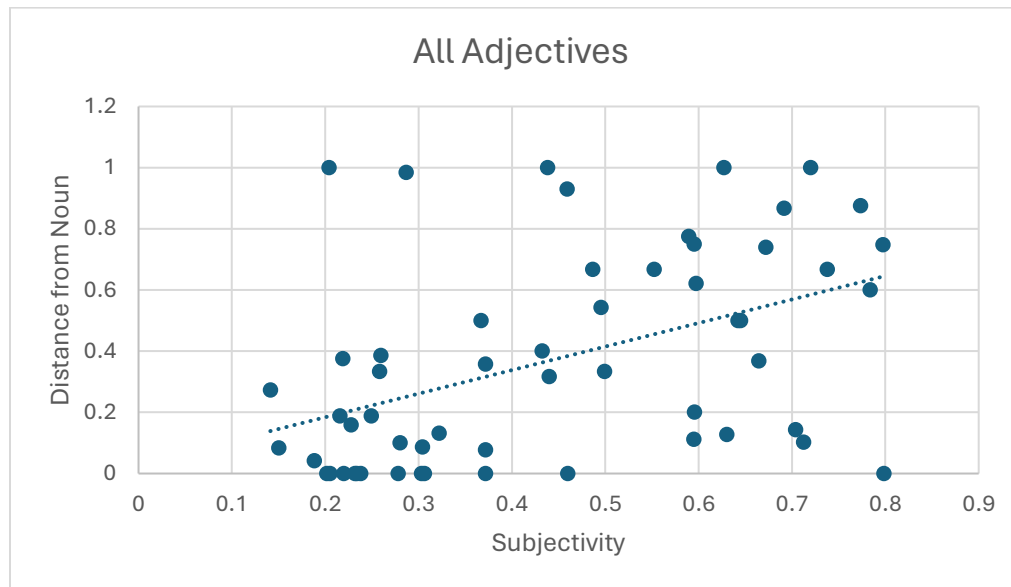
Figure 1 Adjective subjectivity vs distance from the noun with a linear trend line for all target adjectives found in AAN phrases in the corpus.

Given the size of the corpus, certain target adjectives had few data points, potentially affecting the results. To help this, adjectives with less than 10 occurrences were removed. This increased the correlation seen between subjectivity and distance from the noun (r^2 = 0.32, 95% CI [-0.01, 0.67]) (Figure 2).
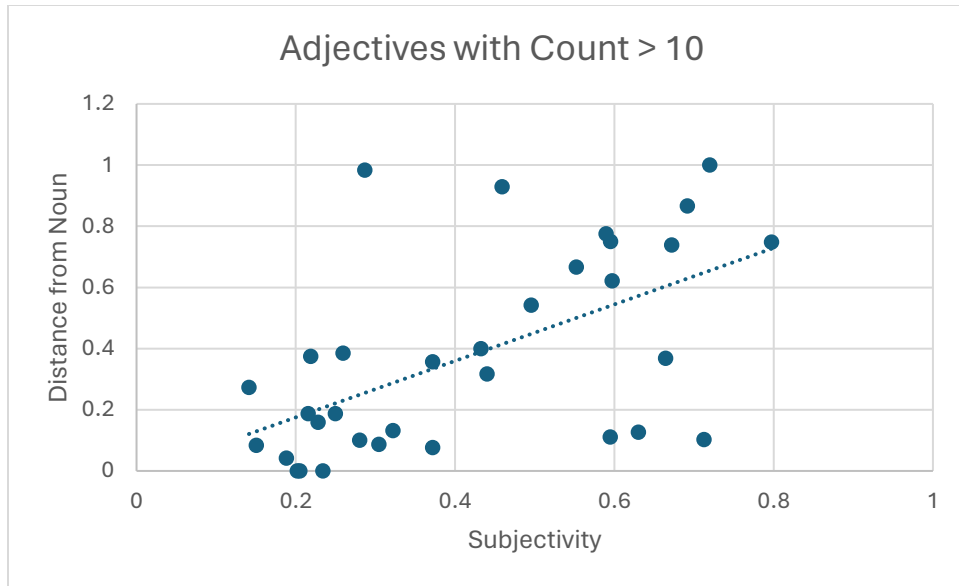
Figure 2 Adjective subjectivity vs distance from the noun with a linear trend line for target adjectives found 10 or more times in AAN phrases in the corpus.

Next, following the approach of Scontras et. al. 2017, superlatives (*last, best, biggest*) were removed from the data set. Superlatives have been found to often break adjective ordering preferences, frequently occurring furthest from noun (Dixon, 1982). In this dataset, the superlatives *last*, *best,* and *biggest* were found and removed. Also, following Scontras et. al. 2017 as an example, outliers were identified as points with a residual greater than three times the standard deviation of the residuals. Using this method, no outliers were identified, so only the superlatives *last, best and biggest* were removed. This strengthens the relationship between distance and subjectivity, but subjectivity still only accounts for less than 50% of the variance in distance (r^2 = 0.39, 95% CI [0.05, 0.73]) (Figure 3).
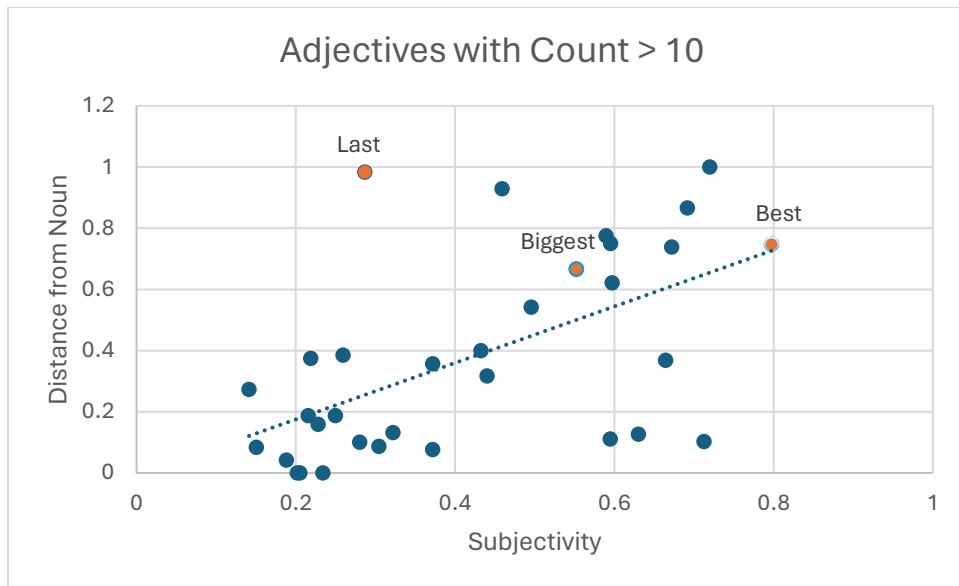
Figure 3 Adjective subjectivity vs distance from the noun with a linear trend line for target adjectives, excluding superlatives found 10 or more times in AAN phrases in the corpus.

Discussion

This project found a weak correlation ($r^2 = 0.39$) between distance and subjectivity. Notably, this is a smaller number compared to the Scontras et. al. 2017 study, which found after removing superlatives and outliers, subjectivity accounted for 70% of variance. It is important to note that this was from an empirical study, asking participants to rate how natural AAN phrases sounded; this project is using corpus data. Scontras et. al. 2017 also looked at corpus data, although with only 26 target adjectives, and got higher statistics ($r^2 = .83$, 95% CI [.63, .90]). Their corpus was over 10 times the size of this presidential speech and debate corpus, which could explain the deviation in results.

Subjectivity Distance

Results

In addition, for AAN phrases both adjectives as target adjectives, the subjectivity difference between the adjectives was also computed. A negative result, where the closer noun was more subjective than the further, would indicate not following the subjectivity hypothesis. Of the 160 instances of AAN phrases with two target adjectives, the mean and median subjectivity difference was greater than zero (Figure 4). Only 22.5% of the phrases had a negative subjectivity difference.
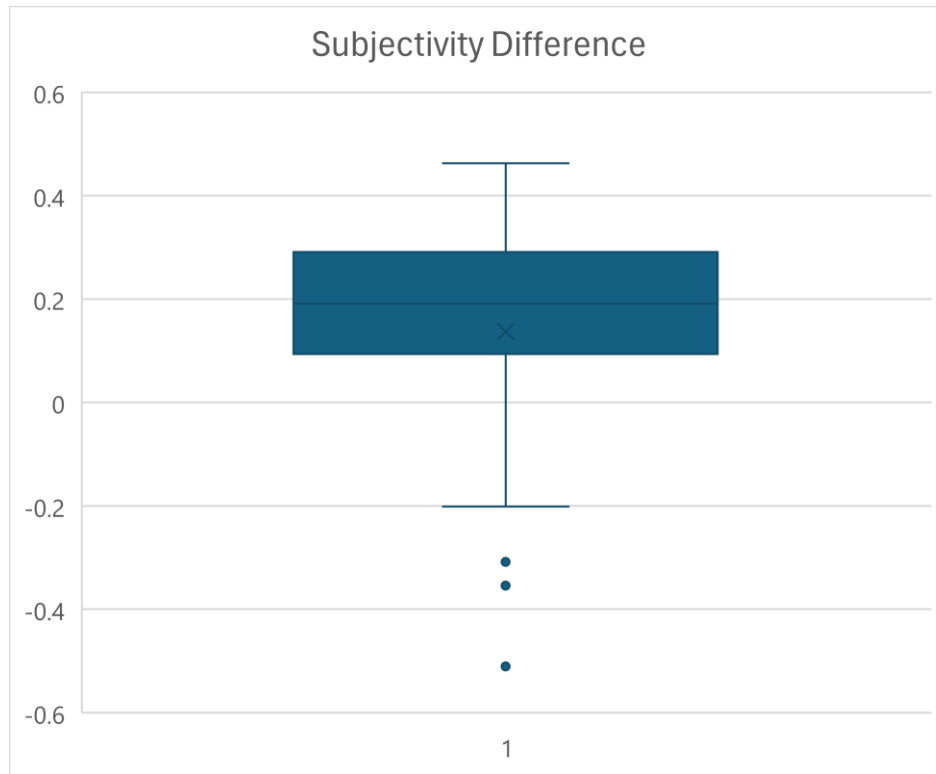
Figure 4 Subjective difference, measured (subjectivity of first adjective – subjectivity of second adjective) for all AAN phrase with two target adjectives

Next, like in the previous section, phrases containing superlatives were removed, resulting in 30 removals. The only superlatives present were *last* and *best*. This increased both the mean and median (mean = 0.23, median = 0.23) (Figure 5). Only 6.2% of the phrases had a negative subjectivity difference.
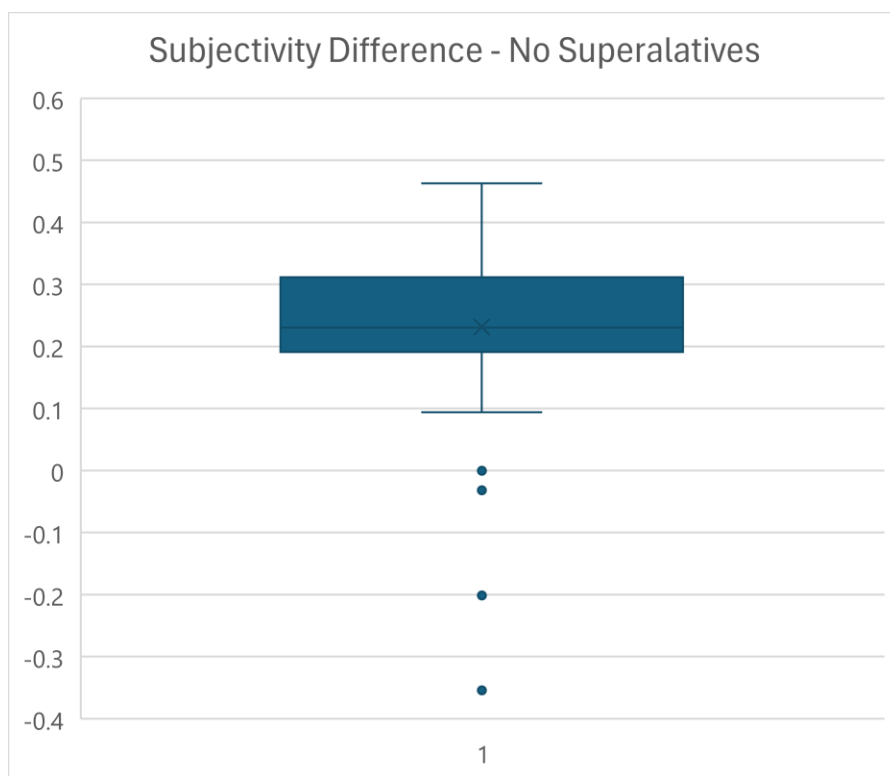
Figure 5 Subjective difference, measured (subjectivity of first adjective – subjectivity of second adjective) for all AAN phrase with two target adjectives, excluding phrases containing superlatives.

Discussion

This analysis supports the subjectivity-based adjective ordering preference more than the previous. After removing superlatives, only a small value, 6.2%, of instances violated this adjective ordering preference indicated by a negative subjectivity difference. Some specific violating instances are considered below in the general discussion.

General Discussion

As discussed above, there was not a strong relationship between adjective subjectivity and distance from the noun in the presidential debates and speeches. This suggests that adjective ordering preferences differ in these contexts.

To investigate a linguistic explanation for this, the adjectives with the largest absolute residuals were considered. After the superlative *last*, the next adjectives were *civilized, current, young,* and *professional*. The subjectivity scores for *current, young,* and *professional* (0.46, 0.63, and 0.60, respectively) are all near 0.5, not indicating strong subjectivity nor objectivity. However, *civilized* has a high subjectivity score of 0.71. The subjectivity-based adjective ordering preferences suggest that *civilized* would be found further from the noun, but in *civilized* had an average distance of 0.10 from the noun for the 39 times it appeared in this corpus. Examples of *civilized* from the corpus include:

- "Four years ago, the thoughts and fears of the **whole civilized world** were centered on another piece of American soil—Pearl Harbor" - Harry S. Truman, September 1, 1945: Announcing the Surrender of Japan
- "There is not a **single civilized power** which has anything whatever to fear from aggressiveness on our part." - Theodore Roosevelt, December 3, 1901: First Annual Message
- "While **other civilized nations** have encountered great expense to enlarge the boundaries of knowledge by undertaking voyages of discovery..." - Thomas Jefferson, January 18, 1803: Special Message to Congress on Indian Policy

In these examples, the adjectives *whole, single,* and *other* are clearly less subjective than *civilized,* but *civilized* appears closest to the noun. One interpretation is that the speaker assumes all listeners agree on what a *civilized world/power/nation* is, so then places *civilized* closest to the noun, as if it is objective.

However, it is not necessarily true that everyone agrees, given the subjectivity of *civilized.* Nonetheless, the speaker maintains this perhaps because a political speech is not a typical conversation, so the RSA framework, the recursive understanding process between speaker and listener, may not fully apply. In these situations, keeping a favorable public image is an importance. Similarly, in presidential debates, the purpose is to persuade the public of one's candidacy. Therefore, speaking as if there is unanimity on the meaning of *civilized* might improve the speaker's persuasiveness to the audience.

On the other hand, the results on the subjectivity difference between adjectives does not support the claim that adjective ordering differs in political spoken language, only a handful of the instances did not follow the subjectivity-based order preference. Additionally, the corpus was not vast. Perhaps with more data, the results would even out to follow the same pattern found in general language.

Conclusion

This project sought to answer if the subjectivity-based adjective ordering preferences are adhered to the same extent in US presidential speeches and debates. While the correlation between adjective subjectivity and distance from the noun was weak, almost all instances where the subjectivity score for both adjectives was available followed the preference of having the less subjective adjective closer to the noun. Because of these conflicting results, it is hard to conclude that the adjective ordering preferences were followed or not.

In future steps, the corpus size could be expanded to include speeches from other political figures (not just presidents), congressional debates, campaigning speeches, and other countries and/or languages. Additionally, the method for extracting the AAN phrases could be improved. The SpaCY library used for part of speech tagging is not completely accurate, and despite attempts to manually clean the data, it is likely that some invalid phrases were overlooked. On the other hand, it is also possible that some valid phrases were not included because they were not caught by the tool. Finally, more specific examples could be analyzed in depth, as the *civilized* examples were in the discussion, to develop more explanations for why there might be differences in adjective ordering preferences.

Bibliography

Buddharat C., Ambele, E., & Boonsuk, Y. 2017. "Uncooperativeness in Political Discourse: Violating Gricean Maxims in Presidential Debate 2016". *Songklanakarin Journal of Social Sciences and Humanities*.

Kachakeche, Zeinab, and Gregory Scontras. 2020. "Adjective Ordering in Arabic: Post-Nominal Structure and Subjectivity-Based Preferences". *Proceedings of the Linguistic Society of America*

Kotowski, Sven and Holden Hartl. 2019. "How real are adjective order constraints? Multiple prenominal adjectives at the grammatical interfaces". *Linguistics*.

Martin, James E. 1969. "Semantic determinants of preferred adjective order". *Journal of Verbal Learning and Verbal Behavior*.

Obeng, Samuel. 1997. "Language and Politics: Indirectness in Political Discourse". Discourse & Society.

Samonte, Suttera, and Gregory Scontras. 2019. "Adjective Ordering in Tagalog: A Cross-Linguistic Comparison of Subjectivity-Based Preferences". *Proceedings of the Linguistic Society of America*

Scontras, Gregory, Judith Degen, and Noah D. Goodman. 2017. "Subjectivity predicts adjective ordering preferences". *Open Mind*.

Scontras, Gregory, Judith Degen, and Noah D. Goodman. 2019. "On the grammatical source of adjective ordering preferences". *Semantics and Pragmatics*.

Wulff, Stefanie. 2003. "A multifactorial corpus analysis of adjective order in English". *International Journal of Corpus Linguistics*.

Appendix

Full data and code is available at https://github.com/gdriskill/LING8803-Project.git

Note: ChatGPT was used after a first draft to condense some sections in order to meet the word limit.