Linguistic Style in Political Discourse: A Quantitative Analysis of Part-of-Speech and Formality Distributions in Joe Biden's 2020 Speeches

Regina Kukuh Prasetyo Ati

Abstract

This study explores the interplay between part-of-speech (POS) tags and formality scores in differentiating Joe Biden's formal and informal speeches. The analysis focuses on Heylighen and Dewaele's (1999) F-score to examine Biden's 2020 Democratic Convention Speech (formal) and Thanksgiving Day Speech (informal). The formality is measured by analyzing word type frequencies from different genres, aligning with the expectation that "the more formal the language excerpt, the higher the value of F" (Heylighen & Dewaele, 1999). The analysis aims to contribute to the understanding of the role of language in shaping formality and understanding how language is used in different contexts. The analysis reveals that Biden's Democratic Convention speech exhibits a higher F-score compared to his Thanksgiving speech. The difference suggests that Biden adapts his language style to suit the specific context or the formality of each occasion.

Keywords: formality, language, word frequencies, context

Introduction

Language is not just a code for transmitting meaning, but also a social tool that we use to enact different social roles and relationships. According to the variety of situations, language use varies depending on the situation in which it occurs, which can be referred to as registers. Drawing on Halliday and Hasan's (1985) *register theory*, it emphasizes how context-specific variations in language are categorized by field (topic), tenor (relationship between participants), and mode (channel of communications). Political communication exemplifies this adaptability, with the field being politics, the tenor is typically formal, and the mode is often spoken. Formality holds particular weight in political discourse, where speakers strive to effectively communicate and convey their messages to diverse audiences. Formal language is generally characterized and often associated with more complex sentence structure and a more extensive vocabulary. On the other hand, informal language is often typically more direct, concise, and familiar.

This study examines part-of-speech (POS) tagging and Heylighen and Dewaele's F-score (1999) to investigate the linguistic register employed by Joe Biden in formal and informal speeches delivered during 2020. POS tags offer a grammatical classification of words, revealing their syntactic functions within sentences. The F-score, in this instance Heylighen and Dewaele's F-score (1999), serves as a quantitative measure of formality based on the relative frequencies of specific word types. The F-score is calculated by summing the relative frequencies of specific word types associated with formality (nouns, adjectives, prepositions, and articles) and subtracting those associated with informality (pronouns, verbs, adverbs, and interjections). The resulting value, normalized by a factor of 2 and adjusted by a constant of 100, serves as a numerical indicator of formality within a text with higher scores indicating more formal writing.

Through analyzing POS tags and formality scores across two contrasting speeches, the writer seeks to identify linguistic features contributing to their perceived formality. The study anticipates revealing how Biden strategically adapts his language to varying contexts, employing more formal language in more formal settings.

Research Question
1. How does the POS tag distribution relate to the F-score?
2. What is the Heylighen and Dewaele F-score for Joe Biden's formal and informal speeches in 2020?

Description of the Data
The data for this study consists of two speeches delivered by Joe Biden in 2020: The Democratic Convention Speech and the Thanksgiving Speech. The data source is the Kaggle repository, which ensures authenticity and provides public accessibility to the transcripts of both speeches.

Methods of Analysis
This study employs a computational methodology, implemented through Python libraries such as spaCy, to explore the quantifiable relationship between formality and the distribution of parts of speech (POS) in text. The analysis adopts the F-score metric, established by Heylighen and Dewaele (1999), to quantify and compare the formality levels of two speeches. The analysis anticipates a higher F-score for the formal speech, reflecting the expected correlation between formality and POS tag frequencies. The methodology can be summarized in the following sequential key steps:

1. **Data Acquisition and Preparation**:
   A corpus of Joe Biden's speeches, categorized as formal: Conventional Democratic Speech and informal: Thanksgiving speech, is subjected to comprehensive data-cleaning procedures to ensure accuracy and consistency.
2. **Part-of-Speech Tagging:**
   This study utilizes spaCy, a well-established Natural Language Processing (NLP) library written in Python, to perform part-of-speech (POS) tagging on the speeches. POS tagging, a fundamental NLP task, involves assigning grammatical labels (e.g., noun, verb, adjective) to individual words based on their context and surrounding linguistic elements.
3. **Formality Score Calculation:**
   The formality of each speech is quantified using the F-score. This score is computed by analyzing the frequency distribution of part-of-speech (POS) tags within each speech, following the formula established by Heylighen and Dewaele (1999, 2002)
4. **Analysis:**
   This study employs a comparative analysis of the calculated F-scores for both speeches. This analysis aims to determine, with statistical significance, whether the formal speech exhibits a demonstrably higher F-score.

Conclusion
This study examined the formality of Joe Biden's 2020 speeches using Heylighen and Dewaele's (1999, 2002) F-score and genre-specific word frequency analysis. As expected, Joe Biden's formal Democratic Convention speech exhibited a significantly higher formality score compared to his informal Thanksgiving speech. This aligns with expectations that Biden adapted his language style to suit the formality of each setting. These findings contribute to our understanding of Joe Biden's communication strategies by highlighting the linguistic choices he employs to adapt his message to different contexts.

References

Halliday, M. A. K. and Hasan, R. 1985. *Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective*. Geelong: Deakin University Press.

Heylighen, F., & Dewaele, J.-M. 1999. *Formality of Language: Definition, measurement and behavioral determinants (Internal Report)*. Center "Leo Apostel", Free University of Brussels.