A Corpus-based Stylistic Analysis of Female Speeches: A Case Study of Michelle Wu's and Hillary Clinton's Speeches

Xudong Sun

Beijing Institute of Technology, Zhuhai, Zhuhai, Guangdong, China

d666666ng@163.com

Abstract. With the advancement of society and the elevation of women's social status, the study of female language has increasingly gained attention. This thesis explores the linguistic features of women's discourse in public speaking through corpus analysis tools. It aims to contribute to the enrichment and advancement of related research fields. This thesis employs corpus analysis to elucidate the speech traits exhibited by the research subjects in terms of overall discourse complexity, vocabulary richness and complexity, as well as personal attributes. Specifically, it focuses on the public speeches of Hillary Clinton and Michelle Wu as its research subjects. By comparing their speech patterns, this study reveals that Michelle Wu demonstrates a higher level of discourse complexity compared to Hillary Clinton, utilizing more diverse vocabulary. However, Hillary Clinton's vocabulary carries a greater depth of information. Upon observation of their speech performances, it is evident that Michelle Wu appropriately integrates slang, while Hillary Clinton excels at evoking the audience's emotions through various rhetorical devices and humor.

Keywords: public speaking, Michelle Wu, Hillary Clinton, corpus.

1 Introduction

With the progress of society and the strengthening of gender awareness, women's social status, professional roles and psychological processes are undergoing changes. The topics related to women are not only increasingly appearing in the fields of politics, sports and culture, but also causing considerable reactions in the fields of education and academia.

At present, many scholars have turned their attention to research on female based stylistic construction and discourse strategies. In research on women's language features, scholars focus on analyzing women's language features through texts extracted from movies, television or media. Zhu and Sun conducted a critical analysis of journey metaphors in public speaking to investigate their role in constructing female identity[1]. Du and Chen examined the discursive strategies employed by female characters, particularly exemplified by the protagonist Offred in the dystopian novel *The Handmaid's*

© The Author(s) 2025

A. L. Baldini et al. (eds.), *Proceedings of the 2025 3rd International Conference on Language, Innovative Education and Cultural Communication (CLEC 2025)*, Advances in Social Science, Education and Humanities Research 938,
https://doi.org/10.2991/978-2-38476-430-3_35

Tale, through linguistic analysis of their verbal expressions[2]. Fan adopted a sociolinguistic perspective to explore how linguistic features in web film dialogues construct feminine identity and reflect patriarchal societal norms[3]. Wang performed a corpus-based comparative study analyzing linguistic stylistic differences in online consumer reviews of women's skincare products between Chinese and American e-commerce platforms[4]. However, research based on publicly speaking corpora often approaches it from the perspective of functional grammar or pragmatics. Therefore, in this context, it is particularly timely and important to use corpora to conduct stylistic analysis of women's speeches.

2 Investigating Linguistic Features by Corpus Tools

A corpus constitutes a systematically sampled and processed textual database that enables computer-aided retrieval and analysis, serving as an instrumental tool for conducting both theoretical investigations and practical applications in linguistic research[5]. This thesis utilizes two corpus analysis tools. AntConc is a multi-platform corpus analysis toolkit based on the Microsoft Windows operating system developed by Dr. Another corpus analysis tool is the readability detection tool for scientific research on the We Research website, which can detect the number of words, sentence length, sentence count, and six related international readability indicators. The methodologies of corpus linguistics and discourse analysis demonstrate rationality and appropriateness in examining linguistic features and exploring ideological factors underlying discursive practices[6].

3 The Personal Speech Styles Reflected by Michelle Wu and Hillary Clinton

3.1 Comparative Analysis of Michelle Wu and Hillary Clinton

The Overall Difficulty of the Discourse. In exploring the overall discourse difficulty and readability of the two speeches, a comparison was made through average sentence length and fog index. Among them, average sentence length (ASL) refers to the average number of words contained in a sentence in a specific corpus, and the difficulty and complexity of the text can be understood based on the average sentence length. Butler once divided sentences into three categories based on their length: short sentences (1-9 words), medium length sentences (10-25 words), and long sentences (25 words or more)[7]. The degree of accessibility is often measured by a formula called Fog Index. Fog Index=0.4 (L+H), L refers to the average length of senses, H refers to the percentage of big words. The higher the fog index, the higher the reading difficulty of the text, with index 10 being the average difficulty. By using the Antconc corpus to count the tokens, and using the corpus analysis tool on the We Research website to count the number of sentences, average sentence length and number of complex words, the fog index is calculated based on existing data. The following table shows the statistical results.

Table 1. Statistics on Discourse Difficulty Related Data.

	Tokens	Complex Words	Sentences	ASL	Fog Index
Hillary	3698	438	205	17.56	11.89
Michelle	2216	201	106	20.57	11.92

According to the above table 1, the average sentence length of both individuals is between 10-25 words, so they belong to medium length sentences. However, Michelle Wu's average sentence length is 20.57, which is significantly higher than Hillary's. The larger the average sentence length, the more evidence the speaker expresses more content in one sentence. Compared to Hillary's speech, Michelle Wu's speech implies that listeners need more energy to understand. In terms of comparison of fog index, there is not much difference between the two, with two fog index slightly greater than 10, so the overall readability of the text is moderate. The fog index of Michelle Wu's speech corpus is slightly higher than that of Hillary, which corresponds to the comparison of average sentence length, still indicating that the discourse difficulty of Michelle Wu's speech is higher than that of Hillary.

The Richness and Complexity of Vocabulary. The measurement of vocabulary usage in Michelle Wu and Hillary Clinton's speeches is mainly considered from three aspects: lexical diversity, lexical complexity and lexical density. Among them, lexical diversity refers to the proportion of non-repeating words in the text, which is detected by TTR in this thesis. TTR refers to the ratio of types to tokens in the selected corpus, which can to some extent reflect the lexical richness of the text. The higher the value of TTR that represents the higher the richness of vocabulary in the text, the more types of vocabulary are used. Lexical complexity is measured by the average word length (MWL), which refers to the average number of letters contained in words in a corpus. The word length of a typical text is around four letters. If it is less than four letters, it means that the words used and the text are relatively concise and simple. If it is much longer than four letters, it means that the words used and the text are more complex[8]. In terms of lexical density (LD), Stubbs pointed out that it can be calculated by the ratio of the number of content words to the tokens in a corpus[9]. According to Halliday's classification of vocabulary, content words include nouns, adjectives, adverbs, and verbs[10]. According to the calculation formula proposed by Ure: lexical density= content words/tokens * 100%[11]. The size of lexical density directly reflects the information carrying capacity of the text, and also indirectly reflects the difficulty of vocabulary[12].

Based on the above three aspects, this thesis uses the AntConc corpus for the statistics of tokens and types, and uses the corpus analysis tool under the We Research website for the statistics of average word length. For the calculation of lexical density, this thesis uses Clwas7 online annotation to assign codes to the selected text, and then uses AntConc for the statistics of content words, and calculates the final figure. The following table shows the relevant data.

Table 2. Relevant Statistics of Vocabulary Information.

	Tokens	Types	MWL	TTR	LD
Hillary	3698	1084	4.42	29.06	47.27%
Michelle	2216	809	4.41	35.56	46.34%

According to the table 2, although Michelle Wu has a small vocabulary, her TTR is much higher than Hillary's, indicating that Michelle Wu uses more diverse vocabulary. In terms of average word length (MWL), the average word length of the two texts is similar, around four letters, so the difficulty of using words in their speeches is relatively low, which is basically in line with the language used in public speeches. In terms of lexical density (LD), neither of them has a high vocabulary density, which is consistent with the results displayed by the average word length. Among them, Hillary's LD is 47.27%, significantly higher than Michelle Wu's 46.34%, indicating that Hillary has a higher information carrying capacity than Michelle Wu, and has certain difficulties in word selection.

Overall, in terms of vocabulary richness and complexity, the difficulty of selecting words for both is not high, which is in line with the language used in public speeches. Compared to the two, Michelle Wu's vocabulary is more diverse, while Hillary's vocabulary has a higher information carrying capacity.

The Exploration of Personal Speech Traits. By observing the speeches, it can be seen that in addition to analyzing the speech styles of Hillary and Michelle Wu from a stylistic perspective, they also have their own unique personal speech styles. From Michelle Wu's speech, she is good at adding some slang expressions at the beginning or end of her speech, such as *day in and day out*, *love knowledge no bounds*, *too pie in the sky*, which helps to add fun and attraction to the speech. Hillary is good at using sarcastic and self deprecating language in her speeches to stimulate the audience's emotions and make the speech more humorous. In addition, by observing the speech, it can be found that Hillary is good at grasping the emotions of the audience, making reasonable pauses, and completely breaking free from the script during the speech, demonstrating a confident and mature image.

(1) It was like Facebook, but the bulletin board didn't steal your personal information.

(2) Because if you thought my emails were scandalous, you should hear my singing voice.

(3) I went for long walks in the woods. Yale students went for long walks in East Rock Park. I spent hours going down a twitter rabbit hole. You spent hours in the Yale memes group. I had my fair share of Chardonnay, you had penny drinks at Walt's. I practiced yoga and alternative nostril breathing, you took psych and the Good Life.

(4) And because of that, in the end, you changed Yale as much as Yale changed you.

In (1), she used the bullet board to mock Facebook for stealing personal information, which resonated with the audience. In (2), Hillary used self deprecation to imply that her singing was even worse than her own scandal, turning negative information into humorous elements and causing laughter from the audience. In addition, Hillary also

pays great attention to the art of language expression. She uses a lot of parallelism and contrastive rhetoric in her speeches, such as the (3). She compares her own life behavior with that of students, and through daily life, resonates with each other, successfully achieving communicative effects. In (4), Hillary also used palindrome rhetoric at the end of her speech, linking students to Yale University and providing a dual response, elevating the theme of the speech. In addition, by observing the speech, it can be found that Hillary is good at grasping the emotions of the audience, making reasonable pauses, and completely breaking free from the script during the speech, demonstrating a confident and mature image.

3.2 Some Advice on Public Speaking

After this comparative analysis, it can be seen that Hillary and Michelle Wu have certain differences in language use, overall discourse structure, and personal speech style. Based on this exploration, the following suggestions have been summarized: In terms of language rhetoric, Michelle Wu can choose to use direct quotations to strengthen the persuasiveness of his views; In terms of emotional mobilization, some humorous effects can be appropriately added to regulate the atmosphere of the scene and gather the attention of the audience; In terms of speech rhythm, appropriate pauses or the addition of some filler words can be used to control the on-site rhythm, thereby presenting a more confident image. For Hillary, in terms of vocabulary selection, increasing the use of synonyms or changing expressions can improve vocabulary richness.

4 Conclusion

By exploring and comparing the speeches of Hillary and Michelle Wu, it can be seen that both of them highlight the many features of women's language. In terms of overall speech style, Michelle Wu's discourse difficulty is higher than Hillary's, and the difficulty of their vocabulary is moderate. However, Michelle Wu's vocabulary is more diverse, while Hillary's vocabulary load is greater. From the perspective of personal speech performance, Michelle Wu's use of slang is prominent in her speeches. Hillary's speech language is more humorous, better at mobilizing the audience's emotions, and the language expression is rich, using various rhetorical devices, and pausing appropriately during the speech process.

From an overall perspective, compared to Hillary, Michelle Wu can appropriately add humorous elements to adjust the atmosphere. For Hillary, further enrichment can be made in vocabulary selection. Overall, they have unique personal styles, which make their speeches more infectious, reflecting the advantages and power of female leaders in public speaking.

The analysis of the different speech features of Michelle Wu and Hillary provides referential significance for future researchers to analyze women's language and corpus, and also provides learning directions for future women when giving speeches, helping more women to examine the speeches from multiple perspectives and increase their confidence. In addition, the number of selected texts in this thesis is limited, and there

is limited analysis of the details, such as keyword matching, which has not been explored yet. Subsequent researchers can conduct more in-depth and comprehensive research and analysis.

References

1. Zhu, L. L., & Sun, Y. (2023). Critical metaphor analysis of female identity construction in English speeches. *New Chu Culture*, (25), 81-84.
2. Du, H. Q., & Chen, A. M. (2024). Female discourse strategies and discourse authority in *The Handmaid's Tale*. *Journal of Changzhou Institute of Technology (Social Science Edition)*, *42* (4), 63-67.
3. Fan, Y. H. (2024). Analysis of female language in online films from a sociolinguistic perspective. *Film Literature*, (7), 48-52.
4. Wang, C. Y. (2023). A contrastive study of linguistic styles in online reviews of female skincare products in China and the United States. In *Comparative and translation studies of English and Chinese* (Vol. 15, p. 189). Shanghai Maritime University.
5. Li, J. X. (2023). A corpus-based textual analysis of *Frankenstein*. *Modern English*, (24), 79-81.
6. Yang, L. Q. (2023). Corpus-based analysis of corporate discourse cases. *Forum on Language and Culture*, (1), 128-139.
7. Butler, C. (1985). *Statistics in linguistics*. Basil Blackwell.
8. Chen, J. S., & Gao, B. (2011). A corpus-based study on translator styles in two English versions of *The Book of Songs*: Focusing on "Guofeng". *Journal of Tianjin Foreign Studies University*, *4*, 36-41.
9. Stubbs, M. (1986). Lexical density: A technique and some findings. In M. Coulthard (Ed.), *Talking about text* (pp. 27-48). University of Birmingham. STUBBS M. Lexical density: A technique and some findings //COULTHARD M. Birmingham: University of Birmingham, 1986: 27-48.
10. Halliday, M. A. K. (1985). *Spoken and written English*. Deakin University Press.
11. Ure, J. N. (1971). Lexical density and register differentiation. In G. Perren & J. L. M. Trim (Eds.), *Applications of linguistics* (pp. 443-452). Cambridge University Press.
12. Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target*, *7* (2), 223-243.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.