

Study on the Correlation between Linguistic Complexity and Audience Recognition in College English Speech Contests

Lin Shi

Department of English, College of Foreign Languages, Ocean University of China, Qingdao, Shandong, China

shilin@stu.ouc.edu.cn

Abstract. From the perspective of linguistic complexity, this paper explores the correlation between linguistic complexity and audience recognition in college English speech contests. By adopting the corpus construction and computer visualized data analysis, this study analyzes the speech of contestants at different levels in FLTRP Cup National English speaking Contest 2019-2020, the most authoritative college English speech contest in China. The study shows that: 1) In college English speech contests, the lexical complexity of the speech is negatively correlated with the recognition degree of the audience (i.e. the final ranking of the competition or the success of the speech). 2) The syntactic complexity of the speech exists reasonable interval to ensure good audience recognition. 3) In college English speech contests, the correlation between the lexical complexity and syntactic complexity of speeches and audience recognition is similar to the correlation between the rhetoric and audience recognition in political speeches which is obtained by previous researchers in the field of political speeches. Therefore, we think this study has a certain practical value. It provides evidence of linguistic complexity for predicting the winner of college English speech contests and helping contestants prepare for the contest better.

Keywords: Corpus; Computer Visualized Data Analysis; Linguistic Complexity; Lexical Complexity; Syntactic Complexity; Audience Recognition.

1. Introduction

Linguistic Complexity is one of the directions in the study of bilingual language features (including fluency, accuracy, and complexity) in the context of the linguistic developmental perspective of complex dynamic systems theory, and Norris & Ortega (2009) argue that linguistic complexity should be viewed as "a dynamic, interconnected, and simultaneously changing set of subsystem collections".

Syntactic Complexity is an important dimension of linguistic complexity and one of the most fundamental elements of linguistic competence. Numerous studies have investigated the relationship between syntactic complexity and second language ability (Lu 2011; Ortega 2015). The results of the existing studies show that syntactic complexity can determine the level of second language ability. Factors affecting syntactic complexity, such as sentence length, sentence structure, and phrase structure, have also been the focus of attention.

Lexical Complexity is another important dimension to measure linguistic complexity. According to Bulté & Housen (2012), lexical complexity is composed of lexical richness and lexical depth. Lexical richness includes the density and diversity of lexical resources; lexical depth includes the difficulty and composition of lexical resources. Lexical richness is a horizontal feature of vocabulary complexity, including the number of word forms and word families, the number of real words and functional words, and the number of near-sense words and antonyms, reflecting the breadth of vocabulary acquired by learners.

The previous stage of research on linguistic complexity mainly divided and treated lexical and syntactic in a compartmentalized manner. In recent years, researchers have increasingly realized that the language system is a whole, and not only syntactic complexity should be studied, but also Lexical Complexity should be emphasized (Skehan 2009; Bulté & Housen 2012; Wang Haihua, Li Beibei, and Xu Lin 2015). On this basis, scholars further proposed that research should focus on the linkage changes of both (Norris & Ortega 2009; Wang Chu-Ming 2008). Since then, some studies break through the tradition of lexical or syntactic divide and conquer and begin to focus on the joint

development of lexicon and syntax (Xu Xiaoyan, Wang Weimin, Xiong Yanyu, Jiang Jing, Pan Xiaoyan, and Sun Nianhong 2013). Through the analysis of a large corpus, researchers found that the linguistic complexity of second language compositions can effectively distinguish learners' language levels, predict composition quality, and become more complex as language levels grow. At the same time, linguistic complexity in second language composition is also influenced by factors such as learners' native language background, task type, and teaching environment (Lu 2011; Bulté & Housen 2014; Crossley & McNamara 2014; Mazgutova & Kormos 2015; Bao Gui 2009; Wang Min and Wang Chuming 2014).

In recent years, research on the further integration of linguistic complexity and analysis of written second language information has gradually emerged. Among them, the effect of speech rhetoric on listeners has been a hot research topic in this field. Researchers in this field have used linguistic complexity as one of the criteria for measuring Speech Style and Rhetoric.

In the foreign academic community, the creation of a corpus of general policy statements of the French government was an early study using corpus construction and computer visualization analysis. However, this study did not further explore the correlation between linguistic complexity and audience approval.

The Opinion Funder linguistic analysis system found a negative correlation between the linguistic complexity of positive tweets tagged with "Obama" and presidential approval ratings, i.e., the lower the linguistic complexity of positive tweets, the better the presidential approval rating. However, the data for this study was collected from the social media platform Tweet, where the average length of tweets is only 11 words and many tweets are not well thought out. Therefore, compared to TV debates, campaign speeches, etc., tweets are less informative. In addition, by conducting a study from a historical perspective and selecting written texts of State of the Union speeches from Washington to Obama, Wasike (2017) explores the correlation between linguistic complexity and re-election from the perspectives of the frequency of long word usage, appellative changes, speech speed, and high-frequency use of words.

There are also many specific analyses of individual speech rhetorical characteristics in studies related to speech text linguistic complexity and audience recognition. For example, by analyzing Hillary's speech rhetoric, Samuel Gunawan (2017) found that Hillary tended to repeat complex reasoning in a large number of long sentences. Jacques Savoy (2018) used corpus and computer visualization analysis to study the political speeches of each candidate in the 2016 U.S. presidential election and found the linguistic style and similarities and differences in rhetorical devices and the impact of such similarities and differences on the effectiveness of political speeches. Savoy (2018) refines the field of study of linguistic complexity from political texts to election speeches, arguing that the speech message consists of both written and oral components. Of these, the oral form of communication is a more direct and free form of interaction that more closely reflects the personal style of each candidate. The written message is prepared in advance by the candidate, and a large portion of the written statements in an election are co-authored by multiple politicians. Therefore, it is important to analyze these two forms of communication separately. The debate process, on the other hand, is a more spontaneous form of communication that provides a truer picture of the language and rhetorical style of the campaigners. On this basis, Savoy used formal campaign debates collected through official television media as a data source to create the Electoral Corpus, which includes both written and short-lived forms of oral communication. And from multiple perspectives such as lexical richness, most frequent lemmas, the mean sentence length (MSL), most specific terms, thematic concentration, POS distribution, grammatical categories, and other perspectives to visually analyze the similarities and differences in the linguistic styles and rhetorical approaches of the campaigners. The study found that in the 2016 presidential primaries, both winners, Trump and Hillary, used the pronoun "I" extensively, while the other candidates did not use it significantly. Trump even showed a distinctly atypical image, using short sentences, reducing his vocabulary, repeating the same arguments in simple words, and preferring dense verb structures. The high frequency of Trump's repetition of the words "Mexico" and "deal" reflects his two main concerns: immigration and

commercial trade agreements. The responses of the other candidates of the same period, however, contained more nouns and names and used more descriptive rhetoric. In contrast, Trump, with his low linguistic complexity, had a more significant advantage. But in this study, some minor figures who did not appear in the televised debates or appeared less frequently were overlooked due to the conditions of data collection. And because different candidates speak for different lengths of time in a debate, there is some effect on factors such as word frequency, which may narrow the difference in the study results.

In summary, the existing studies have the following problems: 1) Some of the studies on speech language complexity and audience recognition have cut the link between vocabulary and syntax and failed to focus on the co-development of both. 2) The creation of a corpus of French government general policy statements is an early study using corpus construction and computer visualization and analysis methods. However, the correlation between linguistic complexity and audience recognition has not been discussed in that study. 3) The reference value of the samples selected for the study is not uniform. In recent years, there has been an increase in the number of studies conducted on political type speeches through online communication channels. Researchers are using more information from social platforms such as Tweet as a data source for their studies and less from blogs or audio and video media such as YouTube. For example, the Opinion Funder system, in examining the relationship between the linguistic complexity of positive tweets and presidential approval ratings, used tweets with an average length of only 11 words as the subject of analysis. These research materials have low reference value compared to TV debates and campaign speeches. 4) Previous research on the correlation between linguistic complexity and audience approval has mainly focused on the field of political speeches, and less on college English speech contests. Therefore, there is still a gap in the field of college English speech contests. 5) Some of the research variables in the study are not unified, which might have some influence on the research results. For example, in Savoy's (2018) study on the analysis of the political speeches of each candidate in the 2016 U.S. presidential election, some secondary figures who did not appear in the TV debates or appeared less frequently were ignored due to the conditions of data collection. And since the length of speeches of different candidates in a debate varies, which has an effect on word frequency, etc., this may narrow the differences in the study results.

Therefore, based on the interdisciplinary techniques of corpus construction and computer visualization data analysis, this study explores the correlation between linguistic complexity and audience recognition in college English speech contests, which includes three main research questions: 1) How does the lexical complexity of a speech affect audience recognition in college English speech contests (i.e., the final ranking or success)? 2) How does the syntactic complexity of speeches affect audience recognition in college English speech contests? 3) Is the correlation between linguistic complexity and audience approval in college English speech contests consistent with the correlation between linguistic complexity and audience approval in political speeches, or is there a similar pattern?

2. Study Design

2.1 Subjects

The subjects were 32 students from the first to the fourth year of various majors from eight "double-first class" universities in China. Among them, 9 were male and 23 were female, aged 19-22 years old. All of them had studied English for 12-16 years and had a good foundation in English.

2.2 Data Collection

The source of data for this study was the FLTRP Cup National English Speaking Contest, the most authoritative English speaking contest available in domestic universities. The researcher collected the text of 32 contestants' speeches with the theme of "Challenge to all", the video of their speeches, and the final results of the contest for data analysis.

The FLTRP Cup National English Speaking Contest is a public service event organized by the Foreign Language Teaching and Research Press in cooperation with the Steering Committee for Teaching Foreign Languages in Colleges and Universities of the Ministry of Education and the Sub-Committee for Teaching English as a Foreign Language in Colleges and Universities of the Ministry of Education for college and university students nationwide. The FLTRP Cup National English-Speaking Contest was founded in 2002 and has become the most popular, largest, and highest-level English-speaking contest in China. All 32 participants came from eight "double-first class" universities in China, and all of them had a good foundation of English language learning, having studied English for 12-16 years, which further ensured the quality of the speech text. Based on the final ranking and advancement stage of the 32 subjects in the speech contest, the data were divided into four groups: 8 in the school-level preliminary group, 8 in the school-level final group, 8 in the provincial group, and 8 in the national group. The researcher used Ant Conc software to import all the texts and create the "Unipus Speech Corpus" with Corpus Files of 32, Word Types of 2047, and Word tokens of 13,870.

2.3 Data Analysis

The corpus and the data analysis software SPSS (Statistical Product and Service Solutions/Statistic Package for the Social Science) were used for the analysis. Among them, the three criteria of V/T, Most Frequent Lemmas (with the choice of First-person I/We), and BW were used to measure lexical complexity; MSL was used to measure syntactic complexity. Two dimensions, lexical complexity and syntactic complexity, were used to analyze the correlation between linguistic complexity and the final ranking of speakers.

The data analysis was based on the default of a positive correlation between final ranking and audience recognition in speech contests, i.e., the higher the final ranking, the higher the audience recognition. The conclusions drawn from the experiments were all derived from the 32 data only and no other influencing factors were considered.

3. Results and Analysis

3.1 Negative Correlation between Linguistic Complexity of Speech and Audience Recognition

3.1.1 Vocabulary Richness = V/T

To understand the conceptual difference between "Word Type" and "Word Token", consider the following sentence: "The law is harsh, but it is the law." In this sentence, "the", "is" and "law" all appear twice, so the sentence includes 6 word types and 9 word tokens.

All the different word types form the vocabulary, which is represented by the letter "V". The text length is represented by the total number of tokens, denoted by the letter "T". Thus, the vocabulary richness is calculated as "Vocabulary Richness = V/T". The smaller the value of vocabulary richness, the richer the vocabulary used by the speaker, i.e., the speech is more extensive, or the speaker prefers to use different words to explain a topic and describe it from multiple perspectives. Larger values of vocabulary richness indicate that the speaker uses more focused words, tends to use focused words, and deepens the impression of the audience by repeating and emphasizing certain key information.

The data of "Word Type", "Word Token", and "Vocabulary Richness V/T" of 32 subjects were selected for the analysis and conclusions were drawn.

In terms of vocabulary richness, the researcher found similar patterns between college English speech contests and political speeches. Since the length of speeches in the Foreign Language Institute speech contest was set at three minutes, this reduced the error caused by the lack of uniformity in the length of speeches in previous studies.

By comparing the mean values of "Word Type" and "Word Token" of the four groups, the researcher analyzed the trend of the mean value of vocabulary richness V/T of the four groups. Table 1 shows that the number of word types in groups one to four did not differ significantly in the specified

period. This indicates that the number of word types to be used by speakers at different levels to express their speech topics within the specified time is basically the same. However, because the number of word tokens decreases from the first group at the national level to the fourth group at the preliminary school level, the "V/T" value, which represents vocabulary richness, gradually increases.

Table 1. Average values of "Word Type", "Word Token" and "V/T" of the four groups

Group	1	2	3	4
Word Types (Average)	218.75	218.75	216.625	212.25
Word Tokens (Average)	456.375	437.429	405.125	392
V/T (Average)	0.48	0.499	0.539	0.542

Through the analysis of speech videos, the researcher found that the higher-level speakers had faster oral expressions, more fluent speech, shorter hesitation time, and higher speech articulation. As a result, the speeches of high-level speakers are usually longer in the same amount of time. Their presentation style receives higher ratings from the audience. In addition, the results of the study showed that the higher the final ranking of the speakers, the lower the "V/T value" of their speeches. This suggests that high-level speakers are better at repeating words to emphasize key messages. This shows that repeating keywords and emphasizing central ideas can improve audience recognition. Meanwhile, fluency in speaking, less hesitation, and pauses, and a decent speaking style can also bring a better feeling to the audience. At the same time, fluency in speaking, less hesitation, and a proper presentation style can also bring a better feeling to the audience.

3.1.2 Most Frequent Lemmas: The Choice of the First-person "I/We" as an Example

Most Frequent Lemmas (MFL) is another important measure of the linguistic complexity of a speech. Lemma is the basic form of a word, such as the singular of a noun or the infinitive form of a verb. Research has shown that similar patterns exist in the MFL of college English speech contests and political speeches. The researcher analyzed the 10 most frequently used Lemmas by 32 speakers. The preposition "the" and auxiliary verb (lemma of the word types am, is, are, was, etc.) frequently appeared in the top two. After that, the frequency of the other words varied.

The researcher analyzed the frequency of using the First-person "I/We" for the 32 subjects. "I" over "We" was marked by the number "1", and "We" over "I" was marked by the number "2". If "I" and "We" were used equally often, they were marked by the number "0".

The result of the data analysis shows an interesting fact about the frequency of using the First-person "I/We". Speakers in the first group were more likely to use "I" than "We", while speakers in the fourth group were generally more likely to use "We" in their speeches. It can be seen that the higher-level participants preferred to use "I" rather than "We". This is because the pronoun "We" is somewhat ambiguous (Who is behind the "We"? Me the future government? Me and the people? Me and the workers? Me and the college students as a group?), while the "I" emphasizes more self-awareness. Therefore, in a particular speech contest situation, the First-person "I" is more conducive to expressing one's point of view, giving people a sense of sincerity and determination, and thus enhancing the audience's recognition.

3.1.3 The Frequency of Big Words (BW)

The Frequency of Big Words (BW) is another important measure of linguistic complexity. As an additional way of measuring the lexical complexity of texts, the frequency of big words (consisting of six or more letters, denoted as BW) can be analyzed (Tausczik, 2010). Texts or conversations with a high percentage of BW tend to be more complex and difficult to understand. Research in cognitive science has also shown that basic vocabulary and elementary words are easier to remember. Thus, the use of simple language in oral communication better helps people to remember key information.

This rhetorical phenomenon has also been recognized by former U.S. presidents, such as President Johnson, who told his speechwriters, "I want words of no more than four letters and paragraphs of no more than four sentences to make up my speech."

The data shows that the BWs of the 32 speakers in the speech contest ranged from 17.20% to 26.67%. Meanwhile, Table 2 shows that BW mean of groups one to four demonstrates a gradually increasing trend.

Table 2. BW mean of groups one to four

Group	1	2	3	4
BW (Average)	19.5	20.6	21	24.02

Based on the three indicators of V/T, frequency of First-person choice, and BW, we can see that the higher-level participants are adept at adopting a more direct communication style. They tend to choose concentrated words and simple words to deliver their speeches, repeatedly emphasize keywords to enhance the audience's memory, and express their ideas in a more direct way to increase the audience's recognition. The lower-level participants tend to use a wider range of words and more complex vocabulary, and the choice of First-person is also more ambiguous, which to a certain extent reduces the audience's recognition and is not conducive to achieving a higher ranking in the speech contest.

In conclusion, the linguistic complexity of speeches and audience recognition in college English speech contests show a negative correlation, i.e., the lower the linguistic complexity, the higher the audience recognition, which is similar to the results of previous studies in the field of political speeches.

3.2 Correlation between the Syntactic Complexity of Speech and Audience Recognition and the Reasonable Interval of MSL

The mean sentence length is one of the criteria to measure Syntactic Complexity. This refers to the POS tagger Toutanova's definition in 2003: "The sentence boundaries" correspond to "strong" punctuation marks (i.e., periods, questions, and exclamation points). Usually, a longer sentence is more complex to understand and less memorable for the audience, which is especially true in oral communication.

According to a comparison of historical data trends in political speeches, the average MSL for "State of the Union" speeches delivered by the Founding Fathers was 39.6 (highest MSL: 44.8 tokens/sentence for Madison), and by Obama, the MSL decreased to 18.5 tokens/sentence (Savoy, 2018). These examples clearly show that the style of political speeches changes over time. Currently, there is a preference for shorter sentences that are easier for the audience to understand.

By analyzing the mean sentence length (MSL) of 32 subjects in speech contests and the mean MSL of the four groups, it was found that in college English speech contests, the MSL values of Group I and Group II participants were more similar and fell within the range of 16.43 tokens/sentence to 23.21 tokens/sentence. In contrast, the MSL values in groups three and four had more extreme values, such as the last three participants in the final ranking, whose MSL values were 29.98 tokens/sentence, 31.21 tokens/sentence, and 29.56 tokens/sentence, respectively. It can be seen that in this set of data, the participants whose MSL values are located above 25 are not ranked high. In college English speech, too high MSL values (>25 tokens/sentence) will reduce the audience's recognition, and long sentences are not conducive to the audience's understanding and memory. This is consistent with the negative correlation pattern between Syntactic Complexity of speech and audience recognition in political speeches. In the fourth group, there are also lower MSL extremes, such as 13.8 tokens/sentence, 13.45 tokens/sentence, etc. In contrast, the speaker with 13.67 tokens/sentence has a higher ranking, and a comparative analysis of the text and video of the three speeches reveals that the first two use a large number of simple sentence templates at the beginning and end to begin and end the speech, and there is Part of the hesitation phenomenon, causing broken sentences, etc., affecting the overall fluency of the speech, to a certain extent affect the audience perception. The latter, on the other hand, used rhetorical questions and prose sentences, used short sentences to render emotions, point out the main idea and repeat keywords, which had the effect of enhancing the

audience's recognition and strengthening the momentum. Accordingly, the researcher deduced that the appropriate MSL value in college English speech contests is an interval value, and from the 32 sets of data, MSL values >25 markers/sentence are too high and MSL values <15 markers/sentence are too low, both of which reduce the audience's recognition, and the appropriate MSL value range is between 15 markers/sentence and 25 markers/sentence. However, there are exceptions if the speaker uses appropriate sentence structure or presentation techniques.

Due to the presence of more extreme values in the fourth group, there was no clear pattern between MSL and audience approval between groups one and four. However, as discussed, there is a correlation between individual MSL and audience approval, and there is a reasonable range of MSL.

4. Conclusion

From the perspective of linguistic complexity, this paper explores the correlation between linguistic complexity and audience recognition in college English speech contests. By adopting the corpus construction and computer visualized data analysis, this study analyzes the speech of contestants at different levels in FLTRP Cup National English speaking Contest 2019-2020, the most authoritative college English speech contest in China. The results of the study show that: 1) linguistic complexity of speech is negatively correlated with audience recognition (i.e., final competition ranking or speech success) in college English speech contests. This pattern is similar to the results of previous studies obtained in the field of political speeches. Based on the analysis of three indicators: vocabulary richness (V/T), frequency of choosing First-person "I/WE" and frequency of using big words (BW), the researcher found that the higher-level participants are adept at adopting a more direct communication style. They tend to choose concentrated words and simple words to deliver their speeches, repeatedly emphasize keywords to enhance the audience's memory, and express their ideas in a more direct way to increase the audience's recognition. The lower-level participants tend to use a wider range of words and more complex vocabulary, and the choice of First-person is also more ambiguous, which to a certain extent reduces the audience's recognition and is not conducive to achieving a higher ranking in the speech contest. (2) In college English speech contests, there is a reasonable interval of syntactic complexity of the speech to ensure good audience recognition. MSL > 25 tokens/sentence is too high and MSL < 15 tokens/sentence is too low, both of which will reduce the audience's approval. The appropriate MSL range is between 15 and 25 tokens/sentence. (3) The linguistic complexity and syntactic complexity of speeches in college English speech contests have some similarities to the correlation patterns between the rhetoric of speeches and audience approval in political speeches derived by previous authors in the field of political speeches. (4) This study has some practical value and relevance. The findings of the study can help people predict the winners of speech contests based on the linguistic complexity of their speeches, or help participants better revise their speeches to improve audience recognition and thus increase the chances of winning the contest.

This study also has some limitations. First, the number of data samples contained in the "Unipus Speech Corpus" is relatively small due to the limitation. More samples are needed to explore more accurate patterns and to further precise the ideal MSL interval. Second, the corpus data analysis established in this study is a textual corpus and does not consider other multimodal factors such as pronunciation and gestures that may affect the results of a speech contest. Therefore, the conclusions of this study need to be confirmed by more theoretical and empirical studies.

References

- [1] Norris, J. & Ortega, L. (2009) Towards an organic approach to investigating CAF in instructed SLA: The case of complexity [J]. *Applied Linguistics* 30: 555-578.
- [2] Lu, X. 2011. A corpus-based evaluation of syntactic complexity measures as indices of college level ESL writers' language development [J]. *TESOL Quarterly* 45: 36-62.
- [3] Ortega, L. 2015. Syntactic complexity in L2 writing: Progress and expansion [J]. *Journal of Second Language Writing* 29: 82-94.

- [4] Bulté, B. & A. Housen. 2012. Defining and operationalizing L2 complexity [A]. In A. Housen, F. Kuiken & I. Vedder (eds.). *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA* [C]. Amsterdam: John Benjamins. 21-46.
- [5] Skehan, P. 2009. Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis [J]. *Applied Linguistics* 31: 510-532.
- [6] Wang Haihua, Li Beibei and Xu Lin, 2015, A case study on the development of written language proficiency of Chinese English learners [J], *Foreign Language Teaching and Research* (1):67-80.
- [7] Wang Chuming, 2008, Language Learning and Interaction [J], *Journal of Foreign Languages* (6):53-60.
- [8] Xu Xiaoyan, Wang Weimin, Xiong Yanyu, Jiang Jing, Pan Xiaoyan, and Sun Nianhong, 2013, A study on the syntactic complexity of English argumentative essays among Chinese English majors [J], *Foreign Language Teaching and Research* (2):264-275.
- [9] Lu, X. 2011. A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development [J]. *TESOL Quarterly* 45: 36-62.
- [10] Bulté, B. & A. Housen. 2014. Conceptualizing and measuring short-term changes in L2 writing complexity [J]. *Journal of Second Language Writing* 26: 42-65.
- [11] Crossley, S. & D. McNamara. 2014. Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners [J]. *Journal of Second Language Writing* 26: 66-79.
- [12] Mazgutova, D. & J. Kormos. 2015. Syntactic and lexical development in an intensive English for Academic Purposes programme [J]. *Journal of Second Language Writing* 29: 3-15.
- [13] Bao Gui, 2009, A study on the change of syntactic complexity of English learners' compositions [J], *Foreign Language Teaching and Research* (4): 291-297.
- [14] Wang Min and Wang Chuming, 2014, The synergistic effect of reading followed by writing [J], *Modern Foreign Languages* (4):501-512.
- [15] Wasike, B. 2017. Charismatic rhetoric, integrative complexity and the U.S. Presidency: An analysis of the State of the Union Address (SOTU) from George Washington to Barack Obama [J]. *The Leadership Quarterly* 28: 812-826.
- [16] Gunawan, S. 2017, Hillary Clinton's presidential campaign rhetoric: Making America while again [J]. *Kasetsart Journal of Social Sciences* 38: 50-55.
- [17] Savoy, J. 2018. Analysis of the style and the rhetoric of the 2016US presidential primaries [J]. *Digital Scholarship in the Humanities* 33: 143-159.
- [18] Tausczik, Y. 2010. The psychological meaning of words: LIWC and computerized text analysis methods [J]. *Journal of Language and Social Psychology* 29: 24-54.
- [19] Toutanova, K. 2003. Feature-rich Part-of-speech Tagging with a Cyclid Dependency Network in Proceedings of Human Language Technology-North American Chapter of the Association for Computational Linguistics [J]. *Applied Linguistics* 28: 432-458.