

Fellow or foe? A quantitative thematic exploration into Putin's and Trump's stylometric features

Yaqin Wang¹ , Ting Zeng^{2*} 

¹ Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies

² School of Foreign Languages, College of Arts and Sciences, Shanghai Polytechnic University

* Corresponding author's email: zengting@sspu.edu.cn

DOI: https://doi.org/10.53482/2023_54_406

ABSTRACT

Thematic concentration, a quantitative linguistic method, can reflect the speech style of a particular person. It may, to some degree, reflect the degree of a speaker's intention to communicate certain themes. There has been limited empirical research on the similarity between Trump and Putin with respect to their linguistic features. Thus, the present study aims to compare Putin's and Trump's stylometric features and political themes based on thematic concentration with a corpus of Putin's, Medvedev's, Trump's, and Obama's speeches. Results show that 1) Both Putin's and Trump's speeches' thematic concentration values are significantly or marginally significantly different from their precedents'. 2) Two leaders pay great attention to the concept of nationalism. 3) Thematic words of their speeches were slightly different across periods, reflecting the influence of external factors on the choice of language. The results of the present study may shed light on the stylometric studies of Putin and Trump.

Keywords: thematic concentration; stylometric features; Putin; Trump; authoritarianism

1 Introduction

Notoriously renowned for the slangy, vulgar, and violent political talk since the advent of the key phrase *Mochit' v sortire* 'to kill somebody in a toilet', the Russian politician Vladimir Putin, has attracted much scholarly attention from linguists and discourse analysts (Glukhova and Sorokina 2018; Sedykh 2016). Despite his offensive linguistic style, Putin was also commented on as a politician with a deliberate choice and strategy that serves political ends by legitimizing jargon or semi-jargon language in the official report (Glukhova and Sorokina 2018; Gorham 2014).

Sounds familiar? Commonly regarded as an object of comparison of Putin's political inclination (Hauser 2018), the former American president, Donald J. Trump, and his team also appear to have

employed a deliberate or idiosyncratic campaigning style and rhetoric (Mercieca 2020; Reyes and Ross 2021). More importantly, Putin, on the one hand, was dubbed as an iron-fist father figure with an inevitably authoritarian inclination under the political scheme of Russia (Gorham 2005), with some political scientists defining his political strategies as “Putinism” (Fish 2017). The ideology of Trump (so-called Trumpology, Trumpism), on the other hand, has been regarded as a type of authoritarian leadership principle (Rivers and Ross 2020) as well. Van Dijk (2008) has pointed out that Putin is used to employing a positive self-presentation and a negative presentation of his opponents. Trump, in a similar vein, unsurprisingly resorted to similar construction of a ‘self-versus opponent’ image (Homolar and Scholz 2019; Ross and Caldwell 2020). Despite many similarities between Putin and Trump in terms of speech strategies, empirical linguistic research into the stylometric features has been much more limited. Exploration of this topic may help to clarify the relationship between Putin's and Trump's speech styles.

As one of the important measurements related to content analysis in quantitative linguistics, thematic concentration can indicate the speech style of a writer or speaker (Čech et al. 2015). As Čech (2016, p. 9, cited from Chen and Liu (2018, p. 68) and reformulated by authors) points out,

“the method of measuring thematic concentration can be classified among the types of textual analysis that are generally referred to as content analysis. In its nature, it is also close to quantitative analysis of the so-called ‘keywords analysis’. However, as is evident from the title of this method, its primary aim is ... to reveal the extent to which the author has addressed the topic(s) on the given theme or themes on the whole. From a more general perspective, it is a method for modeling a particular aspect of speech behavior.”

This method has been used in investigating presidential inaugural speeches (Kubát and Čech 2016) and political debates (Savoy 2018). A number of studies have applied it to investigate linguistic features of official reports and political speeches (Čech 2014; Chen and Liu 2015, 2018; Wang and Liu 2018). Further, Čech (2014) reported significant differences in the levels of thematic concentration between Czechoslovak and Czech presidents from the totalitarian period and the period of democracy respectively. He suggested that the level of thematic concentration may, to some degree, indicate a tendency of ideology, be it a more totalitarian (a higher level of thematic concentration) or a more democratic one (a lower level). Wang and Liu (2018) reported a higher level of thematic concentration in Trump's campaign speeches, which is somehow consistent with the previous conclusion of his political inclination toward authoritarianism. These studies highlight the significance of thematic concentration in stylometric analyses.

Thus, the present study intends to compare Putin and Trump's speech style during their presidency based on the quantitative linguistic method, thematic concentration, by employing three indicators, viz., thematic concentration (TC), secondary thematic concentration (STC), and proportional thematic concentration (PTC). Since the value of thematic concentration is closely related to the indicator of h-point

in scientometrics, which is rather sensitive to the language type (Popescu 2009), we compared theirs with those of their respective political predecessors, Medvedev and Obama¹. Two sets of values, the Putin-Medvedev pair and the Trump-Obama pair, were collected. On top of that, thematic words reflecting the political themes of two political figures, namely, the Putin-Trump pair, were compared.

Research questions are as follows:

1. What is the relationship between Putin's and Medvedev's thematic concentration values?
2. What is the relationship between Trump's thematic concentration value compared with Obama's? Further, is Putin's position in the Putin-Medvedev pair different from Trump's in the Trump-Obama pair?
3. What are the thematic words of Putin and Trump, and what are the political themes they intend to emphasize?

The paper's layout is organized as follows: Section 1 introduces the general background information. Section 2 displays the details of the methods and materials employed in the study. Section 3 presents results and discussion, followed by conclusions and suggestions for further research in Section 4.

2 Methods and Materials

2.1 Materials

The organization of linguistic materials is shown in Table 1, 200 texts and 719,894 tokens in total. Putin's and Medvedev's materials were gleaned from the official website of the President of Russia,² and Trump's and Obama's were from the American Presidency Project.³ Each political figure's speeches during their terms in office were chosen, including addresses to the Federal Assembly, or addresses before a joint session of the congress on the State of the Union, news conferences and remarks at special occasions. For each year, 6-14 texts were selected for each person. The composition of the corpus is displayed in Table 1 and specific information, i.e., date, place and theme, of each text is in Appendix A. It should be noted that the authorship of presidents' or political candidates' speeches is always disputable. President, however, is the one who delivers the speech. He is politically responsible for their speeches and thus can affect the text to some degree (Čech 2014).

¹ It would be more reliable to collect more former presidents' texts as the reference corpus. However, Putin has only one predecessor in the last two decades. Thus, we only chose speeches of Medvedev and Obama for comparison. In the future, texts of Russian politicians other than the president can be gleaned to further the research.

² <http://www.kremlin.ru/>

³ <https://www.presidency.ucsb.edu/>

Table 1: The composition of the corpus.⁴

	Addresses to the Federal Assembly/ the State of the Union	News conference	Remarks at special occasions	Time range	Texts	Tokens
Putin	4	5	41	2017-2021 2017: 12 texts 2018: 11 texts 2019: 12 texts 2020-2021: 15 texts	50	157,051
Medvedev	4	4	42	2008-2012 2008: 9 texts 2009: 12 texts 2010: 11 texts 2011: 11 texts 2012: 7 texts	50	126,514
Trump	3	4	43	2017-2021 2017: 13 texts 2018: 12 texts 2019: 11 texts 2020: 14 texts	50	209,225
Obama	7	6	37	2011-2016 2010-2011: 13 texts 2012: 7 texts 2013: 8 texts 2014: 9 texts 2015: 6 texts 2016: 7 texts	50	227,104
Total	18	19	163	/	200	719,894

2.2 Methods

As an approach to measure the degree of the author's intention to communicate certain themes, thematic concentration (TC) was introduced by Popescu (2007) and further developed by a series of works (e.g., Popescu et al. 2009). The computation of TC is based on the concept of the h-point, which was conceived by Hirsch (2005) for scientometrics and then introduced into linguistics by Popescu (2007). If we rank word frequencies of a text in descending order, we can determine the value of the h-point when the rank of a particular word is equal to its occurrence. Figure 1 shows the position of an h-point in a rank-frequency distribution of a certain text.

⁴ As shown in Appendix A, for Putin and Trump, only 2-3 texts were collected in 2021, thus we combine texts of 2020 and 2021 together. This also holds true for the group of 2010-2011 of Obama's texts (only one text was collected in 2010).

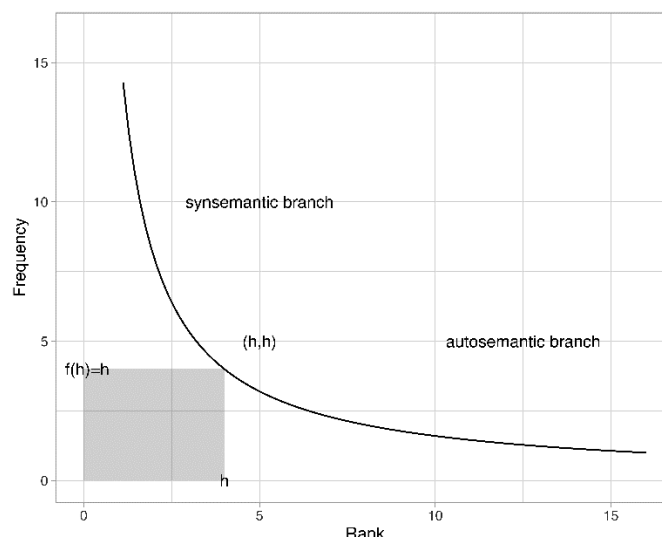


Figure 1: The position of the h-point in a rank-frequency distribution (cited from (Popescu et al. 2009, p. 17).

Popescu et al. (2009) demonstrated that the h-point fuzzily separates the frequent synsemantics (including prepositions, pronouns, particles, articles) from the autosemantics (including nouns, adjectives, and verbs), which build the major vocabulary of the text. Autosemantic words which occur before the h-point indicate that they are frequently used by the author. They represent the text themes (nouns) and descriptions and actions of certain central words (adjectives and verbs). This may signify that the author intends to communicate certain themes with others. The calculation of the h-point in the frequency distribution of lemmas⁵ is shown below (for more details, see Popescu et al. 2009):

$$(1) \quad h = \begin{cases} r_i, & r_i = f(r_i) \\ \frac{f(r_i)r_{i+1} - f(r_{i+1})r_i}{r_{i+1} - r_i + f(r_i) - f(r_{i+1})}, & r_i \neq f(r_i) \end{cases}$$

Based on the value of the h-point, the computation of thematic concentration can be defined as:

$$(2) \quad TC = 2 \sum_{r'}^T \frac{(h-r')f(r')}{h(h-1)f(1)}$$

where $f(1)$ is the frequency of the first rank, T is the number of autosemantics before the h-point, and r' is the average rank ($r' < h$).

⁵ Čech (2014) and Čech et al. (2015) computed the h-point value based on the frequency distribution of lemmas (i.e., canonical forms of words). Hence the current study followed suit.

Table 2: Rank frequency distribution of Putin’s speech of *Парад Победы на Красной площади* ‘Victory Parade on Red Square’ on May 9, 2021.

Rank	Average rank	Frequency	Lemma	English translation
1	1	60	<i>и</i>	and
2	2	20	<i>в</i>	at
3	3	17	<i>наш</i>	our
4	4	14	<i>мы</i>	we
5	5	14	<i>на</i>	on
6	6	10	<i>с</i>	with
7	7.5	9	<i>победа</i>	victory
8	7.5	9	<i>тот</i>	that
9	9.5	8	<i>который</i>	which
10	9.5	8	<i>кто</i>	who
11	11.5	7	<i>к</i>	to
12	11.5	7	<i>народ</i>	people
13	14	6	<i>быть</i>	be
14	14	6	<i>война</i>	war
15	14	6	<i>для</i>	for
16	18	5	<i>весь</i>	all
17	18	5	<i>год</i>	year
18	18	5	<i>за</i>	for
19	18	5	<i>по</i>	by
20	18	5	<i>сила</i>	power
21	21	4	<i>великий</i>	great

For example, the rank frequency distribution of Putin’s speech of *Парад Победы на Красной площади* ‘Victory Parade on Red Square’ is displayed in Table 2. As Table 2 shows, there is no rank of a lemma that exactly equals its corresponding frequency, thus we calculate it by the second part of the Formula (1):

$$h_{sample\ text} = \frac{9 * 9 - 8 * 8}{9 - 8 + 9 - 8} = 8.5$$

Thus, there is one autosemantic word which lies in the pre-h domain, i.e., *победа* ‘victory’, as shown in Table 2. The TC value is calculated as follows according to Formula (2):

$$TC_{sample\ text} = 2 * \left(\frac{(8.5 - 7.5) * 9}{8.5 * (8.5 - 1) * 60} \right) = 0.0047$$

A problem occurs when the TC value of a certain text is 0, which poses a challenge for comparing thematic differences between texts. Therefore, Čech et al. (2015) proposed the indicator of secondary thematic concentration (STC) by doubling the h point.

$$(3) \quad STC = \sum_{r'=1}^{2h} \frac{(2h-r')f(r')}{h(2h-1)f(1)}$$

The STC value of the sample text in Table 2 is displayed as well. 2h point of the text is $8.5 \times 2 = 17$, and there are three autosemantics before 2h point. STC value is:

$$STC = \frac{(17 - 7.5) * 9}{8.5 * (17 - 1) * 60} + \frac{(17 - 11.5) * 7}{8.5 * (17 - 1) * 60} + \frac{(17 - 14) * 6}{8.5 * (17 - 1) * 60} = 0.0174$$

The third formula is called proportional thematic concentration (PTC). It is proposed to eliminate the circumstance where there is only one content word in the pre-h domain in a text (Čech et al. 2015). It is computed as:

$$(4) \quad PTC = \frac{1}{N_h} \sum_{r' < h} f(r')$$

N_h refers to the frequency of all words r_1, \dots, r_h , in the pre-h domain, the sum of $f(r')$ is the frequency of all autosemantic words occurring before the h point. PTC value of the sample text is:

$$PTC = \frac{9}{153} = 0.0588$$

In sum, a higher level of TC, STC, and PTC signify the author's effort in communicating more intensive certain themes with others, while the lower one suggests the diversity of one's themes.

As argued by Čech (2016), TC and STC values are independent of text length of the range $\langle 200, 6500 \rangle$. PTC values are said not to be a suitable tool for comparing texts with a length of $N < 2000$ words. In the present study, the lengths of most texts (171 texts) roughly fall into the interval of $\langle 200, 6500 \rangle$ and more than half of texts' (116 texts) lengths are greater than 2000. Thus, to investigate the influence of text size which may exert on indicators, we carried out three Pearson tests between TC, STC, and PTC and the text size. Results show that the correlation coefficient between TC, STC, and PTC and the text size is low (Pearson $r = -0.13, -0.33, 0.04$ respectively). Thus, we can compare indicators of texts with different sizes.

3 Results and Discussion

This section discusses quantitative results and possible factors for those phenomena. The comparisons of TC, STC, and PTC values in the Putin-Medvedev pair and the Trump-Obama pair are carried out, followed by analyses of the thematic words of the Putin-Trump pair. In both the comparison of thematic concentration and that of thematic words, diachronic comparisons or analyses of their speeches are shown after the general discussion.

3.1 Comparison of Putin's and Medvedev's thematic concentration

Table 3 displays descriptive statistics of three indicators from two Russian presidents during their terms of office.

Table 3: Descriptive statistics of three indicators of thematic concentration of Putin-Medvedev pair.

	TC	STC	PTC
Putin			
Min.	0	0.0054	0
First quartile	0.0098	0.0260	0.0717
Median	0.0234	0.0411	0.1118
Mean	0.0321	0.0432	0.1215
Third quartile	0.0403	0.0562	0.1613
Max.	0.1374	0.1081	0.3060
Standard Deviation	0.0318	0.0222	0.0731
Medvedev			
Min.	0	0.0028	0
First quartile	0.0018	0.0265	0.0340
Median	0.0203	0.0340	0.0878
Mean	0.0201	0.0363	0.0835
Third quartile	0.0324	0.0431	0.1247
Max.	0.0783	0.0937	0.2171
Standard Deviation	0.0190	0.0174	0.0612

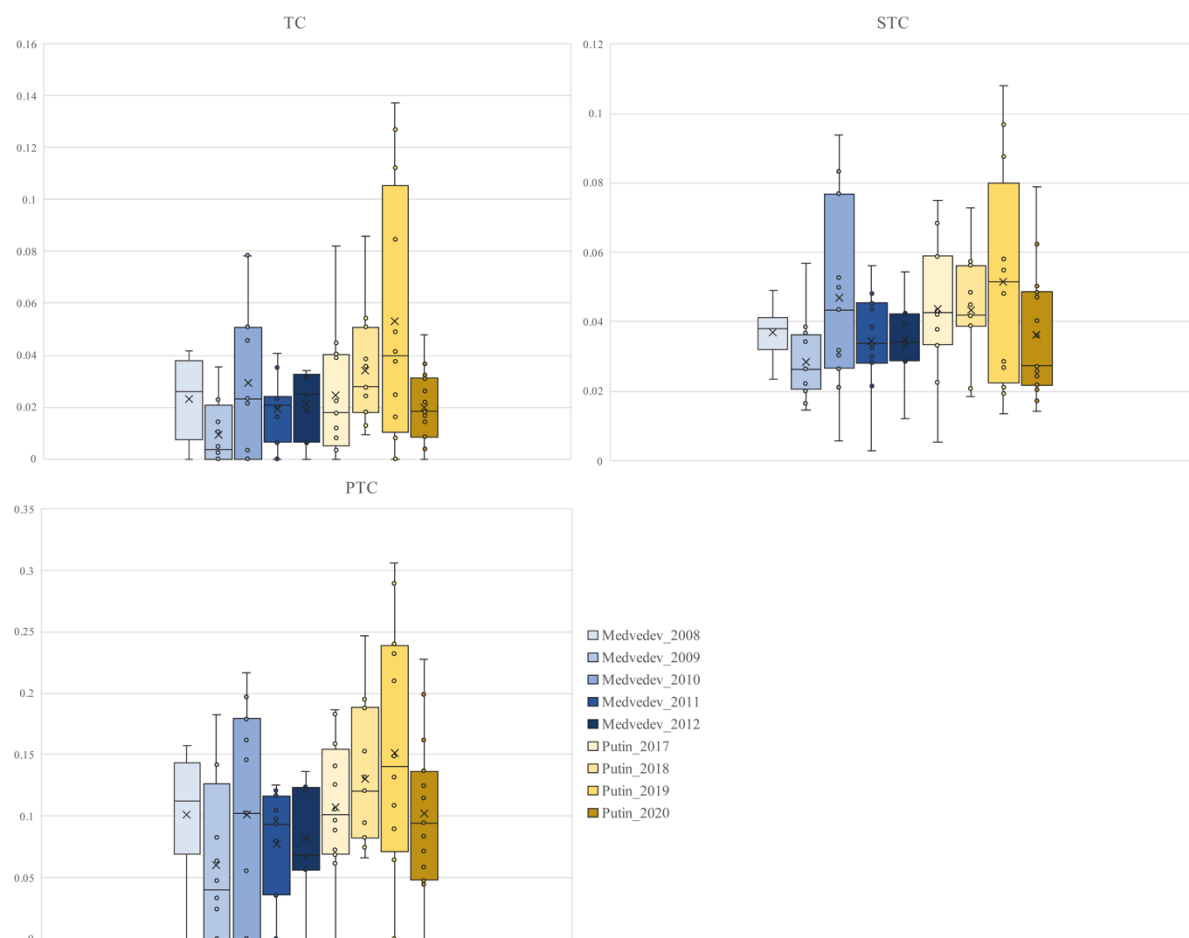


Figure 2: The distribution of three indicators of thematic concentration in the Putin-Medvedev pair. Boxes are the distribution of TC/STC/PTC values of two people each year as the legend displays. The blue series of boxes represents Medvedev's indicators and the yellow one is Putin's. The label "x" on each plot is the average value of each distribution.

As Table 3 and Figure 2 show, the average values of all indicators of Putin are greater than those of Medvedev. Then, a non-parametric test (Mann Whitney U) was carried out on TC values and two t-tests on STC and PTC values, respectively (since the set of TC values does not follow the normal distribution).

Results of the Mann Whitney U test show that values of Putin's TC ($Mdn = 0.0272$) is marginally significantly different from those of Medvedev's ($Mdn = 0.0220$, $U = 966$, $p = .05 < .1$). Regarding the values of the other two indicators, results of t-tests for two independent samples demonstrate that the difference between Putin's STC ($M = 0.0432$, $SD = 0.0222$) is marginally significant from Medvedev's ($M = 0.0363$, $SD = 0.0174$) values ($t(98) = 1.738$, $p = .085 < .1$). PTC values of Putin ($M = 0.1215$, $SD = 0.0731$) are significantly greater than those of Medvedev ($M = 0.0835$, $SD = 0.0612$, $t(98) = 2.823$, $p = .006 < .05$). As Figure 2 shows, most of PTC values are higher in Putin's speeches.

These results indicate that regarding three indicators, Putin's thematic concentration is significantly (or marginally significantly) greater than Medvedev's. Three metrics, especially PTC values, can distinguish two people's degrees of thematic concentration. This implies that, to some degree, Putin's intention to communicate some topics is greater than that of Medvedev. In other words, his discursive practice contains relatively more central themes, while his predecessor's speeches reflect the diversity of themes.

As Čech et al. (2015) commented, texts with $STC < TC$ can be regarded as extremely concentrated texts. We counted the number of texts and found that 9 texts of Putin's meet this requirement, while only 5 ones of Medvedev's do. This shows that, compared with Medvedev's speeches, more of Putin's speeches reach the extreme end of thematic concentration. This, additionally, reflects Putin's intense intention of communicative practice.

Diachronically, we compared Putin's speeches according to chronological order, i.e., based on four sets of speeches ranging from 2017 to 2021. Results of a One-way ANOVA test show no significant differences among speeches from 2017 to 2021 for STC and PTC values ($p_{stc} = .364$, $p_{ptc} = .293$). TC values show significant differences among different periods ($F(3, 46) = 3.123$, $p = .035 < .05$), however, the post-hoc test shows that only TC values of 2019 are significantly different from those of 2020-2021. It can be seen that in Figure 2, TC values for 2019 are greater than those for 2020-2021. This indicates that Putin did show differences across different periods diachronically in terms of the degree of concentration on certain themes.

3.2 Comparison of Trump's and Obama's thematic concentration

Table 4 displays descriptive statistics of three measurements from Trump and Obama during their terms of office.

Table 4: Descriptive statistics of three measurements of the thematic concentration of the Trump-Obama pair.

	TC	STC	PTC
Trump			
Min.	0	0.0115	0
First quartile	0.0035	0.0183	0.0335
Median	0.0095	0.0209	0.0560
Mean	0.0114	0.0225	0.0619
Third quartile	0.0165	0.0256	0.0892
Max.	0.0668	0.0529	0.1361
Standard Deviation	0.0114	0.0076	0.0376
Obama			
Min.	0	0.0025	0
First quartile	0.0021	0.0101	0.0206
Median	0.0045	0.0138	0.0361
Mean	0.0054	0.0143	0.0363
Third quartile	0.0080	0.0174	0.0538
Max.	0.0216	0.0323	0.0999
Standard Deviation	0.0046	0.0060	0.0251

As Table 4 shows, the mean values of three indicators of Trump are greater than those of Obama. Figure 3 displays that most of Trump's TC, STC, and PTC values are greater than those of Obama. Then, one non-parametric test (Mann Whitney U) was carried out on TC values and two t-tests on STC and PTC values, respectively (since only the set of TC values followed the normal distribution).

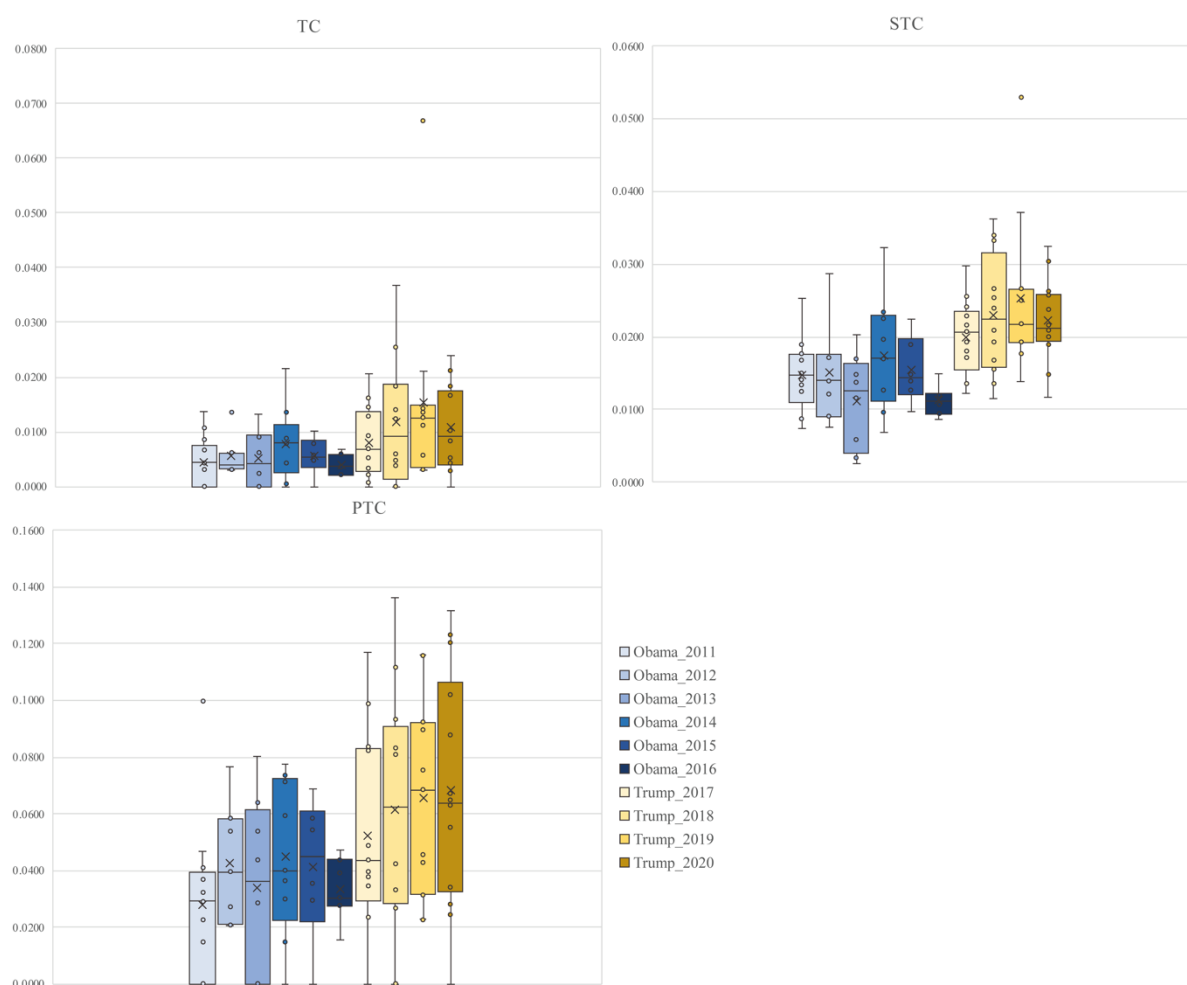


Figure 3: The distribution of three measurements of thematic concentration in the Trump-Obama pair. Boxes are the distribution of TC/STC/PTC values of two people each year as the legend displays. The blue series of boxes represents Obama's indicators and the yellow one represents Trump's. The label "x" on each plot is the average value of each distribution.

Results of the Mann Whitney U test show that values of Trump's TC ($Mdn_{Trump} = 0.0095$) is significantly different from those of Obama's ($Mdn_{Obama} = 0.0045$, $U = 794$, $p = .002 < .01$). Regarding the values of the other two indicators, results of t-tests for two independent samples demonstrate that Trump's STC ($M = 0.0225$, $SD = 0.0076$) is significantly greater than Obama's ($M = 0.0143$, $SD = 0.0060$) values ($t(98) = -5.997$, $p < .0001$). PTC values of Trump ($M = 0.0619$, $SD = 0.0376$) are significantly greater than those of Obama ($M = 0.0363$, $SD = 0.0251$) ($t(98) = -4.013$, $p < .0001$).

These results indicate that Trump's thematic concentration is significantly greater than Obama's regarding three indexes. This implies that, to some degree, Trump's intention to convey certain themes is greater than that of Obama; in other words, his speeches contain relatively more central themes while those of his predecessor reflect the diversity of themes. Here, levels of three indicators in Trump's speeches are significantly higher than those of Obama, suggesting his preference for an authoritarian leadership style. This result is consistent with Wang and Liu (2018)'s findings that the significantly greater TC levels of Trump's campaign speeches than those of Obama and Clinton.

Furthermore, in addition to the level of TC (in Wang and Liu's (2018) research), STC and PTC values applied in the current research also demonstrate a similar tendency of the distribution.⁶ Results suggest that in addresses and remarks other than campaign speeches, Trump, as usual, demonstrates the tendency of concentrating on a handful of political themes. Moreover, 3 texts of Trump whose STC value is smaller than the TC value, while none of Obama's texts does so. As mentioned in Putin-Medvedev pair, 9 texts of Putin whose STC value is smaller than the TC value and 5 for Medvedev. This shows that generally speaking, Russian presidents' extremeness of TC is more evident than that of American presidents.

Likewise, we carried out a statistical test on values of three metrics across different periods. Results demonstrate no significant differences among speeches from 2017 to 2021 for Trump ($p_{tc} = .480$, $p_{stc} = .402$, $p_{ptc} = .718$). This indicates that diachronically, Trump did not show obvious differences in terms of the degree of concentration on certain themes. Compared with Putin's results, Trump's intention to convey certain themes remains consistent no matter when the speech was delivered.

Together with what we have discussed so far, the values of TC, STC, and PTC in Putin's texts are significantly or marginally significantly greater than those of Medvedev's speeches; in a similar vein, those of Trump's are significantly greater than those of Obama's. Both Trump and Putin tend to concentrate on certain central themes compared with their predecessors. Čech (2014) suggested that the level of thematic concentration may, to some degree, indicate a tendency toward ideology. As noted by political scientists, e.g., Medvedev attempted to employ moderate reformism by promoting economic modernization and political liberalization (Noriega 2016). On the contrary, commonly reported as a strong leader with an iron fist, Putin is famous for his so-called father-figure leadership style. As for Trump, During the 2016 election, his authoritarian tendency has been one of the key factors in his winning the presidency (MacWilliams 2016; Homolar and Scholz 2019). The theme intensity of Putin and Trump may reflect their authoritarian leadership style to some extent. Future research, however, including more presidents and texts, is needed to explore this relationship.

⁶ Since Čech (2014) and Wang and Liu (2018) only investigated the level of TC in speeches, this somehow suggests the applicability of STC and PTC values in the thematic concentration comparison.

More importantly, Trump's differences from his former president are more obvious than those of the Putin-Medvedev comparison. This reflects Trump's peculiarities again compared with traditional politicians. Davis (2020: 77) suggested, "neither Trump nor Putin made explicitly calls for authoritarianism..., despite evidence suggesting otherwise." By analyzing the political speeches of the two presidents, Davis then concluded that, though in their idiosyncratic ways, Trump and Putin constructed a kind of power centered around themselves, reflecting features of authoritarian leaders. This, to some extent, indicates that those two political leaders share a similar tendency from the aspect of thematic concentration.

Let us hence propose a question further, what are their thematic words and what kind of political themes do they want to emphasize?

3.3 Comparison of Putin's and Trump's thematic nouns

Due to limited space and the fact that nouns reflect political themes better, we only gleaned thematic nouns based on TC. The total frequencies of words (Frequency), the number of texts they occurred in (Occurrence), and the average value of ranks (Average rank) among all occurrences are shown in Table 5 and Table 6.

Table 5: The relevant information on Putin's thematic nouns.

	Thematic word	Translation	Frequency	Occurrence	Average rank
1	<i>год</i>	year	741	15	11
2	<i>человек</i>	man	544	13	20
3	<i>страна</i>	country	440	14	15
4	<i>Россия</i>	Russia	401	14	10
5	<i>вопрос</i>	question	216	3	35
6	<i>всё</i>	everything	177	3	43
7	<i>развитие</i>	development	166	6	19
8	<i>семья</i>	family	74	2	27
9	<i>работа</i>	work	71	3	19
10	<i>регион</i>	region	62	3	12
11	<i>процент</i>	percent	61	1	38
12	<i>система</i>	system	61	1	24
13	<i>решение</i>	solution	57	1	42
14	<i>восток</i>	East	44	1	15
15	<i>сотрудничество</i>	cooperation	44	3	9
16	<i>оружие</i>	weapon	42	1	33
17	<i>господин</i>	Sir	31	1	12
18	<i>гражданин</i>	citizen	31	1	29
19	<i>эпидемия</i>	epidemic	29	1	18
20	<i>отношение</i>	attitude	26	2	8

21	<i>коллега</i>	colleague	23	1	17
22	<i>проблема</i>	problem	22	1	15
23	<i>экономика</i>	economy	21	2	9
24	<i>интеллект</i>	intelligence	20	1	4
25	<i>Сербия</i>	Serbia	19	1	4
26	<i>Африка</i>	Africa	18	1	3
27	<i>бизнес</i>	business	18	1	8
28	<i>война</i>	war	18	2	6
29	<i>Монголия</i>	Mongolia	18	1	3
30	<i>государство</i>	state	17	1	10
31	<i>соотечественник</i>	compatriot	16	1	6
32	<i>ООН</i>	UN	15	1	8
33	<i>прокуратура</i>	Prosecutor's office	15	1	6
34	<i>спорт</i>	sports	15	1	4
35	<i>лауреат</i>	laureate	13	1	5
36	<i>право</i>	the right	13	1	8
37	<i>премия</i>	prize	12	1	9
38	<i>учитель</i>	teacher	12	1	3
39	<i>число</i>	number	12	1	10
40	<i>победа</i>	victory	9	1	7
41	<i>организация</i>	organization	8	1	7
42	<i>двадцатка</i>	G20	7	1	5

Table 6: The relevant information on Trump's thematic nouns.

	Thematic word	Frequency	Occurrence	Average rank
1	people	766	16	22
2	country	296	8	24
3	America	160	6	16
4	ballot	119	2	25
5	election	116	2	25
6	year	106	3	24
7	nation	92	3	15
8	vote	89	2	31
9	state	85	2	32
10	United	71	3	14
11	States	61	3	17
12	thing	58	1	38
13	Israel	43	1	28
14	Korea	38	1	12
15	tax	37	1	15
16	voter	34	1	30
17	drug	33	1	12

18	Coast	32	1	16
19	Guard	32	1	17
20	Lou	30	1	14
21	Afghanistan	25	1	13
22	price	25	1	18
23	vaccine	25	1	22
24	Dame	21	1	17
25	Notre	21	1	18
26	trade	21	1	17
27	God	19	1	13
28	Matt	18	1	14
29	Justice	17	1	15
30	virus	14	1	13

As shown in Table 5 and 6, it can be seen that in the past five years, there exist similarities and differences between two people's thematic words. For Putin, themes addressed most prominently mainly include the concept of nation and people (*человек* 'man', *Россия* 'Russia', *страна* 'country', *семья* 'family', *гражданин* 'citizen'), socio-economic development (*развитие* 'development', *экономика* 'economy', *бизнес* 'business'), other nations and foreign policy (*Сербия* 'Serbia', *Африка* 'Africa', *Монголия* 'Mongolia', *ООН* 'UN', *двадцатка* 'G20', *сотрудничество* 'cooperation'), security and wars (*оружие* 'weapon', *война* 'war'), epidemic (*эпидемия* 'epidemic'), etc. Putin focused on the idea of a strong, secure Russia (Davis, 2020), which is consistent with the most frequent thematic nouns (*человек* 'man', *страна* 'country', *Россия* 'Russia'). For Trump, he intensified topics related to nation and people (*people*, *country*, *America*, *nation*, *United States*) as well, election (*ballot*, *election*, *vote*, *voter*), economy (*tax*, *trade*, *price*), epidemic (*vaccine*, *virus*), social policy (*state*, *drug*), foreign policy (*Israel*, *Korea*, *Afghanistan*, *guard*), etc. The first two thematic words are consistent with the most frequent content words in his campaign corpus (Homolar and Scholz 2019).

Specifically, four words, namely, *год* 'year', *человек* 'man', *Россия* 'Russia', *страна* 'country', are the most frequent thematic words and occurred in more than 10 texts in Putin's speeches, while total frequencies of *people*, *country*, *America* rank the first three positions for Trump's texts, occurring in 5 or more texts. Both Trump and Putin emphasize the issues related to people and country. The concept of people is one of the basic concepts of political discourse (Yakoba 2017) and is often used as a tool of political manipulation. As stated by Yakoba (2017: 167), in several speeches delivered by Trump, no matter which topic he was talking about, "by emphasizing on the importance of the people, Trump...constructs a basis for creating an impression of concern for the nation." This works well in Putin's case, too, as he addressed the issue of people intensively.

As for the diachronic change in two presidents' thematic words, we calculated total frequencies of words and the number of texts they occurred in each year. As shown in Appendix B, the most frequent thematic nouns in Putin's texts from 2017 to 2019 are *год* 'year' while the most frequent one is *страна* 'country' in the year 2020 and 2021. The theme of nation and people (*Россия* 'Russia', *страна* 'country', *человек* 'man') ranks in the first several positions for five years, which again highlights Putin's intention on emphasizing the concept of country and people when addressing to his audience.

In 2017, development and security (*развитие* 'development', *оружие* 'weapon') were given enough attention, in 2019, the topics on global issues and foreign policy (*восток* 'East', *регион* 'region', *Сербия* 'Serbia', *Африка* 'Africa', *Монголия* 'Mongolia', *сотрудничество* 'cooperation') were repeatedly mentioned by Putin. When in 2020 and 2021, the period of COVID 19, the theme related to the pandemic and socio-economic development (*эпидемия* 'epidemic', *проблема* 'problem', *экономика* 'economy') was mentioned for many times.

For Trump, as shown in Appendix C, the most frequent thematic noun is always *people* from 2017 to 2021. The concept of nation and people (*country*, *America*, *United States*), in addition, is highlighted in 2017 and 2018. Apart from that, global and economic issues (*Korea*, *Afghanistan*, *trade*, *tax*) were emphasized by Trump in 2017 and 2018. In 2019, the concept of people remained to be concentrated by him while the intensity of the concept of nation and country decreased to some degree. In fact, during the 2020 and 2021, i.e., the 2020 US presidential election, Trump turned to topics serving his own political ends, which are essential for promoting himself, viz., the election (*ballot*, *election*, *vote*, *voter*). In contrast, the issue of pandemics (*vaccine*, *virus*) seems to be given less attention. His intensity revolved around the election, or more specifically, legal vs. illegal ballots, the issue he valued much more than the epidemic.

4 Conclusion

In sum, the present study explored the intensity of thematic concentration of Russian and American presidents using quantitative linguistics methods and qualitative analysis. Values of thematic concentration, secondary thematic concentration, and proportional thematic concentration of Putin's speeches are significantly or marginally significantly different from those of Medvedev's texts. All of Trump's three indicators are significantly greater than those of Obama. Diachronically, Putin's speeches contain more central themes in 2019 than in 2020-2021. By contrast, Trump remains a consistent tendency toward conveying a small number of themes in his communicative practice.

The quantitative-linguistic method, thematic concentration, employed in the current study may gain insight into the relationship between Trump and Putin and their predecessors, Obama and Medvedev, respectively, in terms of their choice of language. This also reflects the feasibility of combining the quantitative linguistic metric, thematic concentration in discourse analysis and stylistic studies. Further

research on thematic words can be conducted, such as words, synonyms, and their references to a greater set (or list), usually called *hreb*, proposed by Ziegler and Altmann (2002).

5 Acknowledgements

The authors would like to thank the anonymous reviewers for their helpful suggestions and remarks. This study is supported by the Social Science Foundation of the Education Ministry of China (Grant No. 21YJC740060), the Social Science Foundation of the Education Ministry of China (Grant No. 19YJC740003), the MOE Project of Key Research Institute of Humanities and Social Sciences at Universities in China (22JJD740018), the Philosophy and Social Science Planning Project of Guangdong Province (Grant No. GD23YWY02).

6 References

- Čech, R., Garabík, R., Altmann, G. (2015). Testing the thematic concentration of text. *Journal of Quantitative Linguistics* 22(3), pp. 215–232. <https://doi.org/10.1080/09296174.2015.1037157>.
- Čech, R. (2014). Language and ideology: Quantitative thematic analysis of New Year speeches given by Czechoslovak and Czech presidents (1949–2011). *Quality & Quantity* 48(2), pp.899–910. <https://doi.org/10.1007/s11135-012-9811-3>.
- Čech, R. (2016). *Tematická koncentrace textu v češtině* [The thematic concentration of the text in Czech language]. Praha: ÚFAL.
- Chen, R., Liu, H. (2015). Ideologies of supreme court justices: Quantitative thematic analysis of multiple opinions of “Bush V. Gore 2000.” *Glottology* 6(2), pp. 299–322.
- Chen, R., Liu, H. (2018). Thematic concentration as a discriminating feature of text types. *Journal of Quantitative Linguistics* 25(1), pp. 53–76. <https://doi.org/10.1080/09296174.2017.1339441>.
- Davis, C. M. (2020). *Presidential Authoritarianism in the United States and Russia During the Metamodern Era*. Fort Lauderdale, FL: Nova Southeastern University dissertation.
- Fish, M. S. (2017). The Kremlin Emboldened: What Is Putinism? *Journal of Democracy* 28(4), pp. 61–75.
- Glukhova, I., Sorokina, O. (2018). Linguistic implementation of Persuasive Strategies and Tactics in Political Public Communication (a case study of V. Putin’s election campaign speeches). In: Uslu, F., Güçlü, T., Özdemir, M., Altan, K., Aslan, S. (Eds.). *Proceedings of INTCESS2018-5th International Conference on Education and Social Sciences, 5-7 February 2018, Istanbul, Turkey*.
- Gorham, M. S. (2005). Putin’s Language. *Ab Imperio* 2005(4), pp. 381–401.
- Gorham, M. S. (2014). *After Newspeak: Language Culture and Politics in Russia from Gorbachev to Putin*. Ithaca, NY: Cornell University Press.

- Hauser, M.** (2018). Metapopulism in-between democracy and populism: transformations of Laclau's concept of populism with Trump and Putin. *Distinktion: Journal of Social Theory* 19(1), pp. 68–87. <https://doi.org/10.1080/1600910X.2018.1455599>.
- Hirsch, J. E.** (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences* 102(46), pp. 16569–16572. <https://doi.org/10.1073/pnas.0507655102>.
- Homolar, A., Scholz, R.** (2019). The power of Trump-speak: Populist crisis narratives and ontological security. *Cambridge Review of International Affairs* 32(3), pp. 344–364. <https://doi.org/10.1080/09557571.2019.1575796>.
- Kubát, M., Čech, R.** (2016). Quantitative Analysis of US Presidential Inaugural Addresses. *Glottometrics* 34, pp. 14–27.
- MacWilliams, M. C.** (2016). Who decides when the party doesn't? Authoritarian voters and the rise of Donald Trump. *PS: Political Science & Politics* 49(4), pp. 716–721. <https://doi.org/10.1017/S1049096516001463>.
- Mercieca, J.** (2020). *Demagogue for president: the rhetorical genius of Donald Trump*. College Station, TX: Texas A&M University Press.
- Noriega, A. C.** (2016). The Putin System: Russian Authoritarianism Today. *Revista Mexicana de Análisis Político y Administración Pública* 5(1), pp. 75–92.
- Popescu, I. I., Altmann, G., Grzybek, P., Jayaram B. D., Köhler, R., Krupa, V., Macutek, J., Pustet, R., Uhlirva, L., Vidya, M. N.** (2009). *Word frequency studies*. Berlin & New York: De Gruyter Mouton.
- Popescu, I. I.** (2007). Text ranking by the weight of highly frequent words. In Peter Grzybek & Reinhard Köhler (Eds.). *Exact Methods in the Study of Language and Text: Dedicated to Gabriel Altmann on the Occasion of his 75th Birthday*, pp. 555–566. Berlin & Boston: De Gruyter Mouton.
- Popescu, I. I.** (2009). Thematic concentration of the text. In: Popescu, I., Altmann, G., Grzybek, P., Jayaram B. D., Köhler, R., Krupa, V., Macutek, J., Pustet, R., Uhlirva, L., Vidya, M. N. (Eds.). *Word Frequency Studies*, pp. 95–100. Berlin & New York: De Gruyter Mouton.
- Reyes, A., Ross, A.** (2021). From the White House with anger: Conversational features in President Trump's official communication. *Language & Communication* 77, pp. 46–55. <https://doi.org/10.1016/j.langcom.2020.12.003>.
- Rivers, D. J., Ross, A. S.** (2020). Authority (de) legitimation in the border wall Twitter discourse of President Trump. *Journal of Language and Politics* 19(5), pp. 831–856. <https://doi.org/10.1075/jlp.19105.riv>.
- Ross, A. S., Caldwell, D.** (2020). 'Going negative': An APPRAISAL analysis of the rhetoric of Donald Trump on Twitter. *Language & Communication* 70, pp. 13–27. <https://doi.org/10.1016/j.langcom.2019.09.003>.
- Savoy, J.** (2018). Analysis of the style and the rhetoric of the 2016 US presidential primaries. *Digital Scholarship in the Humanities* 33(1), pp. 143–159. <https://doi.org/10.1093/lc/fqx007>.
- Sedykh, A. P.** (2016). К вопросу об идиополитическом дискурсе В.В. Путина [On the problem of the idiopolitical discourse of V.V. Putin in the Russian language]. *Political linguistics* 1(55), pp. 35–41.

Van Dijk, T. A. (2008). *Discourse and power*. New York: Macmillan International Higher Education.

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), pp. 299–323. <https://doi.org/10.1177/0957926517734659>.

Yakoba, I. A. (2017). Деконструкция дискурса Дональда Трампа (на примере его предвыборных выступлений 2016 г.) [Deconstruction of Donald Trump discourse (on example of his pre-election speeches in 2016) in the Russian language]. *Discourse-Pi* 14(1), pp. 164–170.

Ziegler, A., Altmann, G. (2002). *Denotative Textanalyse: ein textlinguistisches Arbeitsbuch*. Vienna: Praesens.

7 Appendix

Due to limited space, appendix can be found at

https://osf.io/xf38v?view_only=68b2f6e335aa4dd8b661209e2e29a889