

## SYNTACTIC COMPLEXITY AND POLITICAL IDEOLOGY: A STUDY OF CZECHOSLOVAK AND CZECH PRESIDENTIAL SPEECHES

MIROSLAV KUBÁT<sup>1</sup> – MICHAELA NOGOLOVÁ<sup>2</sup> – XINYING CHEN<sup>3</sup>  
– ŽANETA STIBORSKÁ<sup>4</sup>

<sup>1</sup>Department of Czech Language, Faculty of Arts, University of Ostrava, Ostrava, Czech Republic (ORCID: 0000-0002-3398-3125)

<sup>2</sup>Department of Czech Language, Faculty of Arts, University of Ostrava, Ostrava, Czech Republic (ORCID: 0000-0001-7604-9765)

<sup>3</sup>Department of Czech Language, Faculty of Arts, University of Ostrava, Ostrava, Czech Republic (ORCID: 0000-0002-5052-4991)

<sup>4</sup>Department of Czech Language, Faculty of Arts, University of Ostrava, Ostrava, Czech Republic (ORCID: 0009-0005-5137-5130)

KUBÁT, Miroslav – NOGOLOVÁ, Michaela – CHEN, Xinying – STIBORSKÁ, Žaneta: Syntactic Complexity and Political Ideology: A Study of Czechoslovak and Czech Presidential Speeches. *Journal of Linguistics*, 2025, Vol. 76, No 1, pp. 345 – 354.

**Abstract:** This study analyzes syntactic complexity in Czechoslovak and Czech presidential speeches, using a corpus of New Year's addresses spanning nearly a century. Applying quantitative stylometric methods, we measure average sentence length (ASL), average clause length (ACL), mean dependency distance (MDD), and mean hierarchical distance (MHD) to compare syntactic structures across democratic and communist regimes. The results show that democratic presidents generally use more complex sentence structures, while communist-era speeches are syntactically simpler. However, individual differences are also observed inside groups. Husák exhibits higher complexity among communist leaders, and modern democratic presidents (Klaus, Zeman, Pavel) show a trend toward simplification. These findings confirm ideological influences on presidential rhetoric and highlight a broader shift toward linguistic accessibility in contemporary political speech.

**Keywords:** political discourse, syntactic complexity, presidential speeches, ideology, Czechoslovakia

### 1 INTRODUCTION

The analysis of political speeches has a well-established tradition in stylometry, with presidential addresses receiving particular attention. As pivotal rhetorical events, these speeches not only convey national visions and political ideologies but also reflect broader linguistic and social dynamics. Research in this field has primarily examined lexical features, thematic content, and discourse structures, with U.S. presidential speeches dominating the literature (e.g. Liu 2012; Lim 2004; Savoy 2010, 2016). Beyond the United States, scholars have explored political discourse in other national contexts, such as Italy (Tuzzi, Popescu and Altmann 2010) or Russia

(Kuznetsova 2016), demonstrating how political communication adapts to historical and ideological shifts.

This study examines New Year's presidential speeches delivered by Czechoslovak and Czech presidents, a unique and continuous corpus spanning nearly a century. These annual addresses provide an opportunity for heads of state to engage with the public, summarize national progress, and outline future policy directions. The longevity and regularity of these speeches make them an ideal dataset for investigating the changes of linguistic patterns in political rhetoric over time.

Previous research on Czechoslovak and Czech presidential speeches has primarily focused on lexical and thematic analyses. Čech (2014) conducted a quantitative study analyzing the thematic concentration of these speeches, arguing that totalitarian and democratic leaders exhibited distinct linguistic patterns due to their differing ideological orientations. Additional studies, such as David et al. (2013), explored thematic structuring and ideological markers in presidential rhetoric. Čech (2011) analyzed the frequency structure of these speeches, demonstrating how certain lexical characteristics remain stable while others change in response to political shifts. A related study by Jičínský and Marek (2017) combined phonetic and textual analyses to investigate pronunciation tendencies and stylistic variations in Czech presidential speeches. Their findings indicated that ideological shifts influenced word choice and affected voice characteristics as well.

Building on these findings, Kubát, Mačutek and Čech (2021) further examined structural patterns of presidential speeches, employing multiple quantitative measures such as moving-average type-token ratio, mean verb distance, and cluster analysis of frequently used words. Their research confirmed that democratic-era speeches displayed greater lexical diversity, while communist-era addresses exhibited repetitive patterns, aligning with ideological constraints of the time. These studies underscore the importance of linguistic analysis in political speech research, illustrating how stylistic and structural patterns are shaped by ideological contexts. Despite the wealth of research on lexical and thematic aspects of presidential speeches, syntactic complexity remains an underexplored dimension, particularly within the context of Czechoslovak and Czech presidential rhetoric. This study seeks to address this gap by providing a comprehensive syntactic analysis, offering new insights into the relationship between political ideology and linguistic structuring in presidential discourse.

Syntactic complexity, broadly defined as the structural sophistication of sentences, plays a crucial role in shaping political communication. More complex syntactic structures may signal rhetorical sophistication, authority, and persuasive intent, while simpler structures may reflect an effort to communicate more directly with a broader audience. Research in political communication suggests that

authoritarian regimes often favor simplistic sentence structures to facilitate ideological indoctrination, whereas democratic leaders tend to employ more complex linguistic constructions to appeal to diverse audiences (Van Dijk 2006; Fairclough 1989). Therefore, the differences between communist and democratic leaders’ speeches should be reflected in their syntactic complexity, with democratic rhetoric typically employing more complex sentence structures while authoritarian communication favors simplified, repetitive patterns for ideological clarity.

2 MATERIAL

The corpus under analysis consists of 95 annual speeches delivered by all twelve Czechoslovak and Czech presidents. This tradition began in 1935 when the first Czechoslovak President, Tomáš Garrigue Masaryk, addressed the nation on the occasion of the New Year. However, Masaryk delivered only one such speech, as he resigned from office in December 1935 due to age and health concerns. Since then, these speeches have become a consistent annual tradition, providing an important window into political discourse over time.

The texts were primarily obtained from Český rozhlas, the Czech public radio broadcaster, (accessible at: <https://interaktivni.rozhlas.cz/prezidentske-projevy/>). This archive includes both textual transcriptions and audio recordings of the speeches. Additionally, speeches delivered by the most recent presidents, Miloš Zeman and Petr Pavel, were sourced from the official website of the Office of the President of the Czech Republic (<https://www.hrad.cz/>). The overview of the material can be seen in Tab. 1.

group	president	number of texts	years
Democrats	Masaryk	1	1935
Democrats	Beneš	11	1936–1938, 1941–1948
–	Hácha	7	1939–1945
Communists	Gottwald	5	1949–1953
Communists	Zápotocký	4	1954–1957
Communists	Novotný	11	1958–1968
Communists	Svoboda	6	1969–1974
Communists	Husák	15	1975–1989
Democrats	Havel	13	1990–2003
Democrats	Klaus	10	2004–2013
Democrats	Zeman	10	2014–2023
Democrats	Pavel	2	2024–2025

Tab. 1. Overview of material

### 3 METHODOLOGY

In this analysis, we followed a structured approach. The speeches were parsed using UDPipe 2.0 (Straka 2018) with Universal Dependencies (UD) 2.15 models (Zeman et al. 2024), a widely used framework for syntactic annotation. The resulting dependency structures were then converted into the Surface Syntactic Universal Dependencies (SUD) scheme (Gerdes et al. 2018), which provides a more syntactically oriented representation of sentence structure.

To ensure data integrity and enable valid comparisons, we included only sentences that (i) contained a predicate (i.e., a finite verb or auxiliary) as the root of the sentence and (ii) did not contain abbreviations, numerical digits, or special characters. We computed these syntactic indices – average sentence length (ASL), average clause length (ACL), mean dependency distance (MDD) and mean hierarchical distance (MHD).

ASL was computed using two complementary approaches: (i) dividing the total number of words by the number of sentences and (ii) dividing the total number of clauses by the number of sentences. The first measure reflects overall sentence length, while the second captures clause density within sentences.

ACL was determined by dividing the total number of words by the total number of clauses, offering insight into clause-level complexity.

MDD, based on Liu (2008), quantifies syntactic complexity by measuring the average dependency distance (DD) across all words in a text, excluding punctuation and root words. The DD of a word is defined as the absolute difference between its id (position of the word in the sentence) and the id of its syntactic parent. The sum of all DDs in a text was divided by the total number of dependent words (i.e., total words minus the number of sentences), as formalized in following formula:

$$\text{MDD} = \frac{\sum_{i=1}^{n-s} |DD_i|}{n - s}$$

where  $n$  is the number of words,  $s$  is the number of included sentences, and  $DD_i$  is the dependency distance of the  $i$ -th word.

MHD, introduced by Jing and Liu (2015), was computed analogously to MDD but using hierarchical distances (HDs) instead of dependency distances. The HD of a word represents the number of dependency edges between the word and the root of the sentence. MHD provides a deeper structural perspective, capturing the degree of syntactic embedding in a sentence.

To assess the significance of the results, we conducted statistical comparisons across the following groups:

1. Democratic-era (41 speeches) vs. Communist-era (47 speeches) vs. Háchá's 7 speeches.
2. Individual presidents compared against one another.

Before conducting statistical tests, we evaluated the normality of each group's distribution using the Shapiro-Wilk test (Shapiro and Wilk 1965). If normality was violated in at least one group, the Mann-Whitney U test (Mann and Whitney 1947) was applied as a non-parametric alternative. If both groups followed a normal distribution, we employed the independent samples t-test to compare means.

## 4 RESULTS

### 4.1 Democratic-era vs. communist-era speeches

The findings reveal differences in syntactic complexity across the ideological contexts of democratic and communist-era presidential speeches (see Tab. 2). Democratic speeches exhibit a tendency toward higher sentence length, both in terms of words and clauses. In contrast, communist-era speeches feature shorter sentences, suggesting a more constrained syntactic style. However, only sentence length measured in number of clauses shows a statistically significant difference ( $p \leq 0.05$ ).

Interestingly, while communist-era speeches contain shorter sentences overall, they also feature longer clauses compared to democratic-era speeches ( $p \leq 0.05$ ), indicating a shift in the internal structuring of sentences. This suggests that instead of expanding sentences with additional clauses, communist-era speeches tend to rely on more complex clause structures.

Further structural differences are evident in mean dependency distance (MDD) and mean hierarchical distance (MHD). Democratic speeches display greater syntactic depth and larger dependency distances, indicating a more complex structure. However, only MHD demonstrates a statistically significant difference ( $p \leq 0.05$ ), reinforcing the notion that democratic speeches tend to be syntactically more sophisticated.

The case of Emil Hácha, whose speeches do not fit into the democratic or communist categories offers additional insights. Statistically, Hácha's speeches show a significantly higher average sentence length in clauses compared to communist-era presidents ( $p \leq 0.05$ ), while his average clause length in words is significantly lower than that of communist leaders ( $p \leq 0.05$ ). These findings suggest that his speeches share more similarities with democratic discourse. However, in terms of MHD, Hácha's speeches stand out as significantly lower than both democratic and communist presidents ( $p \leq 0.05$ ), positioning his rhetorical style as distinct from both groups.

index	Democrats	sd	Communists	sd	Hácha	sd
ASL (words)	19.243	5.880	18.596	2.462	17.157	2.394
ASL (clauses)	2.429	0.573	1.888	0.344	2.154	0.363
ACL	7.895	1.184	10.010	1.456	8.032	0.688
MDD	2.306	0.188	2.174	0.111	2.199	0.207
MHD	4.339	0.578	4.226	0.440	3.747	0.413

**Tab. 2.** Average values (ASL, ACL, MDD, MHD) of democrats, communists and Hácha and their standard deviations (sd)

## 4.2 Individual presidents compared against one another

As can be seen in Tab. 3 and 4, syntactic complexity varies across individual presidents, reflecting differences in historical context, political ideology, and rhetorical strategy.

Among all presidents, Beneš consistently exhibits the highest syntactic complexity across multiple indices. His speeches feature significantly longer sentences (ASL), more clauses per sentence, and greater hierarchical depth (MHD) than those of most other presidents ( $p < 0.001$  in comparisons with Klaus, Zeman, Pavel, among others). This complex rhetorical style reflects his diplomatic background and the high formal demands of early democratic leadership.

Havel also ranks among the most syntactically complex speakers, particularly in clause density (ASL in clauses) and sentence depth (MHD). His background as a playwright and philosopher likely contributed to his use of layered, introspective sentence structures. Havel's values are statistically higher than those of most communist-era and modern presidents ( $p < 0.01$  in several cases), further distinguishing his rhetorical profile within the post-1989 democratic period.

In contrast, more recent presidents (Klaus, Zeman, and Pavel) exhibit a reduction in syntactic complexity. Klaus and Zeman both use shorter sentences than Beneš and Havel ( $p < 0.001$  in multiple comparisons), indicating a shift toward more accessible political communication. Pavel, the most recent president, has the shortest ASL and lowest MDD in the corpus, with statistically significant differences from nearly all earlier presidents. These findings reflect a broader movement toward direct, media-adapted rhetoric in the contemporary political sphere.

Masaryk delivered only one New Year's address before resigning due to age and poor health, limiting statistical comparison. Although his background as a philosopher suggests a capacity for complex rhetoric, the speech ranks only moderately in complexity, possibly reflecting his condition at the time. Caution is therefore needed in interpreting this as representative of his typical style.

Among communist-era presidents, Husák stands out for using longer clauses and more complex sentence structures. His average clause length (ACL) is the highest of all presidents in the corpus. Unlike earlier communist leaders who used

shorter, slogan-like sentences, Husák’s speeches are more formal and technocratic. It could be explained by his academic background and the context of the normalization era of the 1970s and 1980s. His speeches often include bureaucratic phrasing and abstract discussions of economic planning.

Early communist presidents show varied rhetorical profiles, likely shaped by their backgrounds and ideological roles. Gottwald, with limited formal education and a background in manual labor, uses short, slogan-like sentences reflecting a simplified, mobilizing style. Zápotocký, in contrast, shows higher complexity, possibly due to his experience as a writer, suggesting a somewhat more elaborative approach, though still within ideological constraints.

Novotný and Svoboda also differ in their syntactic patterns. Novotný’s relatively high values for sentence length and structural depth may reflect a more technocratic style, consistent with his engineering background. Svoboda, a military commander, ranks among the simplest speakers, favoring short, directive sentences. This may stem from a preference for clarity shaped by his military experience.

Hácha (1939–1945), whose presidency took place under foreign occupation, consistently ranks in the middle range of syntactic complexity, indicating a rhetorical style shaped by political ambiguity and limited agency.

	ASL (words)	sd	ASL (clauses)	sd	ACL	sd
Masaryk	16.000	0.000	1.833	0.000	8.727	0.000
Beneš	24.092	7.886	2.632	0.808	9.149	1.092
Hácha	17.157	2.394	2.154	0.363	8.032	0.688
Gottwald	17.997	1.188	1.888	0.111	9.530	0.170
Zápotocký	18.880	1.229	1.839	0.219	10.348	1.133
Novotný	19.882	1.763	2.310	0.330	8.693	0.910
Svoboda	15.970	4.539	1.652	0.227	9.586	1.722
Husák	18.828	1.597	1.684	0.133	11.215	1.008
Havel	22.225	3.437	2.712	0.382	8.207	0.675
Klaus	14.316	0.977	1.929	0.158	7.465	0.788
Zeman	16.424	2.270	2.501	0.434	6.622	0.551
Pavel	13.551	0.052	1.921	0.011	7.056	0.069

**Tab. 3.** Average values (ASL, ACL) of individual presidents and their standard deviations (sd)

	<b>MDD</b>	<b>sd</b>	<b>MHD</b>	<b>sd</b>
Masaryk	2.211	0.000	3.689	0.000
Beneš	2.475	0.174	4.598	0.420
Hácha	2.199	0.207	3.747	0.413
Gottwald	2.192	0.102	3.976	0.267
Zápotocký	2.330	0.086	4.284	0.145
Novotný	2.216	0.067	4.356	0.405
Svoboda	2.061	0.109	3.818	0.616
Husák	2.141	0.090	4.36	0.384
Havel	2.367	0.137	4.848	0.428
Klaus	2.273	0.141	3.775	0.251
Zeman	2.136	0.106	4.113	0.490
Pavel	2.041	0.036	3.871	0.090

**Tab. 4.** Average values (MDD, MHD) of individual presidents and their standard deviations (sd)

## 5 CONCLUSION

This study examined the relationship between syntactic complexity and political ideology in Czechoslovak and Czech presidential speeches, analyzing a nearly century-long corpus of New Year's addresses. By investigating multiple measures of syntactic complexity – including average sentence length (ASL), average clause length (ACL), mean dependency distance (MDD), and mean hierarchical distance (MHD) – the findings reveal distinctions between democratic and communist-era presidential discourse.

The results indicate that democratic-era presidents generally exhibit higher syntactic complexity across multiple metrics. Beneš and Havel stand out with the longest and most complex sentence structures, demonstrating a preference for rhetorically sophisticated discourse, which aligns with their intellectual and philosophical backgrounds. Conversely, communist-era leaders, particularly Svoboda and early communist presidents, exhibit lower syntactic complexity, favoring shorter sentences and reduced syntactic depth, a characteristic often associated with authoritarian discourse that prioritizes ideological clarity and mass accessibility. Hácha represents a transitional case, with syntactic complexity values that place him between democratic and communist leaders.

Over time, a trend toward simplification is observed in modern presidential rhetoric. Recent democratic presidents, such as Klaus, Zeman, and Pavel, exhibit significantly lower syntactic complexity compared to earlier democratic leaders. Their speeches contain shorter sentences, reduced clause length, and flatter syntactic structures, indicating a shift toward more direct and accessible political communication.



The results show that political ideology plays a key role in shaping syntactic choices in presidential speeches. The findings reveal that authoritarian regimes tend to favor structurally simpler and more constrained discourse, while democratic leaders historically have employed more complex syntactic constructions. However, the increasing simplification in modern democratic rhetoric suggests that other factors – such as media evolution, changing audience expectations, and the influence of digital communication – may now be driving linguistic change in political speech.

At the same time, the results reveal that these broader trends are significantly modulated by individual factors. Differences in educational background, professional training, and personal rhetorical style account for much of the variation observed within each ideological group. These findings indicate that although ideological context and historical shifts shape the general trajectory of political language, individual characteristics continue to play a crucial role in determining syntactic style.

## ACKNOWLEDGEMENTS

This research is supported by Grant SGS04/FF/2025, University of Ostrava.

## References

- 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.
- David, J., Čech, R., Davidová Glogarová, J., Radková, L., and Šústková, H. (2013). *Slovo a text v historickém kontextu – perspektivy historickosemantické analýzy jazyka*. Brno: Host, 324 p.
- Gerdes, K., Guillaume, B., Kahane, S., and Perrier, G. (2018). SUD or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pp. 66–74, Brussels, Belgium. Association for Computational Linguistics. Accessible at: <https://aclanthology.org/W18-6008/>.
- Jičinský, M., and Marek, J. (2017). New year's day speeches of Czech presidents: phonetic analysis and text analysis. In: K. Saeed – W. Homenda – R. Chaki (eds.): *Computer Information Systems and Industrial Management*. Cham: Springer, pp. 110–121.
- Jing, Y., and Liu, H. (2015). Mean Hierarchical Distance Augmenting Mean Dependency Distance. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, pp. 161–170, Uppsala, Sweden. Accessible at: <https://aclanthology.org/W15-2119/>.
- Kubát, M., Mačutek, J., and Čech, R. (2021). Communists spoke differently: An analysis of Czechoslovak and Czech annual presidential speeches. *Digital Scholarship in the Humanities*, 36(1), pp. 138–152.

Kuznetsova, J. (2016). Modern Russian history through the New Year addresses. In: S. Kübler–M. Dickinson (eds.): *Proceedings of Computer Linguistics Fest 2016*. Bloomington, IN: Indiana University, pp. 34–38.

Lim, E. T. (2004). Five trends in presidential rhetoric: an analysis of rhetoric from George Washington to Bill Clinton. *Presidential Studies Quarterly*, 32(2), pp. 328–348.

Liu, F. (2012). Genre analysis of American presidential inaugural speech. *Theory and Practice in Language Studies*, 2(11), pp. 2407–2411.

Liu, H. (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, 9(2), pp. 159–191.

Mann, H. B., and Whitney, D. R. (1947). On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, 18, pp. 50–60.

Straka, M. (2018). UDPipe 2.0 Prototype at CoNLL 2018 UD Shared Task. In *Proceedings of CoNLL 2018: The SIGNLL Conference on Computational Natural Language Learning*, pp. 197–207. Association for Computational Linguistics, Stroudsburg, PA, USA. Accessible at: <https://aclanthology.org/K18-2020/>.

Savoy, J. (2010). Lexical analysis of US political speeches. *Journal of Quantitative Linguistics*, 17(2), pp. 123–141.

Savoy, J. (2016). Text representation strategies: an example with the State of the Union addresses. *Journal of the Association for Information Science and Technology*, 67(8), pp. 1858–1870.

Shapiro, S. S., and Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52, pp. 591–611.

Tuzzi, A., Popescu, I.-I., and Altmann, G. (2010). *Quantitative Analysis of Italian Texts*. Lüdenscheid: RAM-Verlag, 161 p.

Van Dijk, T. A. (2006). Ideology and discourse analysis. *Journal of Political Ideologies*, 11(2), pp. 115–140.

Zeman, D. et al. (2024). Universal Dependencies 2.15, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Prague: Charles University. Accessible at: <http://hdl.handle.net/11234/1-5787>.