

ANALYSIS OF CZECH PRESIDENTIAL SPEECHES WITH NLP TOOLS

Nikola Kaspříková

Abstract

Recent presidential election in the Czech Republic has raised a lot of interest in the public and in the media. The tools for natural language processing available in statistical software offer a possibility to perform a formal analysis of the speech. The paper presents some results of a sentiment analysis and topic modeling of seven inaugural addresses of the Czech presidents. It follows from the sentiment polarity analysis of the speeches that all the speeches have positive polarity.

Topic analysis has suggested that topic focused on president was the most likely one in the two speeches given by Miloš Zeman. The second topic focused on the Czech Republic is the most covered one in the first four speeches (two speeches by Václav Havel and two speeches by Václav Klaus). The third topic which may be considered as philosophical and humanitarian one is represented most in the first speech by Václav Havel and in the speech by Petr Pavel and is nearly not represented in speeches by Václav Klaus.

Key words: NLP, Czech presidents, sentiment analysis, topics analysis

JEL Code: D72, C80

Introduction

Recent president vote in the Czech Republic has raised a lot of interest in the public and in the media. The presidential inaugural speech has been observed by many people and received a lot of comments. The tools for natural language processing (NLP) available in statistical software offer a possibility to perform a formal analysis of the speech. The presidential inaugural addresses attract interest of many researchers, see for example recent papers (Fowobaje et al., 2022) and (Han and Lim, 2021).

The paper (Kaspříková, 2023) addresses the formal analysis using natural language processing techniques of the two presidential inaugural speeches given by the recent president, the first Czech president elected directly by the public. The tokenization, lemmatization, identification

of the parts of speech and frequency analysis of words and co-occurrences by parts of speech types is among the techniques used for the analysis. The results show that the co-occurrences found by NLP could be used for automated keywords identification.

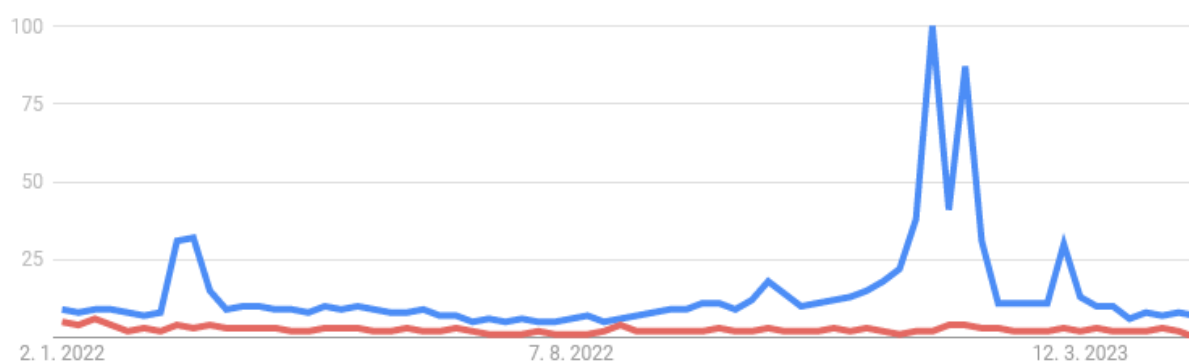
This paper presents some results of a basic analysis of seven inaugural speeches of the Czech presidents with the aim to analyse sentiment polarity of the speeches and to provide a basic topic modeling.

1 Presidents in the Czech Republic

The president is often mentioned in media in the Czech Republic, especially in times around election of the president. The question is how to objectively find out how citizens perceive the importance of the president. There are opinion surveys which measure the popularity of the president and government members. One of the ways of getting an insight into the popularity of president is to use the Google Trends data. The Figure 1 shows the popularity of two search terms, “prezident” (president) and “vláda” (government) in Google Trends in the Czech Republic since January 2022. The maximum number of weekly searches is assigned the value 100 and the other values are shown relative to the maximum, for example value 50 reflects 50 percent of the maximum number of searches. The frequencies of the two search terms may reflect the level of interest in president and in the government.

It follows from Figure 1 that the search term president is much more popular than government all along the selected period. There are clear peaks in popularity of the president search term in time of election of the president and around the time of inauguration of the president.

Fig. 1: Popularity of Google searches – „prezident“ (“president”, in blue) versus „vláda“ (“government”, in red)



Source: Google Trends

There have been four people serving as Czech president since the founding of the Czech Republic. The Table 1 shows the complete list of the presidents of the Czech Republic. The first three people who served as the presidents in the Czech Republic were elected for two consecutive periods. Václav Havel and Václav Klaus were elected by the Parliament, Miloš Zeman was the first one to be elected in direct vote by the public.

Tab. 1: Czech Presidents

President	Starting in year
Václav Havel	1993
Václav Havel	1998
Václav Klaus	2003
Václav Klaus	2008
Miloš Zeman	2013
Miloš Zeman	2018
Petr Pavel	2023

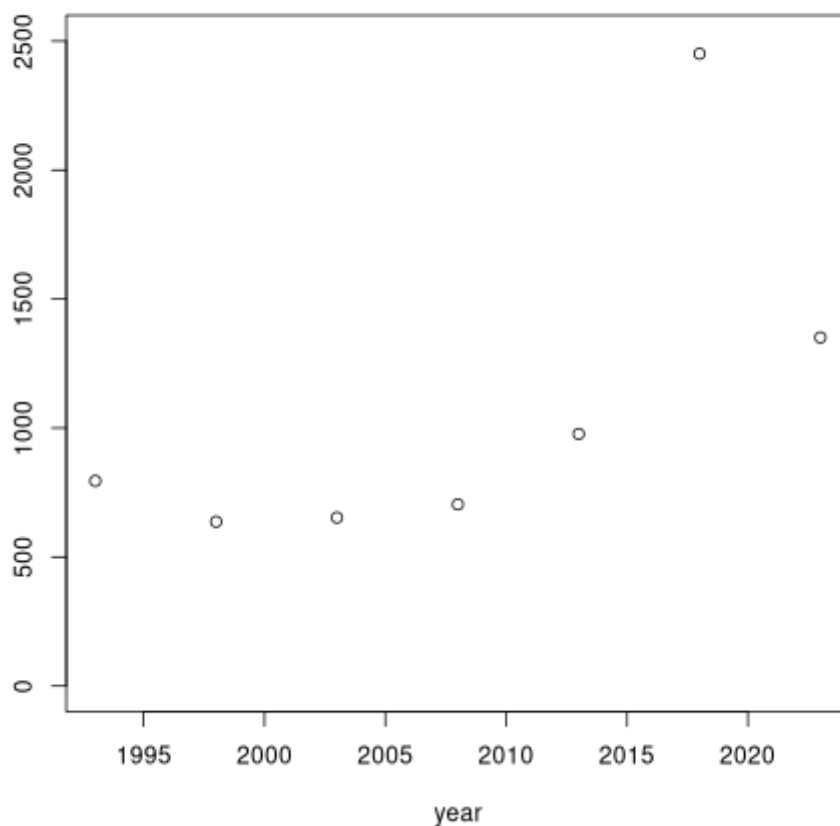
Source: https://cs.wikipedia.org/wiki/Seznam_prezident%C5%AF_%C4%8Cesk%C3%A9_republiky

The presidents in the Czech Republic have been always successful in the vote for their second term. The current president at the beginning of his first term now. There are seven inaugural speeches of the Czech presidents available from 1993 to 2023, which gives rather short history compared with other countries.

2 Data and Methods

The seven inaugural speeches may be obtained from (Wikisource.org, 2023). The speeches do not have equal length. Figure 2 shows the length of the inaugural speeches measured by the number of nonpunctuation terms, see also Table 2. The first five speeches are roughly of the same length. The second speech by Miloš Zeman is rather long and the 2023 speech by Petr Pavel is the second longest.

Fig. 2: Length of speech as the number of nonpunctuation terms



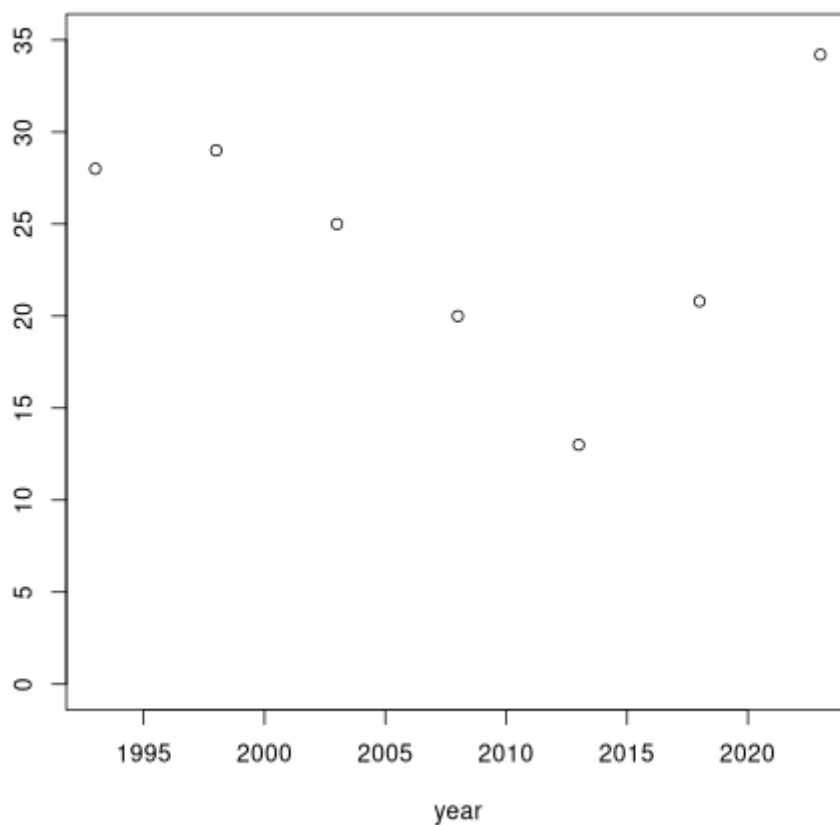
Source: own calculation using udpipe package

The udpipe package (Wijffels, 2023) for the R statistical software (R Core Team, 2023) is used for the dictionary-based sentiment analysis. The Czech SubLex dictionary (Veselovská and Bojar, 2013) is used. The topic modeling is performed using the tools in the topicmodels R package (Grün and Hornik, 2011). For the description of NLP methods see also (Jurafsky and Martin, 2023).

3 Sentiment Analysis

Figure 3 shows the sentiment polarity scores of the inaugural speeches of the Czech presidents. Clearly all the speeches have positive polarity. The highest sentiment polarity score is in the recent speech given by Petr Pavel in 2023, followed by the two Václav Havel addresses. The smallest polarity score, but still positive, was observed in the first speech by Miloš Zeman. Table 2 presents the detailed values of the sentiment polarity scores and an average sentiment polarity score per nonpunctuation term.

Fig. 3: Polarity scores of inaugural speeches



Source: own calculation using udpipe package and Czech SubLex dictionary

5 Topic Analysis

The topic modeling is performed based on adjectives and nouns. The Latent Dirichlet allocation (LDA) method is used for the analysis, see (Blei et al., 2003), (DiMaggio et al., 2013) and (Griffiths and Steyvers, 2004). The model with three topics as been used based mostly on interpretability. Table 3 shows 100%*probability of topics within each speech.

Tab. 2: Sentiment polarity scores

Inaugural speech	Sentiment polarity score	Number of terms	Average polarity score per term
Václav Havel 1993	28	795	0.035
Václav Havel 1998	29	637	0.046
Václav Klaus 2003	25	653	0.038
Václav Klaus 2008	20	704	0.028
Miloš Zeman 2013	13	977	0.013
Miloš Zeman 2018	20.8	2451	0.008
Petr Pavel 2023	30.4	1351	0.025

Source: own calculation using udpipe package and Czech SubLex dictionary

Tab. 3: Topics by speeches

Inaugural speech	Topic 1	Topic 2	Topic 3
Václav Havel 1993	0	60	40
Václav Havel 1998	0	83	17
Václav Klaus 2003	44	56	0
Václav Klaus 2008	40	56	3
Miloš Zeman 2013	73	11	16
Miloš Zeman 2018	81	5	14
Petr Pavel 2023	31	26	43

Source: own calculation using topicmodels package

The first topic identified is the most likely one in the two speeches given by Miloš Zeman. The second topic is the most covered one in the first four speeches (two speeches by

Václav Havel and two speeches by Václav Klaus). The third topic is represented most in the first speech by Václav Havel and in the speech by Petr Pavel and is nearly not represented in Václav Klaus speeches.

Regarding the interpretation of the three topics identified, Table 4 may provide some suggestions. Table 4 shows in descending order the six most likely terms for each topic with the probability of terms by topic.

Tab. 4: Terms by topic

Topic 1		Topic 2		Topic 3	
prezident	0.045	země	0.078	člověk	0.066
strana	0.043	dobrý	0.066	hodnota	0.042
český	0.040	republika	0.049	nový	0.042
občan	0.038	český	0.044	krok	0.033
řád	0.038	politický	0.034	problém	0.033
vážený	0.035	velký	0.034	způsob	0.033

Source: own calculation using topicmodels package

The most likely term for the first topic is president followed by (political) party, Czech, citizen, like and dear. This topic seems focused on the role of the president. The second topic is characterised by terms country, good, republic, Czech, political and great. This topic may be focused on the Czech Republic. The third topic seems more philosophical or humanitarian, with human being the top term for the topic, followed by value, new, step, problem, way and truth (probability 0.029, not shown in the table).

Conclusion

Seven inaugural addresses of the presidents of the Czech Republic have been analysed with basic NLP techniques. The sentiment polarity analysis has shown that all addresses are positive, the recent 2023 speech by Petr Pavel has the highest polarity score.

Topic analysis has suggested that topic focused on president was the most likely one in the two speeches given by Miloš Zeman. The second topic focused on the Czech Republic is

the most covered one in the first four speeches (two speeches by Václav Havel and two speeches by Václav Klaus). The third topic which may be considered as philosophical and humanitarian one is represented most in the first speech by Václav Havel and in the speech by Petr Pavel and is nearly not represented in speeches by Václav Klaus.

It may be interesting as the next step of research to compare the characteristics of the Czech presidential inaugural speeches with the addresses of presidents in other countries.

Acknowledgment

The paper has been prepared using the institutional support for long term development at Prague University of Economics and Business.

References

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *J. Mach. Learn. Res.*, 3, 993–1022.
- DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the Sociological Perspective on Culture: Application to newspaper coverage of U.S. Government Arts Funding. *Poetics*, 41(6), 570–606. <https://doi.org/10.1016/j.poetic.2013.08.004>
- Fowobaje, K. R., Mashood, L. O., Ekholuenetale, M., & Ibidoja, O. J. (2022). Qualitative content analysis of Nigerian heads-of-state and presidents' inaugural addresses: Text mining, topic modeling and sentiment analysis. *SN Social Sciences*, 2(12). <https://doi.org/10.1007/s43545-022-00570-x>
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl_1), 5228–5235. <https://doi.org/10.1073/pnas.0307752101>
- Grün, B., & Hornik, K. (2011). topicmodels: An r package for fitting topic models. *Journal of Statistical Software*, 40(13). <https://doi.org/10.18637/jss.v040.i13>
- Han, H., & Lim, S. (2021). Inaugural speech classification with named entities and key phrases. *2021 IEEE International Conference on Big Data and Smart Computing (BigComp)*. <https://doi.org/10.1109/bigcomp51126.2021.00029>
- Jurafsky, D., & Martin, J. H. (2022). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Pearson.
- Kaspříková, N. (2023). *Czech presidential inaugural addresses in 2013-18: a natural language processing analysis*. Submitted to Proceedings of SMIS 2023 conference.

R Core Team. (2023). R: A language and environment for statistical computing. [Computer software]. Retrieved from www.r-project.org

Veselovská, K., & Bojar, O. (2013). *Czech sublex 1.0*. CZ. <https://lindat.mff.cuni.cz/repository/xmlui/handle/11858/00-097C-0000-0022-FF60-B>

Wijffels, J. (2023). *udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit*. R package version 0.8.11. Available at: <https://CRAN.R-project.org/package=udpipe>.

Wikisource.org. (2023). *Kategorie:Inaugurační Projevy prezidentů české republiky – wikizdroje*. – Wikizdroje.

https://cs.wikisource.org/wiki/Kategorie:Inaugura%C4%8Dn%C3%AD_projevy_prezident%C5%AF_%C4%8Cesk%C3%A9_republiky

Contact

Nikola Kaspříková

Katedra matematiky VŠE v Praze

Nám. W. Churchilla 4, Praha

nikola.kasprikova@vse.cz