

Initial Analysis of Presidential Candidate Speeches

Research Article

Eliška Syrovátková, Jaroslav Krbec*
Palacký University Olomouc, Department of General Linguistics

Received: May 2019; Accepted: November 2019

Abstract: In January 2018, the President of the Czech Republic was elected. Before that, each of the candidates communicated their intention to run for the office in a different kind of speech. By using selected characteristics, we evaluate and compare these candidate speeches. Subsequently, we reflect on the possibilities of correlating the results of the election with data collected during the analysis.

Keywords: presidential election, thematic analysis, thematic concentration, speech analysis, topic analysis, vocabulary richness

INTRODUCTION

Democratic elections are always a strong driver in society, not only because of their results, but also with regard to the campaign process itself. The same is true about the presidential elections in a country. The presidential election in the Czech Republic took place in January 2018. Ten candidates had previously declared their interest in the presidential office in their introductory candidate speeches. In this article, we focus our attention on these. Our goal is to compare candidate speeches and their properties based on selected basic, thematic and quantitative characteristics.

First, we briefly describe the analytical steps and the very selection of the analysed material. Since the form of these candidate manifestations varies considerably, we devote the next part of the paper to their basic descriptions. Subsequently, we move on to identifying general topics and topics that were specific for each candidate. In the next comparison, we focus our attention on the application of quantitative linguistic indices. We capture the vocabulary richness, the thematic concentration (TC) and keywords. Based on the obtained data, we consider the possibility of certain trends and potential correlations with election results.¹ Since the article is

conceived primarily as an output of analytical activity, the theoretical basis used is continuously mentioned as the background of particular analytical steps.

ANALYSED DATA

Concerning the criteria to be considered when focussing on certain topics, Kořenský et al. (2013) emphasise, above all, aspects of social relevance and associated novelty or social severity. This is also reflected in common research interests in the areas of political and media discourse.

In accordance with these aspects, our area of interest and the material analysed consist of candidate speeches, where speakers officially declare their intention to run for the post of the President of the Czech Republic. All of the candidates did so publicly at press conferences. The only exception in this regard was Miloš Zeman, who first announced his intention to run for the post to his supporters on the occasion of a private speech (represented in the tables as SP); the next day, he held an official press conference for the public (represented in the tables as OP). Since the recordings of the performances were available, both the speeches of Zeman were analysed,

* Corresponding author: Jaroslav Krbec, e-mail: jaroslav.krbec01@upol.cz

1 For possible evaluation by readers, the results of the elections are attached in the Appendix section (Tables A1 and A2).

Initial Analysis of Presidential Candidate Speeches

taking into account the possibility of their comparison. Only in the case of Petr Hannig, the record from the press conference was not traced. Therefore, the transcript of the speech from the press conference, which the candidate posted on his website, was used for the purpose of analysis. Other candidates' speeches were traced and processed into a text form. This was done by the implementation of specific analytical procedures.

ANALYTICAL STEPS

In cases of critical discourse analyses, it is primarily the linguistic nature of the investigation that is usually considered to be a fundamental building or the dominant component of approaches (Wodak, Meyer 2009, 28). The research activity of critical discourse analysis is then viewed in contrast to other types of analysis, be it the various forms and approaches of the grounded theory or conversational-and-textual analyses. Although the limits in the context of microstructure analysis of excerpted and decontextualised speeches are evident, the effects of discourse cannot be understood without examining what is happening when people speak or write (Fairclough 2003, 15). Thus, in the case of the sample, we remain primarily in the field of methods dependent on the application of linguistic concepts. However, given the characteristics of the material under investigation, we also include a comparison of the basic characteristics of the speeches and their thematic analysis, which appears to be crucial in relation to political discourse. We choose the applied analytical steps after considering the specific characteristics of the sample with regard to the practical implementation of the initial analysis. This should lead to the revelation of trends for potentially more detailed examination in the future.

Our goal is to capture the basic characteristics of introductory speeches by the candidate and their comparison. Subsequently, we would like to find out whether there is a correlation between certain characteristics of the speeches and the election results.

The analysed material was processed on several levels, in which the focus changed from dataset characteristics to dependent method adaptation, and data operationalisation was continuously enabled. Most of the performances were transcribed in collaboration with a study group. All of the participants thus had the opportunity to study the speeches first. Then, the process continued chiefly with the identification and classification of topics. The original segmentation by content-pragmatic units led to the identification of repeating topics that differed in their degrees of meaning complexity.

Based on the group dialogue, the individual thematic units were further identified and subdivided into several groups according to the semantic proximity (Table 2). In addition to identifying the general topics, some topics that were specific to a smaller range of candidates or individuals were identified. Content analysis was then used to identify specific topics, taking into account brief formal-content evaluations and descriptions of the speeches developed by the study group. The involvement of a larger number of evaluators was chosen considering the possibilities for legitimising of procedures and avoidance of specific interpretative errors.

The description of the basic characteristics and the identification of topics are precursors to the comparison of the lexical aspects of individual speeches. This process also offers the initial processing of a dataset, which potentially allows focussing on certain trends, while partially reducing the information complexity in order to facilitate movement at different levels of description. The lexical characteristics and comparison of the individual speeches serve to evaluate the vocabulary used by candidates. The classification principle is the semantic opposition of expressive and neutral expression.

For further sample processing, we also use quantitative methods. These are traditionally used as additional tools for text characterisation when working with linguistic data, in addition to serving as tools for the testing of hypotheses about properties of the language

| Candidate's name | Date of Speech | Length of Speech (min) | Transcribed Speech Length | Speech Pace (char.:secs) |
|--------------------|----------------|------------------------|---------------------------|--------------------------|
| Jiří Drahoš | 28.03.2017 | 8:16 (496 s) | 4044 char. | 8,153 |
| Pavel Fischer | 05.10.2017 | 7:43 (463 s) | 6029 char. | 13,021 |
| Marek Hilšer | 30.07.2016 | 9:05 (545 s) | 4970 char. | 9,119 |
| Michal Horáček | 03.11.2016 | 10:45 (645 s) | 7082 char. | 10,979 |
| Vratislav Kulhánek | 29.06.2017 | 3:34 (214 s) | 1817 char. | 8,49 |
| Mirek Topolánek | 07.11.2017 | 22:58 (1338 s) | 16143 char. | 11,714 |
| Miloš Zeman OP | 09.03.2017 | 16:52 (1012 s) | 7741 char. | 7,649 |
| Miloš Zeman SP | 10.03.2017 | 6:23 (383 s) | 2374 char. | 6,198 |
| Jiří Hynek | 21.08.2017 | 8:42 (522 s) | 5064 char. | 9,701 |
| Petr Hannig | 19.07.2017 | | 4603 char. | |

Table 1: Basic characteristics of candidate speeches

| Topics | Miloš Zeman (OP) | Miloš Zeman (SP) | Michal Horáček | Jiří Drahoš | Pavel Fischer | Vratislav Kulhánek | Marek Hilšer | Jiří Hynek | Mirek Topolánek | Petr Hannig |
|-----------------------------|------------------|------------------|----------------|-------------|---------------|--------------------|--------------|------------|-----------------|-------------|
| Self-presentation | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Candidate's motivation | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Candidate's Values | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Campaign Character | ✓ | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | |
| President's Role | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Czech Republic in the World | | | ✓ | ✓ | | ✓ | ✓ | | | ✓ |
| Evaluation of Competitors | | ✓ | ✓ | | | | | | ✓ | ✓ |
| Own Health | ✓ | | ✓ | | | | | ✓ | | |

Table 2: Common topics in speeches

system (Čech et al. 2014, 7). Considering the focus of this paper, the first of these conceptions is particularly important for us. Thus, we are able to expand the potential of comparing and monitoring trends arising from the characteristics of the language used in speeches.

Although the impact of the context is fundamental, there is no clear procedure to establish the principles of its integration with the final interpretation (Wodak, Meyer 2013, 30). The conclusion of the article is thus conceived as a summary and, at the same time, an abductive interpretation of the identified phenomena and trends. It offers an overview of possible correlations of election results with the identified characteristics of the texts examined.

BASIC CHARACTERISTICS OF SPEECHES

If we evaluate the basic characteristics of the speeches, at first glance, we see considerable differences in the overall concept of the performances. Speeches can be divided according to whether the candidates alone held the role of spokesperson or have been partly replaced by other speakers. In order to partially standardise the analysed material for comparison purposes, we select only those sections where candidates act as speakers themselves. This is how we compare individual performances without introducing moderators of press conferences, speeches of other co-workers of the candidates or questions of the press and the answers.²

As is evident from Table 1, the shortest candidate speech was given by Vratislav Kulhánek, which is related to the overall concept. The candidate is originally mentioned by another speaker, and he himself is also quite brief. The second shortest speech was given by Miloš Zeman in his speech for supporters. Here, the key factor

is the lowest speech pace in terms of the ratio of speech length and the extent of its written variant. However, some forced pauses are caused by the reactions of the audience, especially by the long applause when announcing the candidacy. Although Miloš Zeman's speeches are different in duration, their speech rate is still the lowest compared to other speeches. As we can see, there is an apparent consistency of style and expression. Mirek Topolánek presented the longest speech with an above-average speech pace. Pavel Fischer spoke the fastest. The other candidates do not deviate significantly in this comparison. In relation to the pace of speech, one can observe the trend, whereby the speech that is primarily written (in the cases of Topolánek, Fischer and Horáček) shows the highest values of this index. Thus, we could find here a connection between prepared speeches and the speed of their delivery.

Since these are only the basic characteristics, the possibility of correlation with the election results is practically unrealistic. Let us point out at least those deviations that, to some extent, reflect the order of candidates in the elections. After all, we do so with an intention of gradual development, as the following parts of the paper show.

Considering the length of speeches, we can observe an association of the poor election result with extreme values. This is especially true for the results of M. Topolánek and V. Kulhánek. The values of the transcribed speeches, in this case, show the same extreme, as expected. As mentioned earlier, we can also label the speech of M. Zeman for his supporters as distinctive. In this case, the speech is the second shortest with the lowest speech pace. It could also be an interesting aspect with reference to affecting the audience and, therefore, in terms of influencing the election results. However, in contrast, a speech of P. Fischer was the fastest of all,

2 Due to the diverse nature of the original communications, we do not apply the traditionally used syllable segmentation method for speech pace measurement. For the purposes of this paper, we consider the applied approach sufficient for overview concerning both written and spoken speeches.

Initial Analysis of Presidential Candidate Speeches

yet in the election, Fischer finished in third place. As the values for other speeches show no significant deviations, there is no possibility to look for other potential correlations without additional data. We will focus on this data further in this paper.

IDENTIFICATION OF TOPICS

Although the announcement of a presidential candidacy is a very specific communication act, we can recognise quite different ways of speech conceptions on the level of topic selection. The choice of topics is one of the aspects that fundamentally affects the concept communicated.

For purposes of analysis, we first identify individual thematic areas, then we compare the frequencies of occurrence of specific topics and compare their use within speeches. In the tables, topics are listed first from the more general ones. The following topics are those that cover significant length within particular speeches. Especially for topics with a higher degree of generality, content overlay can be apparent. These are mainly the topics covering self-presentation, the candidate's motivation and the candidate's values. However, as part of the identification, we found that these topics are usually addressed somewhat separately.

As can be seen in Table 2, at least to some extent, all of the candidates used their speech to present themselves to the public. Again, the manner of this topic presentation is highly variable. Theoretically, any aspect of the speech could be considered as self-presentation, but we are specifically concerned with the question on how candidates talk about themselves. To some extent, the basic characteristics of the speeches that were mentioned earlier are also reflected here. If we look at the opposite extremes, we can see that Mirek Topolánek used the longest time period for his own presentation.³ Additionally, he divided information about himself into certain time periods (*I have divided these 28 years into four periods of 7 years...*).⁴ On the other hand, a very short time period is used by Vratislav Kulhánek for self-presentation, who is presented as a candidate by another speaker.

In the vast majority of speeches, topics where candidates state their motivations and values are developed.

In particular, they mention truth, reason, respect, democracy or mutual respect. The only exception here are the speeches of Miloš Zeman. He does not explicitly deal with motivation to his candidacy in the speech for his supporters. Moreover, he does not specify his own values in the official speech the next day.

Only two speeches lack at least a basic outline of the characteristics of an election campaign. These are a speech by Miloš Zeman for supporters and a brief speech by Vratislav Kulhánek. The topic of campaign characteristics is most often elaborated with regard to the way candidates intend to acquire their nomination.

There can also be seen an outline of another particular campaign characteristic: this involves the potential evaluation of other candidates. For example, Jiří Drahoš refuses to comment on his competitors (*I will not define myself as an opposition to someone else...*),⁵ and Miloš Zeman, in his official speech, even promises not to verbally attack or respond to attacks of competitors (*I promise not to contest any of my counter-candidates and if he contests me, I will not respond to these attacks.*)⁶

The very role of the President is also decisively discussed, and almost all candidates commented on this topic. Only in speeches of Miloš Zeman did the topic remain omitted. But in the official speech, he justifies why he does not talk about the subject (*I would feel like a fool if I said what my idea of the presidency looks like. That's what I said four years ago, and I've been trying to fulfil that idea ever since. Voters will therefore compare the idea of individual candidates with the actual performance of the presidential office.*)⁷

Five of the candidates mentioned the position of the Czech Republic within the European or the world political scene. None of them, however, was too specific in their attitudes. Thus, the statements were mainly about state sovereignty (Drahoš: *One should not bow to any side of the world, neither to the West nor to the East, North or South.*),⁸ the nature of the existing alliances (Hilšer: *I want to make sure that the Czech Republic is a dignified and above all self-confident partner in the European Union, that it is a dignified and self-confident partner in other organizations that associate democratic countries...*)⁹ or the reputation of the Czech Republic in the world, as in the cases of Kulhánek (*We are one of the larger states*

3 From now on, we use the English translation of citations in brackets. The original citations in Czech language are further referred to as footnotes using candidate initials for identification.

4 M. T.: "...já jsem těch 28 posledních let rozdělil do čtyř sedmiletých období..."

5 J. D.: "Nebudu se vůči nikomu vymezovat..."

6 M. Z.: "Zavazuji se, že nebudu útočit na žádného ze svých protikandidátů a pokud on bude útočit na mě, nebudu na tyto útoky odpovídat."

7 M. Z.: "Já bych si připadal jako blázen, kdybych říkal, jakou já mám představu o výkonu prezidentské funkce. To jsem říkal před čtyřmi lety a snažil jsem se od té doby tuto představu naplnit. Voliči tedy budou porovnávat představu jednotlivých kandidátů s reálným výkonem prezidentské funkce."

8 J. D.: "Člověk se nemá hrbít před žádnou světovou stranou, ani před východem západem severem nebo jihem."

9 M. H.: "Já se chci zasadit o to aby Česká republika byla důstojným a především sebevědomým partnerem v Evropské unii, aby byla důstojným a sebevědomým partnerem i v dalších organizacích které sdružují demokratické země..."

within Europe. And a state that, I think, had perhaps, perhaps still has, but still has huge reserves in gaining an even better reputation in Europe.)¹⁰ and Hannig (*We are always told that we are small and that we must always rely on someone.*)¹¹

Only four of the speakers commented on the other candidates. Miloš Zeman did so indirectly (*...defeated by those who cannot say their own opinion, they cover themselves in political correctness, which means nothing but phrases, and have the charisma of a pickle bottle*),¹² as did Mirek Topolánek (*I am not a rabbit in a hat that someone has pulled out and he is bouncing on some foreign notes.*)¹³

In connection with the personality and evaluation of Miloš Zeman, Michal Horáček expressed his opinion (*I believe that the function of the President of the Czech Republic can be performed in a more decent, competent and helpful way than before*),¹⁴ Mirek Topolánek was very explicit and expressed his opinions about Miloš Zeman, his own past in politics and current activities (*...and I also want him to rest*).¹⁵ Petr Hannig said, he did not want to be a competitor of the current President and added a brief evaluation (*Many people agree with President Zeman's policy, but they do not want to vote for him because of the form of his behaviour, as they say.*)¹⁶

Usually, presidential candidates do not discuss questions of their own health, except for Jiří Hynek (*I have a physical condition for it, my health is fine, I have enough elan and energy.*)¹⁷ and Miloš Zeman. He focussed his official speech especially on his health issues. Michal Horáček only drew attention to his own publicly available medical reports.

The specificity of speeches is further reflected by the occasional choice of unique topics. They are often associated with the professional or personal interests of candidates, or they are otherwise related to their past. In this way, Vratislav Kulhánek spent almost half of his candidate speech talking about the importance of sport. Jiří Hynek placed special emphasis on the topic of national defence, while Pavel Fischer emphasised the importance of the Senate and its representatives. With regard to the content and concept of the speech, the relationship with Marek Dalík can be described as an important specific topic in the speech of Mirek Topolánek. Important issues in the speech of Petr Hannig were the threats associated with the migration wave and the possibility of leaving the European Union.

Considering the basic characteristics of the presidential speeches and the identified topics, the speeches of Miloš Zeman are very different. He conceived his speech for supporters primarily as a thanks, while, in the end, he referred indirectly to one of his competitors and his charisma (see earlier example). On the contrary, the official speech is much longer. The candidate discusses more topics. He does so with a special regard to his health condition. Furthermore, the character of the campaign and the motivation of the candidate are outlined.

Now, we briefly evaluate the possibilities of correlation among the basic characteristics themselves and the thematic areas of the speeches with the success of the candidates in the elections. The candidates who accentuated specific and serious topics (Kulhánek, Hannig, Hynek and Topolánek) seem to have the lowest number of votes. Although Pavel Fischer devoted considerable space to the importance of the Senate, he remained in an area that is relatively uncontroversial with regard to the situational context and is ranked third in overall vote percentage.

Neither the topic of self-health nor the topics related to the personalities of other candidates seem to show a significant correlation with the election results. However, the candidates who declared that they would not define themselves against other candidates, namely Zeman and Drahoš, received the most votes. Only Miloš Zeman, who won the election, did not introduce the idea of the presidential role. He also did not speak about his values in his official speech.

The results of the election do not reflect the time and content range used for self-presentation. We may see a certain connection in the case of Vratislav Kulhánek, who dealt with this topic in a very limited manner. However, in the case of Mirek Topolánek, with the longest self-presentation, the election results are not corresponding.

VOCABULARY RICHNESS

Vocabulary richness measurement is one of the oldest and most traditional fields in quantitative linguistics. The concept of vocabulary richness measure is based on the fact that each person uses specific individual vocabulary' (Kubát, Milička 2013, 1). The complexity of a language and the ability to formulate ideas help to determine vocabulary richness. 'Authors clearly differ in the sizes and structures of their vocabularies—some have large

10 V. K.: "Jsme jeden z větších států v rámci Evropy. A stát, který, myslím, že měl snad, snad pořád ještě má, ale má ještě obrovské rezervy v získání ještě lepší pověsti v Evropě."

11 P. H.: "Je nám neustále vtoukáno do hlavy, že jsme malí, a že se musíme vždy o někoho opírat."

12 M. Z.: "...poražení těmi, kdo nedokáží říci svůj vlastní názor, zahalují se do politické korektnosti, která neznamena nic jiného než fráze, a mají charisma lahve od okurek."

13 M. T.: "Nejsem králík v klobouku kterého někdo vytáhl a ten teď hopsá podle nějakých cizích not."

14 M. H.: "Věřím, že funkci prezidenta České republiky lze vykonávat slušněji, kompetentněji a vstřícněji než dosud."

15 M. T.: "...a také chci, aby si odpočinul."

16 P. H.: "Mnozí lidé souhlasí s politikou prezidenta Zemana, nechťějí ho však volit, kvůli formě jeho chování, jak říkají."

17 J. H.: "Mám na to fyzickou kondici, funguje mi zdraví, mám na to dostatek elánu a energie."

Initial Analysis of Presidential Candidate Speeches

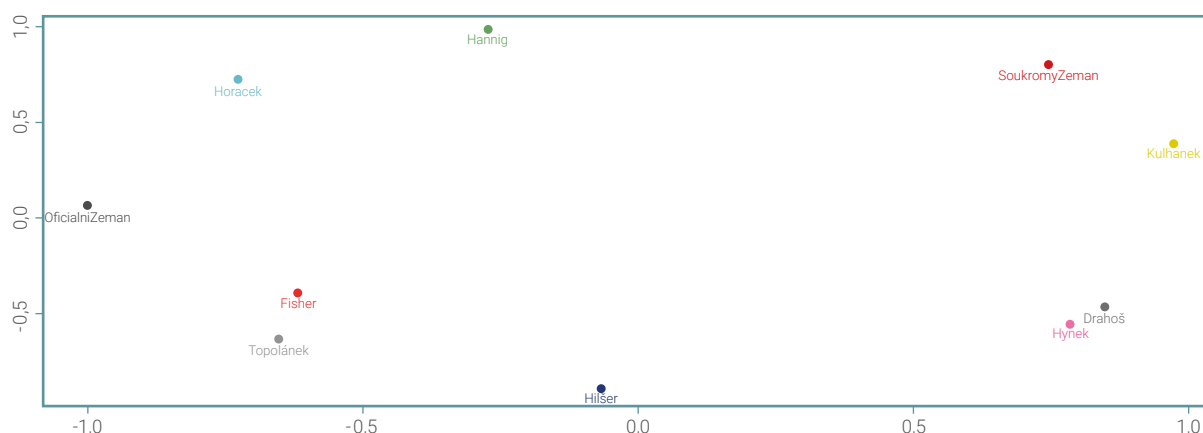


Figure 1: Multidimensional scaling (MDS) map of the results of indices TTR, entropy, RR and Gini coefficient.

vocabularies and use many relatively infrequent words and others have smaller vocabularies and use many more frequent words. This has led to the reasonable assumption, often unstated, that vocabulary richness or concentration provides a kind of authorial word print that can distinguish authors from each other' (Hoover 2003, 151). But it has to be noted that not everybody agrees on the existence of vocabulary richness. Some authors consider this to be a so-called information flow. This presumes that every speaker knows his audience, knows the opportunity and precisely knows the length of the text that he/she creates. The author of a text stratifies lexical units strategically throughout the text to maximise the impact and the understanding of its message. But in our paper, we continue in the tradition of Czech quantitative linguists, who often analyse vocabulary richness and consider it to be existing.

According to Čech, it must be remembered that vocabulary richness was conceptually defined by scientists themselves, which implies that 'it is possible to quantify the vocabulary richness by many different indicators with no "natural" criterion given to decide which method of measurement is better than another' (Čech et al. 2014, 30). For the analysis of the vocabulary richness of presidential candidates, a combination of indices (repeat rate [RR], entropy, Gini coefficient and type-token ratio [TTR]) was chosen. In our opinion, this combination is best suited to the methods that analysed the candidate speeches in the first part of our article, which we want to complete.

The index RR expresses the rate of concentration of the text 'in relation to the lexicon which was used: the higher the RR value, the smaller "distribution of vocabulary", therefore the lesser the vocabulary richness' (Čech 2016, 34). Entropy is generally the name for lack of order or predictability. Therefore, in the context of vocabulary richness, 'the high value of entropy [...] is a sign of extensive vocabulary richness' (Čech 2016, 34). The Gini coefficient is used mostly in economics (also called the Gini index or Gini ratio), its value represents the dispersion to which a Lorenz curve is often used as a graphical representation of diversification (used primarily in economics) and it works with the cumulative distribution function. The value of the Gini coefficient always lies in the span [0,1],

where zero denotes the maximum diversification (Čech 2016, 41). 'One of the oldest and easiest ways of vocabulary richness measure is the type-token ratio (TTR). The TTR index is based on the simple ratio between the number of types and tokens in text. The resulting value shows how much vocabulary varies (the more vocabulary variation in a text, the higher is the TTR)' (Kubát, Milička 2013, 1). All the indices that could be used for expressing vocabulary richness depend on text length. Various ways exist on how to solve this problem; in our paper, we have chosen the method of lemmatisation of text, followed by randomly shortening all texts to the same length using the Quantitative Index Text Analyser (QUITA) tool, which was also used for calculating the indices themselves.

We used multidimensional scaling (MDS) map (Figure 1) for the graphical representation of results, because its biggest advantage is in showing the texts that are similar to each other. Although it does not show us what these similarities are based on, we have all the numerical results available, thus we can find out what we need. Furthermore, this MDS was created (also in QUITA) using cosine distance, which means that the differences between numbers are not compared, but the distances between the texts in the map are made using ratios between numeric results. Therefore, we get a much more accurate data display, given the varying lengths of speech of all candidates. For better graph readability, it will be best to assign names to colours: leftmost in black colour is the official speech of President Miloš Zeman; in turquoise is the speech of Michal Horáček; red represents the speech of Pavel Fischer; grey is for Mirek Topolánek; green is for Petr Hannig; in blue is the speech of Marek Hilšer; a slightly light red colour is used for the private speech of President Miloš Zeman; purple is for Jiří Hynek; the slightly light black colour is for Jiří Drahoš; and yellow is for Vratislav Kulháněk.

It is evident that the speeches of Hynek and Drahoš are the most similar, followed by those of Fisher and Topolánek, and then followed by Zeman's private speech and Kulháněk's speech. It can be assumed that these closest pairs will have similarly rich or poor vocabulary to express themselves. The proximity of the candidates

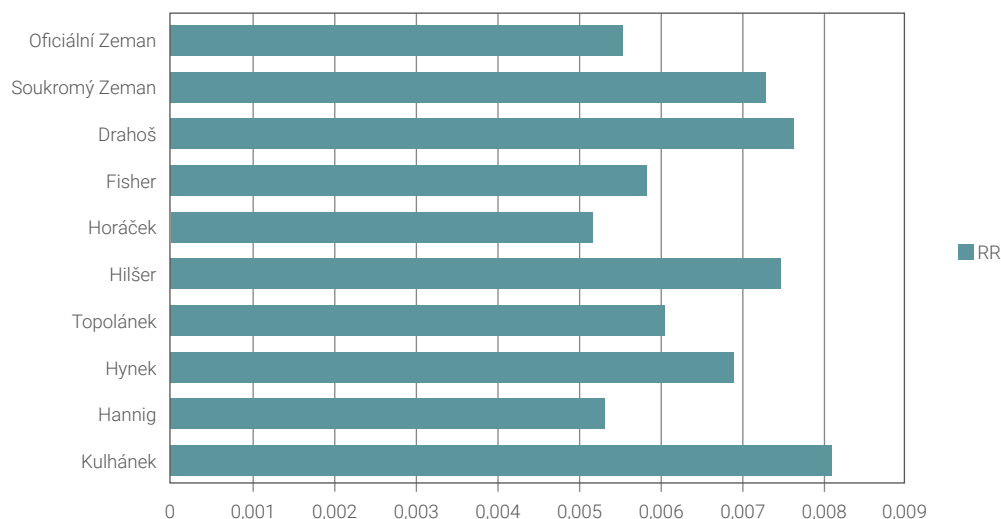


Figure 2: Repeat rate index.

can be searched back in the results of the analysis of the indices, e.g. the results of the RR index, which we can see in Figure 2.

The results of the word repetition index can be interpreted taking into account the qualitative part of the analysis, which is one of the purposes of our paper. For example, the quantitatively expressed proximity of Jiří Drahoš's speech and Miloš Zeman's private speech can be interpreted with the knowledge of their certain ease and the fact that they both spoke in a familiar setting and to almost-domestic audiences. In this context, it is not surprising, on the other hand, that the RR index differs from those of Marek Hilšer and Michal Horáček, because they were making their speeches in different settings. Both candidates held a press conference, but Horáček's one was very formal, and the speech was read, while Hilšer did not read his speech, spoke spontaneously and several anacolutha appeared in the speech.

No candidate spoke long enough to reduce his chances of using a new word. Nevertheless, there was a concern that different lengths of speeches were reflected in the completely different results of the word's wealth, namely the RR index. However, as is evident from Table 2 and Figure 2, differences in the length of the speeches do not correspond to differences in the word wealth; we can see that they do not match even differences in the RR index. Therefore, if the manifestations are not so different in their construction and richness of the formulation, it is possible to look for the distinctions in their contents.

THEMATIC CONCENTRATION

'Every author of any text focuses on a topic or topics which are represented by several autosemantic words. Thematic concentration measures how intensively the

author concentrates on the main theme(s) of the text' (Kubát, Čech 2016, 151). In other words, the main sense of the TC of the text 'is not to reveal the main themes of the text represented by the given linguistic units (although it allows) [...], but to describe the extent to which the author focuses on the subject or topics in general' (Čech et al. 2014, 14). Measurement of TC fully relies on the h-index (Hirsch's index¹⁸) and the words above. 'If we sort words by their decreasing frequency, h-point is defined as the place where the word order equals its frequency' (Čech et al. 2014, 15). Above the h-point, in theory, the words should be primarily synsemantic, being autosemantic beneath it.

If some autosemantics leak into the area of synsemantics, these words should be considered thematic words, 'a kind of anomaly that reflects the specific characteristics of the examined text, namely the strong "focus" (or "concentration") of the author on a particular topic (or topics), represented by the autosemantics found in the synsemantic area' (Čech et al. 2014, 16). However, it is very important to bear in mind that the h-index is a 'fuzzy boundary' (Kubát, Čech 2016, 153), it is not a rigid bound.

We are interested in TC in our paper, mainly because it is a quantitative metric closely related to the previous qualitative analysis of candidates for the President of the Czech Republic. We would like to be able to quantify the results that we were able to gain from analysing the same text using qualitative metrics. Now, with the help of QUITA, we can compare the numerical results of TC of individual speeches. In Figure 3, the candidates are lined up according to their success in the elections.

A low value of TC is indicative of nearly polythematic text and fragmented attention of the author, who addresses more themes superficially and does not go into depth in several key themes of his speech. Figure 2 shows that

18 Originally, Hirsch's index was used in scientometry to determine the citation response of scientific articles. In the Web of Science, the author's h-index corresponds to the serial number of the publication in the list of all author publications sorted in descending order by the number of citations in which the number of citations equals or exceeds its serial number. Thus, the author has a higher h-index value with a larger number of articles cited.

Initial Analysis of Presidential Candidate Speeches

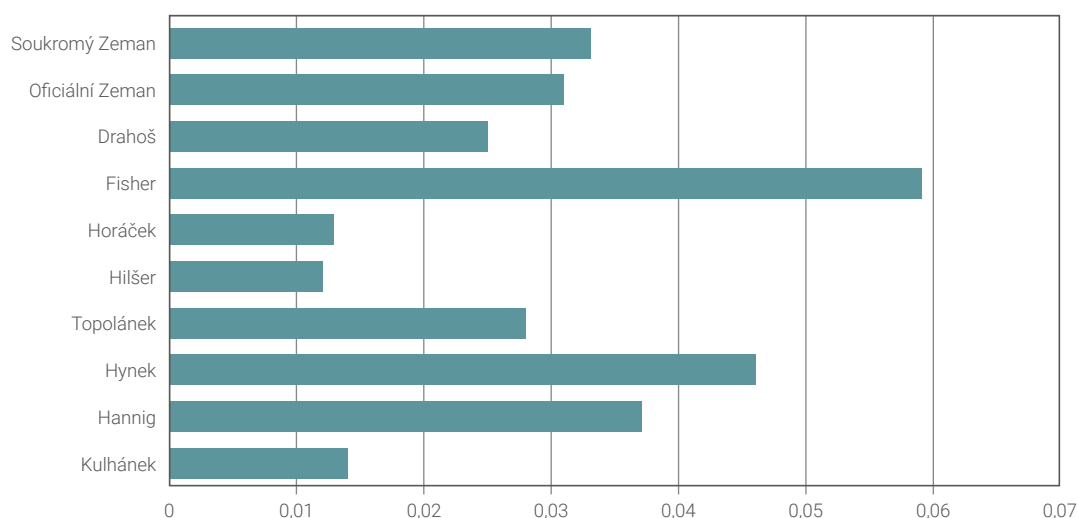


Figure 3: The results of thematic concentrations of presidential candidates' speeches.

scattered attention can be related to the speeches of Marek Hilšer, Michal Horáček and Vratislav Kulhánek. On the other hand, Pavel Fischer's speech is the most thematically concentrated one. Jiří Drahoš and Mirek Topolánek concentrate on themes in their speeches similarly, and it is worth noting that although they are always meant for different kinds of listeners and with different contents, the TC in both the speeches of Miloš Zeman is also very similar. This may be due to the fact that Miloš Zeman's rhetoric is generally characterised by stability over time, and therefore, his speeches show similar results, although they are created for different audiences with different contents. The TC of Miloš Zeman is closest to that of Petr Hannig (whose ambition was not to become a competitor of Miloš Zeman); his speech was mainly written, which may have affected the author's concentration. The second highest TC was of Jiří Hynek, which may be due to his focus on the army and military defence of the Czech Republic.

The high concentration on one topic in the case of Pavel Fischer is mainly because he devotes much of his speech to the support he has received from the Senate and Senators. He mentions that not all senators who support him have given him a signature, so he says he has wider support than may be apparent. Therefore, he not only mentions the Senators in his speech, but he explicitly praises them. On the other hand, Horáček and Hilšer devote their time to many topics, but each in a different way. With Marek Hilšer, this variance is caused by a general description of his life at the beginning of the speech; he also quite generally formulates his visions for the time of the presidency. Michal Horáček is more specific in formulations but covers more topics than Hilšer, talks directly about his relationship with voters and international politics, describes the financing of his campaign (this topic was not openly discussed by other candidates), which also disturbs the attention.

KEYWORDS

We have been able to find out the extent to which candidates were thematically focussed during their speeches, but what was most often thematised in their speeches has not yet been quantified. In our paper, we choose the analytic tool developed by the Czech National Corpus to do this quantification. The name of the tool is KWords, and the Czech National Corpus has developed it to identify the keywords of any text, regardless of the analyst's opinion. The analysed text is compared with the SYN2015 corpus containing 100 million words in fiction, journalistic texts and specialised literature (i.e. professional, non-fiction, administrative, and memoirs and biographies). Then, it evaluates the words that are significantly more frequent in the analysed text than in the common Czech text. The SYN2015 corpus was chosen as a reference frame because its content reveals what the Czech language has looked like in the past few years. Again, there are limits in that it is not possible to obtain pure text topics by this procedure. But with KWords, it is possible to work with keywords more deeply than by merely stating their simple enumeration. Using the keyword distribution in the text, it is possible to interpret the intentions of the speaker and, perhaps, even to set themes—different from those of the qualitative analysis methods. Interpretation is always limited to the first 10 keywords and their distribution in speech. The candidates are once more lined up according to their success in elections.

The diversity of the two speeches by Miloš Zeman has been mentioned many times. As we can see in Figure 4,¹⁹ there are words such as *free* (svobodný), *political* (politický), *proclamation* (prohlášení), *sentence* (věta) and *opinion* (názor), which could be linked primarily to politics, but in the public speech, there are words that relate not only to the presidential candidacy but mainly to his health.

19 All keywords in Miloš Zeman's private speech (English): 1. free; 2. sentence; 3. political; 4. proclamation; 5. want; 6. say; 7. opinion; 8. be; 9. this; and 10. have.

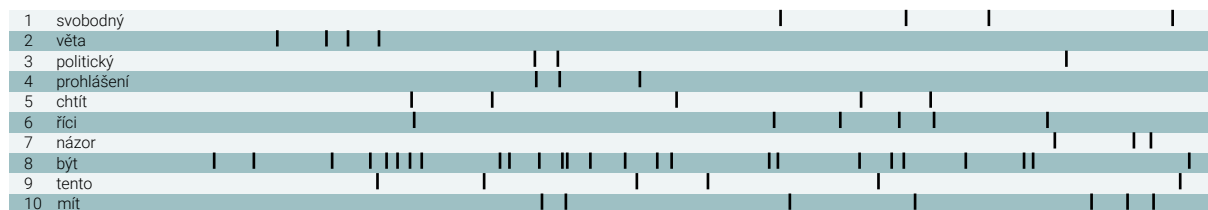


Figure 4: Keywords in Miloš Zeman's private speech.

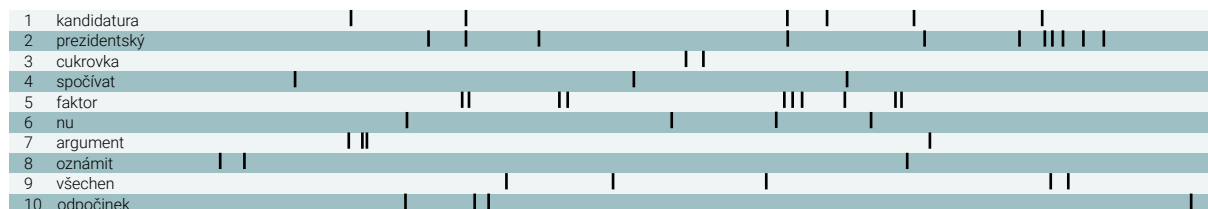


Figure 5: Keywords in Miloš Zeman's public speech.

In Figure 5,²⁰ we can see words such as *candidacy* (*kandidatura*), *presidential* (*prezidentský*) and *announce* (*oznámit*); all of them refer to Miloš Zeman's candidacy in the elections, but there are also words, such as *diabetes* (*cukrovka*) and *a rest* (*odpočinek*), connected to Miloš Zeman's state of health. The analysis of these aspects has been discussed in the earlier part of our article.

Jiří Drahoš's speech (Figure 6) is a proof that the graphical representation of the distribution of keywords²¹ is just as important as determining the keywords themselves, as well as previous knowledge of the circumstances of the speech. Without this knowledge, excessive use of *thank you* could be considered significant. But as the distribution depicts, the increased incidence of *thank you* at the very end of the text is only a way of trying to break the strong applause of his listeners and get in a few words to conclude. In general, the keywords in

Jiří Drahoš's speech are mainly the topic of his work at the Czech Academy of Sciences—*academy* (*akademie*), *chairman* (*předseda*). To some extent, it is interesting that Drahoš puts so much emphasis on himself, on his previous job and post in the academy, does not mention any of his supporters, anyone who himself was obliged during his candidacy, besides the citizens. He was nominated for a presidential election owing to their signatures. He presents himself as a strong independent candidate throughout his speech and creates an image of himself as a key factor in his candidature.

On the contrary, another candidate, Pavel Fischer, as has already been said, speaks primarily about his supporters. According to the keywords²² in Pavel Fischer's speech (Figure 7), this was a purely political speech reflecting his previous diplomatic profession. The words *ally* (*spojenec*), *debate* (*debata*), *reform* (*reforma*)

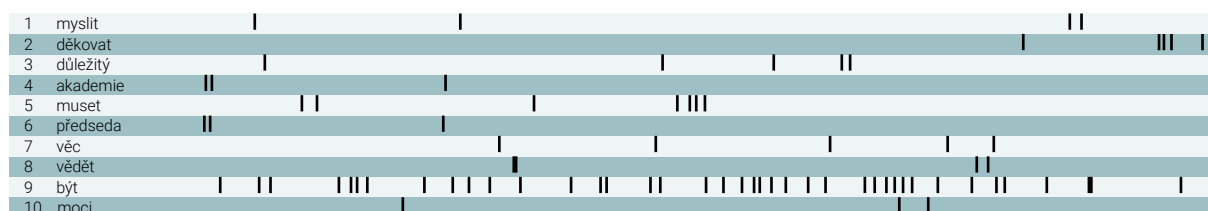


Figure 6: Keywords in Jiří Drahoš's speech.

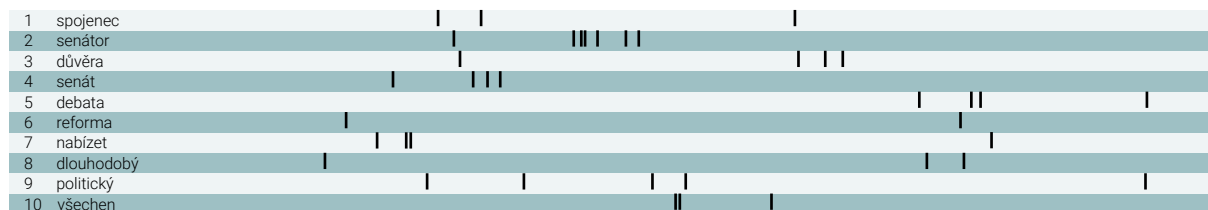


Figure 7: Keywords in Pavel Fischer's speech.

20 All keywords in Miloš Zeman's public speech in English: 1. candidacy, 2. presidential, 3. diabetes, 4. consist in, 5. factor, 6. Czech interjection, 7. argument, 8. announce, 9. all, 10. a rest.

21 All keywords in Jiří Drahoš's speech in English: 1. think, 2. thank, 3. important, 4. academy, 5. have to, 6. chairman, 7. thing, 8. know, 9. be, 10. be able to.

22 All keywords in Pavel Fischer's speech in English: 1. ally, 2. senator, 3. trust, 4. Senate, 5. debate, 6. reform,

Initial Analysis of Presidential Candidate Speeches

and *political* (politický) clearly refer to the topic of possible future presidency. The remaining keywords belong to the main supporters of Pavel Fischer—the senators. The words *senator* (senátor) and *senate* (senát) were mentioned more than 10 times during the whole speech, and thus, there is no doubt to whom Pavel Fischer’s gratitude belongs.

He is not the only candidate who got into the presidential election owing to the signatures of senators or deputies, but he is the only one who gives them a large space in his speech. He praises the senators immensely and considers them to be exceptional citizens and people. *The senators have succeeded in elections, have experience, experience the battle of personal integrity and battle with what has to be addressed in the region. Senators excelled as teachers or doctors, as mayors or professionals. The senator is primarily a personality and a personage, and only then a person is public and political, and therefore I respect their support and interest.*²³ Due to this perception of the senators, Pavel Fisher also claims the support of the citizens who are represented by these senators. He presents himself as a man firmly seated in the current system and the political situation to which he is planning to follow up.

In a previous analysis of Michal Horáček’s speech, mostly his personal conception of the presidency has been identified, and the keywords²⁴ (Figure 8) illustrate this image. Republic (*republika*), President (*prezident*) and state (*stát*) relate to the office itself, and task (*úkol*), want (*chtít*), mine (*můj*) and interest (*zájem*) refers to

Michal Horáček’s determination to acquire this office. Michal Horáček also presents himself as a strong and independent candidate, supported by citizens directly, but unlike Jiří Drahoš, he does not concentrate so much on himself. He focusses mainly on the presidential office and the Czech Republic itself. In his speech, he literally distributes what would be his task if he became President. He puts no emphasis on what he was but stresses on what lies before him.

Marek Hilšer emphasises the topic of the ordinary citizen, according to his keywords²⁵ (Figure 9), *citizen* (občan), *civic* (občanský), along with *accept* (uznávat), *bother* (vadit), *all* (všechn), *trust* (důvěra), *politics* (politika) and *discuss* (hovořit), which indicates that his topic involves civil society in top policy issues. In the case of Marek Hilšer, we can notice the high incidence of the word *want*, along with its interesting distribution—it has a regular spread throughout the length of the speech. Sometimes it repeats only a few words apart. Throughout the speech, he repeats what he wants to accomplish. Not only is it possible to conclude that Hilšer is presenting himself as a decisive person, who goes to candidacy for himself, without senators and without a political party, but at the same time, it is also a technique in which the speaker constantly repeats the important message that he wants to convey to his listeners during his speech. If the speaker wants to influence the listener and create a relatively clear impression of the message, there are two basic ways to accomplish this. Either the concepts are concentrated in one place, they are repeated several

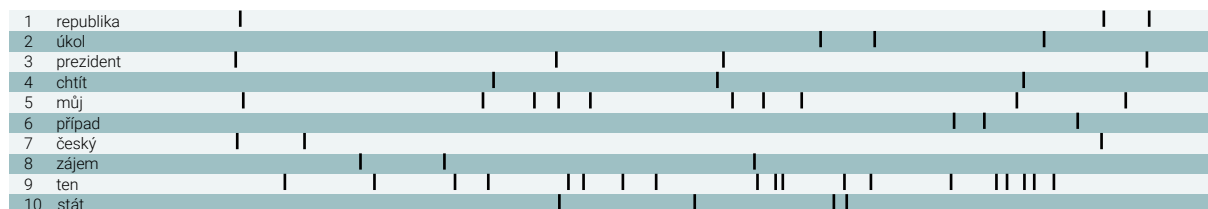


Figure 8: Keywords in Michal Horáček’s speech.

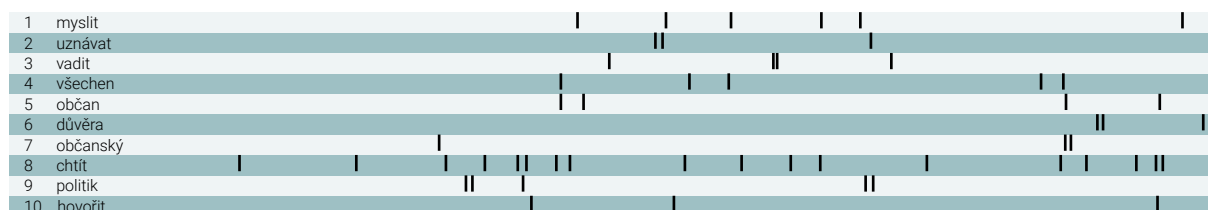


Figure 9: Keywords in Marek Hilšer’s speech.

7. offer, 8. long-term, 9. political, 10. all.

23 P. F.: “*Senátoři a senátorky prošli přímou volbou mají za sebou zkušenost, konfrontace té osobní integrity a konfrontace s tím co se v daném regionu bylo třeba řešit. Senátoři vynikli jako kantoři nebo lékaři jako starostové nebo odborní pracovníci. Senátor je především osobnost a figura a teprve potom člověk veřejný a politický, i proto si jejich podpory a zájmu vážím.*”

24 All keywords in Michal Horáček’s speech in English: 1. republic, 2. task, 3. president, 4. want, 5. mine, 6. case, 7. Czech, 8. interest, 9. this, 10. state.

25 All keywords in Marek Hilšer’s speech in English: 1. think, 2. accept, 3. bother, 4. all, 5. citizen, 6. trust, 7. civic, 8. want, 9. politics, 10. discuss.

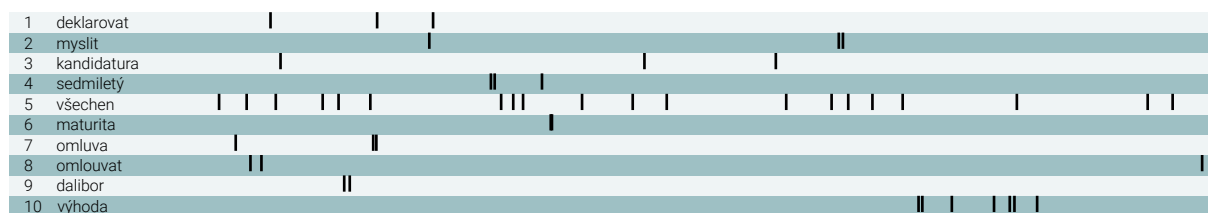


Figure 10: Keywords in Mirek Topolánek's speech.

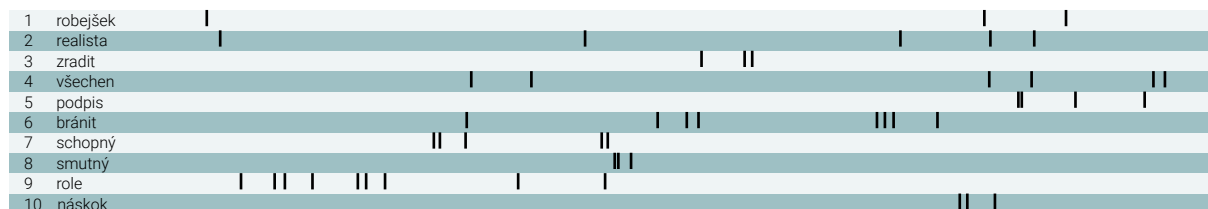


Figure 11: Keywords in Jiří Hynek's speech.

times, and thus they can be stored in the listener's memory, or the concepts will be repeated repeatedly and periodically, with some spacing, in order for the listener to memorise them.

In this way, Hilšer seeks, on the one side, to give the impression to the listener that he is the same as them, the same ordinary *citizen*, but at the same time, he is very determined to act decisively, resolutely and forcefully.

When we perceive individual candidates and their speeches not only as independent speakers, but put their words into context with those of the other candidates, it is possible to notice that some words are repeated, as we can see in Mirek Topolánek's keywords²⁶ in Figure 10. *Declare* (deklarovat) and *think* (myslit) have proved to be quite common words in the context of the speeches of Czech presidential candidates, and *candidacy* (kandidatura) is of course also on the list. But *apology* (omluva) and *try to excuse* (omlouvat) are found for the first time, as is the name Dalibor, which refers to Topolánek's co-worker Dalibor Veřmiřovský.

Although Mirek Topolánek is not widely known for his gentle way of expression, his speech can be regarded as conciliatory. It is strongly felt that this political event is a well-known setting for Topolánek to return to (after all, at the beginning of his press conference, he interacts with the journalists present there as with old friends). He must be

openly threatened with things from his political past, but it also suggests some humbleness to the President's office.

Candidate Hynek's political affiliation and beliefs can be read from the keywords²⁷ of his speech (Figure 11): *Robejšek* (the name of the chairman of the political party that supported Hynek in his candidacy); *realist* (realista, a reference to the aforementioned political party called Realists). Given that Realists are mainly involved in security and military issues, which Hynek also talks about in his speech, *to defend* (bránit) is not a surprising keyword in terms of military defence. This word is frequent in the last third of his speech, that is, where Hynek is presenting part of the political programme of Realists Party—the introduction of compulsory military training for men. Without doubt, he is the presidential candidate which is a representative of the political party in all its consequences.

The very emotional word *betrayal* is used in the context of the so-called Munich betrayal, which Hynek expresses unflatteringly (*when you want to earn respect, you never get it by kissing someone's butt*²⁸).

Like Hynek, Hannig recalls the historical anniversary of the Czech Republic in his speech, especially the ones during the wartime. Although it is not so noticeable in his keywords²⁹ (Figure 12) as much as it is in Jiří Hynek's keywords, Petr Hannig also speaks for his political party. He



Figure 12: Keywords in Petr Hannig's speech.

26 All keywords in Mirek Topolánek's speech in English: 1. declare, 2. think, 3. candidacy, 4. seven-year, 5. all, 6. graduation, 7. apology, 8. try to excuse, 9. (name) Dalibor, 10. throw away.

27 All key words in Jiří Hynek's speech in English: 1. (name) Robejšek, 2. realist, 3. betrayal, 4. all, 5. signature, 6. defend, 7. capable, 8. sad, 9. role, 10. head start.

28 J.H.: "Když chcete získat respekt, tak ho nikdy nezískáte tím, že někomu polezete do zadku"

29 All key words in Petr Hannig's speech in English: 1. candidacy, 2. illegal, 3. sensible, 4. appreciate, 5. inclination,

Initial Analysis of Presidential Candidate Speeches

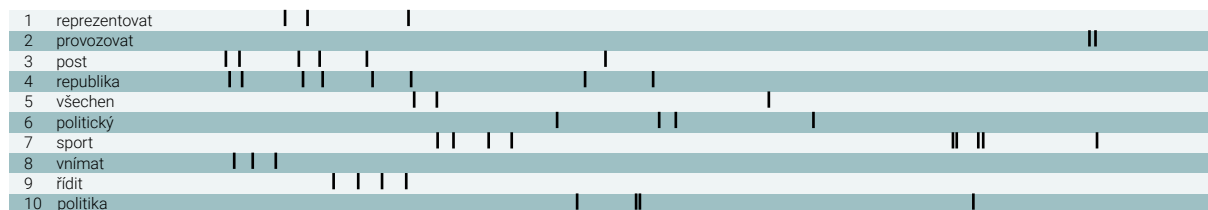


Figure 13: Keywords in Vratislav Kulhánek's speech.

points out that he only agreed with the candidacy, and the speech also focusses primarily on the security of the Republic and its defence with the help of the army: *illegal* (ilegální), *preserve* (zachovat).

In all these points, the speeches by Petr Hannig and Jiří Hynek are very close, unlike the speeches of the others; they devote themselves to narrow themes, which is reflected in the TC of their texts, according to which Hynek is the second highest and Petr Hannig the third highest of all candidates (Figure 3). Otherwise, the keywords of his speech show that he has also devoted himself to formulating his politics – *candidacy* (kandidatura), *inclination* (orientace), *Europe* (Evropa), *President* (prezident) and *politics* (politika).

The speech and overall self-presentation of Vratislav Kulhánek (Figure 13) is appropriately captured by the keyword³⁰ *sport* (sport), because Kulhánek greatly emphasised his connection with sport in terms of physical activity and the paragons associated with sport and in which he himself believes. As for the policy theme, the words *represent* (reprezentovat), *post* (post), *republic* (republika), *political* (politický) and *politics* (politika) also appear, and they are no different from their competitors. This is a bit of a surprise given how little he devoted himself to political topics in his speech. He was really primarily focussed on the issue of sport, and in the context of other speeches, it is a very rare exception. Indeed, throughout his speech, he seems to be running for the President of the Czech Republic so that children have a place to play.

CONCLUSION

Even though our sample data and their interpretations are limited, we will now summarise the fundamental deviations and trends that can be traced from the introductory speeches, and thus contribute in some way to the debate on the analysis of political speeches. In our paper, we focussed on analysing the presidential candidates' speeches at the beginning of their presidential election campaign in 2018. First of all, it was crucial to determine the basic characteristics of the manifestations, so that they could be taken into account in subsequent analytical steps. Among other things, the length of the speech in its spoken and written forms, along with the pace of the speaker, was identified. This identification was very important; without it, there would be many distortions (as

it was in the speech of Miloš Zeman for his supporters, which was characterised by the lowest rate of speech and the length of speech). For our analysis, the content level of the speeches was crucial. We were looking for topics that the candidates devoted themselves to. While developing the analysis, we also investigated the vocabulary richness of individual candidates, quantified their thematic concentration and, using the keywords found in their speeches, complemented the picture of their central themes.

In the qualitative identification of individual topics, it turned out that most of the candidates described their motivations and values during their speech. In addition to the current President, who said all this four years ago, all candidates introduced their vision of the presidential role and policy. Jiří Drahoš, who was second in the election, spoke mainly about state sovereignty. He spoke in a familiar environment, he did not repeat very much during the speech and according to the TC values, he was able to cover several topics in his short speech, for instance, the topics of the candidacy and his work at the Czech Academy of Sciences.

Pavel Fischer, who finished third in the election, spoke for a relatively long time about the Senate, its importance and the praise of the Senators themselves. Compared to other candidates, his speech shows the highest TC, which means concentration in one direction, in this case to Senators, who made his candidacy for the President legitimate with their signatures. Fischer does not present himself to voters as a leader, but as a team player.

Michal Horáček first spoke publicly with the announcement of his candidacy for the office of the President. He intended to pay and lead the whole campaign himself. He did not hesitate to comment on his opponents (in contrast, Drahoš and Zeman explicitly avoided it) and was very specific in his opinions. Horáček's speech was the second longest and was carefully structured. That allowed him to cover as many topics as possible within a limited time. The TC was the second lowest, which indicates a great polythematicity. His statements were decisive, he used words such as *radically* (radikálně), *I don't intend to* (nehodlám). It gives the impression of a strong and decisive leader.

Marek Hilšer appeared to be a more cautious leader, and he was generally very conciliatory, almost as if he didn't want to offend anyone; however, he could still give the impression that he was firmly decisive. He presented

6. Europe, 7. current, 8. preserve, 9. president, 10. Politics.

30 All key words in Vratislav Kulhánek's speech in English: 1. represent, 2. to run, 3. office, 4. republic, 5. all, 6. political, 7. sport, 8. perceive, 9. manage, 10. politics.

himself extensively to the audience and listed the things he wanted to do. Despite the absence of harsh words, he was able to present himself as a decisive man with a vision that was not supported by any political party. He promoted the involvement of ordinary citizens in politics. Thematically, the speech was the most fragmented of all.

Mirek Topolánek's speech was very different from that of other candidates, perhaps due to his long past in politics. He explicitly commented on Miloš Zeman that he would like him to *rest* and his personal theme was a relationship with Marek Dalík. Nothing similar to the speech of another candidate. But in terms of vocabulary richness, he was very close to Pavel Fischer. He was not as resolute as previous candidates, even though he presented himself very strongly throughout the speech. Furthermore, he named and thanked his sponsors at the end of the conference.

The candidates who ended up in the last three places had a lot in common. Jiří Hynek, Petr Hannig and Vratislav Kulhánek seemed very passive. Both Jiří Hynek and Petr Hannig were sent to their elections by their political parties, which they emphasise several times in their speeches. Both were focussed on the subjects of defence, national security and the military, and both recalled the sad historical experience of our country in their speeches. Vratislav Kulhánek talked about the need for collective sports and their difficult accessibility, which was the main content of his speech. But as far as vocabulary richness and TC are concerned, their values were not as close as one could expect from the initial analysis of their speeches.

In our paper, we managed to evaluate presidential candidates mainly in terms of their topics and vocabulary, which we interpreted using the characteristics previously identified. In general, the success of the strategy of looking at the candidacy from a purely personal point of view can be observed, but we have narrowed our analysis by using only a limited number of analytical methods and examining only selected parts of candidate speeches; therefore, this analysis cannot be considered as a summary; on the contrary, it could serve as a partial contribution, which we add to the current debate on the interpretation of political manifestations.

ACKNOWLEDGEMENTS

This work was supported by the student project 'Comprehensive text analysis on the example of the Czech presidential campaign 2018' at the Faculty of Arts of the Palacky University (IGA_FF_2017_035).

REFERENCES

Kořenský, J., Hoffmannová, J., Müllerová, O., 2013. *Analyza komunikačního procesu a textu: učební text pro výběrové semináře filologických oborů*. Olomouc: Univerzita Palackého v Olomouci, Qfwfq.
 Wodak, R., Meyer, M. (Eds.), 2009. *Methods of critical discourse analysis*. 2nd ed. Thousand Oaks, CA: Sage.

Fairclough, N., 2003. *Analysing discourse textual analysis for social research*. London: Routledge.
 Čech, R., Popescu, I., Altmann, G., 2014. *Metody kvantitativní analýzy (nejen) básnických textů*. Olomouc: Univerzita Palackého v Olomouci, Qfwfq.
 Kubát, M., Milička J., 2013. Vocabulary richness measure in genres. *Journal of Quantitative Linguistics*, 4, 339–349.
 Hoover, D., 2003. Another perspective on vocabulary richness. *Computers and the Humanities*, 37, 151–178.
 Čech, R., 2016. *Tematická koncentrace textu v češtině*. Prague: Ústav formální a aplikované lingvistiky.
 Kubát, M., Čech, R., 2016. Thematic concentration and vocabulary richness. In Kelih, E., Knight, R., Mačutek, J., Wilson, A. (Eds.), *Issues in quantitative linguistics 4*. Lüdenscheid: RAM, pp. 150–159.
 Kubát, M., Matlach, V., Čech, R., 2014. *QUITA – Quantitative index text analyzer*. available at: < <https://www.quitaonline.com/> >.
 Cvrček, V., Vondříčka, P., 2013. *KWords*. FF UK, Praha, available at: < <http://kwords.korpus.cz> >.

APPENDIX

For the sake of completeness, this section presents the results of the first and second rounds of the election of the President of the Czech Republic in 2018.

| Candidate | Number of votes | Number of votes in % |
|--------------------|-----------------|----------------------|
| Miloš Zeman | 1 985 547 | 38,56 % |
| Jiří Drahoš | 1 369 601 | 26,60 % |
| Pavel Fischer | 526 694 | 10,23 % |
| Michal Horáček | 472 643 | 9,18 % |
| Marek Hilšer | 454 949 | 8,83 % |
| Mirek Topolánek | 221 689 | 4,30 % |
| Jiří Hynek | 63 348 | 1,23 % |
| Petr Hannig | 29 228 | 0,56 % |
| Vratislav Kulhánek | 24 442 | 0,47 % |

Table A1: The results of the first round of the election of the President of the Czech Republic in 2018

| Candidate | Number of votes | Number of votes in % |
|-------------|-----------------|----------------------|
| Miloš Zeman | 2 853 390 | 51,36 % |
| Jiří Drahoš | 2 701 206 | 48,63 % |

Table A2: The results of the second round of the election of the President of the Czech Republic in 2018.