



**HAL**  
open science

# New Year's Day Speeches of Czech Presidents: Phonetic Analysis and Text Analysis

Milan Jičínský, Jaroslav Marek

► **To cite this version:**

Milan Jičínský, Jaroslav Marek. New Year's Day Speeches of Czech Presidents: Phonetic Analysis and Text Analysis. 16th IFIP International Conference on Computer Information Systems and Industrial Management (CISIM), Jun 2017, Bialystok, Poland. pp.110-121, 10.1007/978-3-319-59105-6\_10. hal-01656221

**HAL Id: hal-01656221**

**<https://inria.hal.science/hal-01656221v1>**

Submitted on 5 Dec 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# New Year's Day speeches of Czech presidents: phonetic analysis and text analysis

Milan Jičínský, Jaroslav Marek

Univerzita Pardubice, Fakulta elektrotechniky a informatiky  
Studentská 95, 530 02 Pardubice I, Czech Republic  
milan.jicinsky@student.upce.cz, Jaroslav.Marek@upce.cz

**Abstract.** The aim of our study is verification of programmed algorithms of phonetic analysis using concrete data, and reassurance that it works as also sought after. For our testing, the appropriate recordings of New Year's Day speeches of Czech and Czechoslovak presidents are available. The very first available recording of presidential speech comes from 1935. All transcripts and recordings of the last 87 speeches are located on the web page [www.rozhlas.cz](http://www.rozhlas.cz). The primary goal of this paper is to analyze voice characteristics of the speaker (log energy, speech velocity and Zero crossing rate). Especially words "with greatest energy" will be found. There will be a list of words having the highest energy values. The most interesting results will be presented by graphical tools. Using a software, capable of text analysis, transcript characteristics such as most frequent words, length of words, total number of words and different words will be computed. The most frequent words will be presented. Political speeches often become the subject of various analyses. Our calculation allows a new perspective on speeches. It is interesting to compare the most frequent semantic words and words with the greatest energy. The results can be historically important. It allows an extraction of new information from available data and scientifically different approach.

**Keywords:** New Year's Day speeches of Czech presidents, intensity of speech, vocabulary richness, speech velocity, Zero crossing rate, length of words

## 1 Introduction

Nowadays analysis of speech is very popular. It started in the second half of 20th century when basic signal characteristics were discovered. Fundamentals are well described in [1-3]. This field of study evolved very quickly so there are many applications from key word detection [4] across transcription of fluent speech [5] to recognition of speaker [6]. This article is reserved for those who want to study phonetic analysis and its fundamentals. Even historians whose field of study is 20th century in the Czech Republic can appreciate the most frequented words and words with the highest energy. Linguists can be satisfied with changes in individual speeches whether it's written or spoken.

Many authors aim their work on linguistic analysis of political speeches. For example, articles End-of-year speeches of Italian presidents or inaugural speeches of US

presidents were researched in [7, 8]. Relationship between ideology and language and thematic concentration of Czechoslovak New Year's Day speeches is analyzed in article [9]. Of course, it is very interesting to study influences of ideology, originality of author and his abilities to differ from uniformity. The most frequent words can provide an information about recent years because they react on the most important events. Some of those words will be listed.

Main goal of this publication is to present words that were said with the greatest energy. Words with the greatest energy allows to track what president emphasized on during a reading. Emphasis of the speaker will be probably on positive words. The only exception could be the time of war. Ideological words may be emphasized in some speeches. There will be one more characteristic calculated for each speech – speech velocity.

## 2 Voice characteristics

Analysis of recordings of New Year's Day speeches will be introduced in this chapter. The intensity of voice (energy) and speech velocity will represent voice characteristic of speaker. Energy tells something about how much emphasis speaker uses and speech velocity shows how fast speaker speaks. These variables can be influenced e.g. by age or by sickness. Then the words having the greatest energy can be found. It could be interesting to compare these words with most frequent thematic words. President didn't have to be an author of written text. But he could highlight any words he wanted to. It depended on what he considered to be important. This is the example of individuality. Then ZCR (zero crossing rate) characteristics will be shown.



Figure 1: “Dear fellow citizens” – Václav Havel (1998). Source: own.

### 2.1 Obtaining data and its processing

Source data have been obtained from website [www.rozhlas.cz](http://www.rozhlas.cz). Speeches are recorded with useful software Audacity. Sampling rate of each speech is 8 kHz. Each recording is modified because the original ones contain a music before the speech starts. Calculations of voice parameters are realized in MATLAB. Scheme of processing can be simplified as on Figure 3.

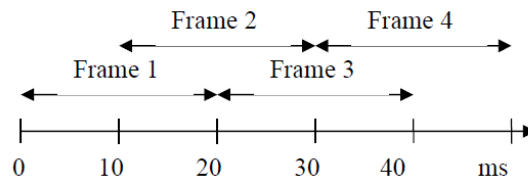


Figure 2: Segmentation of recording into frames. Source: own.

Segmentation means that the record is divided into frames of the same length (typically 20 ms long). Frames are overlapping each other right in the half (in this case). Overlapping is recommended to the fact that parameters can be changed in jumps. So this enhancement improves the dramatic changes and it can describe even changes near an edge of frame without loss of useful data. After the segmentation follows a parameterization. Feature vector values of each frame is computed during the parameterization. Features can be divided into: basic, spectral, cepstral and dynamic. ZCR and energy ranks among the basic features. Feature extraction and segmentation is also discussed in [2].

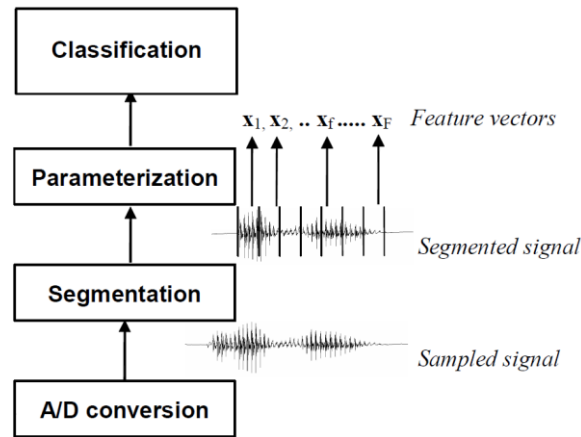


Figure 3: Audio processing. Source: own.

## 2.2 Intensity of voice

The intensity of voice is characterized by energy. So the energy is a key parameter which defines the intensity of voice. Energy is defined as the sum of squared values of samples within one frame. Logarithm function is used for better range of energy values. In this case Log energy of ordinary noise is around 5. Whenever speech is contained in recording, values of energy are greater for those frames. Typically, the energy of speaking person can reach even value of 15. It depends on how loud speaker speaks. Log energy is defined as

$$E = \log\left(\sum_{n=0}^{L-1} x^2(n)\right), \quad (1)$$

where L is the frame length, concretely the number of samples contained in the frame.  $x(n)$  is the designation for the current sample value.

In comparison of all presidents it's evident to see that president Hácha spoke not as loud as others. He had no emphasis. This could be caused by political situation. Hácha used to be a president during the hardest time of the Czechoslovak history. He was helpless president of protectorate state. The only thing he could do was to make peo-

ple feel calm and safe, even if it wasn't possible. As for president Husák, very significant decrease of energy was observed between years 1978 and 1979.

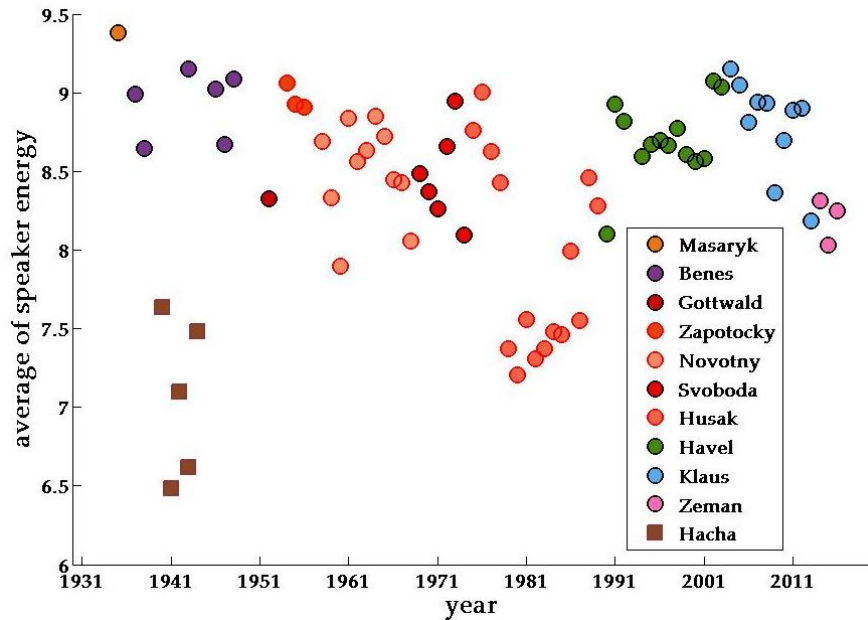


Figure 4: Log energy: mean value. Source: own.

### 2.3 ZCR

Zero crossing rate is a parameter that characterizes changing of sign from negative to positive or back. Zero crossing rate is related to the frequency. There is one value of ZCR for each frame (the same as for the energy). The principle of ZCR can be easily explained with Figure 5. ZCR value is equal to the count of all dots. The dots are placed to the points where signal intersects x axis and changes the sign.

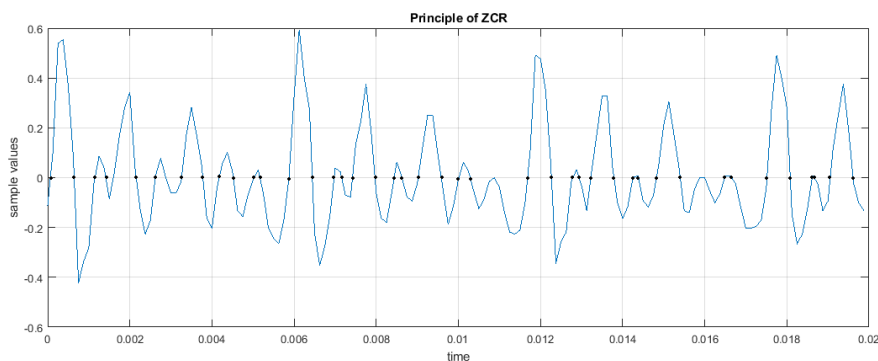


Figure 5: Explanation of ZCR - 20ms frame containing of phoneme "a". Source: own.

It's often used for voice activity detection [11] – to find out if human speech in record is present or not. As for voiced signal ZCR values are typically low. Noises and unvoiced signals have higher values. This method is sensitive to noises and direct component shifts. It even allows us to find out if concrete phoneme is voiced (b,d,g,z,v,h,...) or unvoiced (p,t,k,s,f,ch,c,...). Especially the sibilance (s,c,š,č,...) have higher ZCR values.

$$ZCR = 1/2 \sum_{n=1}^{L-1} |\text{sgn } x(n) - \text{sgn } x(n-1)|. \quad (2)$$

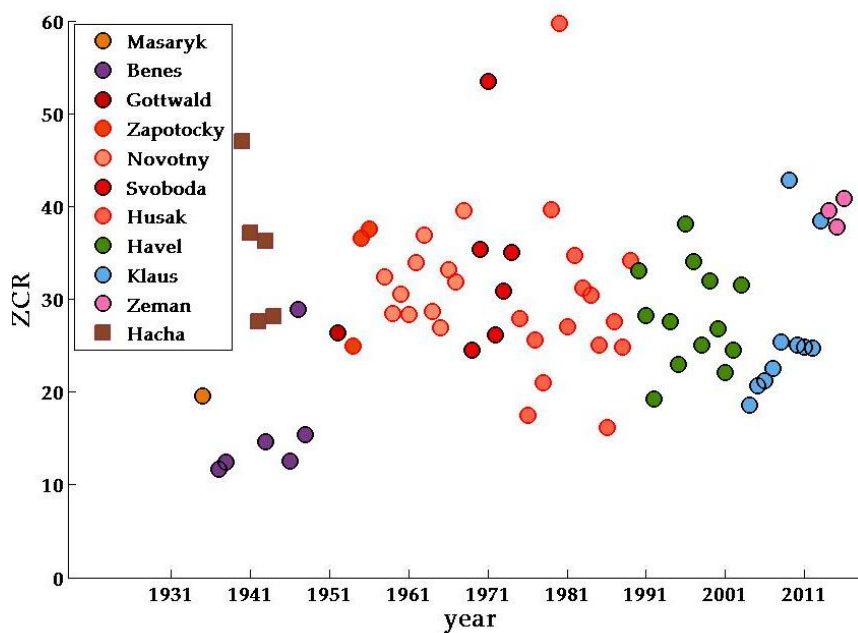


Figure 6: Zero crossing rate: mean value. Source: own.

Data variability is relatively high. So, the mean value of ZCR is not that good to represent individual speaker. Better results can be obtained using ZCR dynamically. That means ZCR of each frame is used. Then search for dynamic changes instead of treating it as one static value. It's preferable to use it for each frame.

## 2.4 Speech velocity

For the purposes of the article there is a created parameter that can be used to link results of text and voice interpretation into one value that characterizes the speaker. It's called speech velocity. This mean value represents how many words the speaker pronounces during the time of one second. The speech of president Husák from 1989 is significantly the slowest. President Háša is speaking relatively slowly too. On the

other hand, the fastest tempo of speaking can be recognized in speeches from 1938, 1943 (Beneš), 1959 (Novotný) and 1996 (Havel).

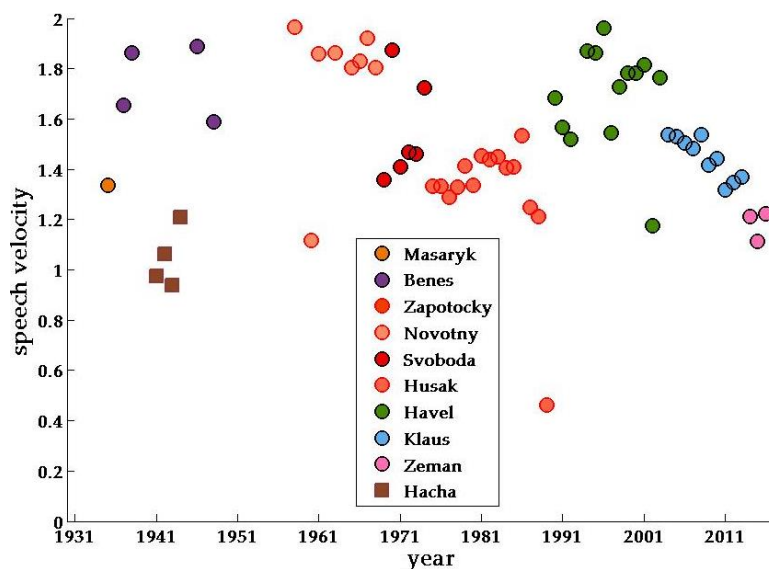


Figure 7: Speech velocity: mean value. Source: own.

### 3 Characteristics of written text

All studies are realized for 87 speeches of Czechoslovak presidents, Czech presidents or Czechoslovak prime ministers. The unique situation happened due to the World War II. The Czechoslovak Republic had two presidents. President Beneš left his country and exiled to the Great Britain. But he was still very politically active. Then Hácha was chosen to be a president of protectorate. So, both groups of speeches were analyzed between 1940 and 1945. On the other hand, president in exile is considered to be more important subject of our analysis.

We can expect changes in using of different length words during a long history of New Year's Day speeches. Therefore, the first aim of our calculation is to determine average of word length. Calculations of text parameters are made using software based on Java [10] called „Statistika v lexikální analýze“. This GUI (Graphical User Interface) has been created during diploma thesis. It makes easier the whole text processing. The software allows to analyze frequent letters and words, length of words, aggregation and alliteration and some other features. The original purpose of existence of software is analysis of poems and its translations as in [12].

#### 3.1 Mean of word length

On the Figure 8 can be seen mean values of length of words of analyzed texts. Length of words of communistic presidents (Zápotocký, Novotný, Husák) is much

greater than nowadays (Havel, Klaus, Zeman). President Beneš used the shortest words. The greatest variability can be seen at Svoboda and Hácha. Estimate of expected value is given by

$$\bar{x} = 1/n \sum_{i=1}^k if_i, \quad (3)$$

where  $i = 1, \dots, k$  is length of word,  $k$  is length of the longest word.

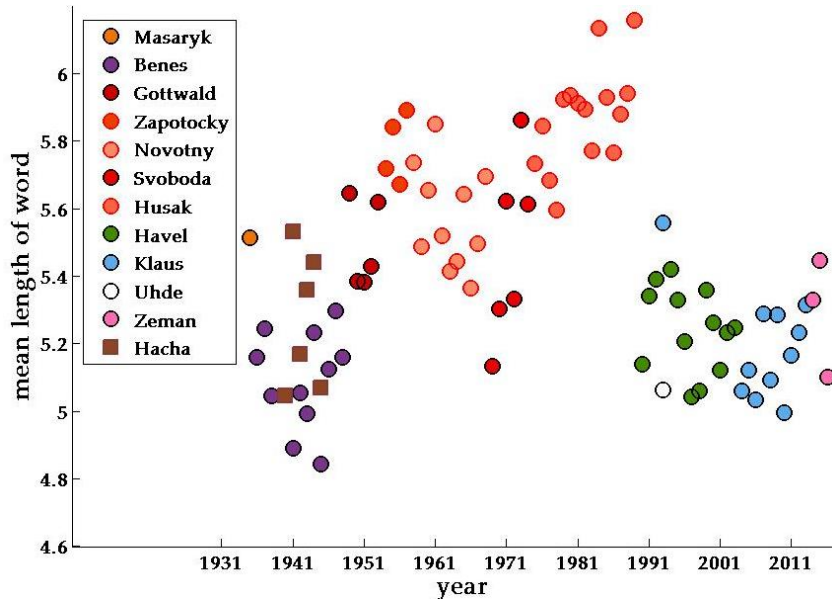


Figure 8: Length of words: mean value. Source: own.

### 3.2 The most frequent words

Conjunctions and prepositions of course ranks among the most frequent words. Conjunction “a” is the most frequent in all speeches except Novotný (1964) – “v”, Svoboda (1973) – “v” and Hácha (1944) – “se”. Figure 10 is the list of sorted conjunctions and prepositions used from first to fourth position. The common words can be seen. The most frequent words with meaning will be presented in Figure 10 too. These words differ much more than prepositions. Presidents react on current political events such as crisis, protectorate, war or return of democracy. Meaning words can provide a quick preview of content. Comparison with inaugural speeches of US presidents can be interesting. As for words with meaning, for example Roosevelt (1933) said words: HAVE, NATIONAL and Truman (1949) used the words: WORD, HAVE, NATIONS, PEACE, FREEDOM, PEOPLE, FREE, UNITED, MORE, SECURITY, DEMOCRACY [7]. See Figure 9 or 10.



<b>Tomáš Garrigue Masaryk (1850-1937)</b>				
1935: a, se, na, v	národů	států	pokoj	V HOSPODÁŘSKÉM i politickém
<b>Edvard Beneš (1884-1948)</b>				
1936: a, aby, v, i	národ	prezident	stát	-
1937: a, v, se, že	demokracie	evropě	dnešní	BUDEME
1938: a, v, je, co	rok	státu	dobře	NA POLI politickém
1941: a, v, se, na	bude	zase	vánoce	-
1942: a, se, to, že	války	vítěství	dnes	-
1943: a, že, v, se	bude	německo	dnes	KDYBY SNAD tak zvaná slovenská vláda
1944: a, v, to, se	naší	svazu	sovětského	-
1945: a, v, na, že	bude	dnes	roku	-
1946: a, v, se, to	nové	dnes	války	po utrpeních VÁLEČNÝCH z rukou nepřátel
1947: a, v, na, o	zjevy	bude	dnešní	o zrození ČLOVĚKA nového
1948: a, i, na, se	život	člověk	roku	I TO bude našemu životu prospěšné
<b>Emil Hácha (1872-1945)</b>				
1939: -	-	-	-	NEMŮŽE
1940: a, z, i, v	srdce	dětem	večer	V PLNÉM kruhu svých nejbližších
1941: a, v, se, na	národní	dobu	národa	SVĚTOVÝ přerod
1942: a, na, v, jim	vánoce	radosti	dnes	vedle ZÁŘÍČÍHO vánočního stromku
1943: se, v, a, všech	říše	vánoce	děj	NEVÁHEJTE
1944: a, na, že, v	říše	národu	spoluobčané	-
1945: a, se, že, v	válka	vánoce	radosti	-
<b>Klement Gottwald (1896-1953)</b>				
1949: a, v, k, na	našeho	plánu	roku	-
1950: a, o, v, že	procent	roku	více	-
1951: a, v, na, o	roce	naše	minulém	-
1952: a, v, na, se	roce	více	průmyslu	UBRÁNILI mír
1953: a, v, že, na	míru	sovětského	války	-
<b>Antonín Zápotocký (1884-1957)</b>				
1954: a, v, i, na	hospodářství	rutno	mír	KULTURNÍ úroveň našeho lidu
1955: a, i, v, se	míru	národní	republiky	naše mírumilovné BUDOVANÍ
1956: a, i, v, se	budování	míru	republiky	zaměstnancům V LIDOVÉ i státní správě
1957: a, i, v, na	lidu	roku	uhlí	-
<b>Antonín Novotný (1904-1975)</b>				
1958: a, v, se, na	strany	lidu	kommunistické	PŘILOŽÍME-LI v těchto ruku
1959: a, v, na, o	výsledky	práce	dnes	cílem je BUDOVIAT socialistický národ
1960: a, v, se, na	zemědělství	roce	československa	jaké JSOU dosavadní výsledky
1961: a, v, se, na	socialistické	roce	společnosti	UPEVNŮVAT naši československou
1962: a, v, na, se	hospodářství	roku	rok	OBRAZ růstu životní úrovně
1963: a, na, v, se	spředu	usnesení	socialistické	JEJICH cílem je vytvořit
1964: a, v, na, i	hospodářství	národní	roce	CO DNES můžeme říci
1965: a, v, i, na	hospodářství	společnosti	roce	VYBUDOVALI jsme socialismus
1966: a, v, na, se	společnosti	hospodářství	vývoj	BUDOU-LI chítit národní výbory
1967: a, v, na, že	hospodářství	životní	roce	UVOLNĚNÍ mezinárodního napětí
1968: a, v, i, na	republiky	socialismu	národní	PUJDE-LI i letos hospodářský vývoj
<b>Luďvík Svoboda (1895-1979)</b>				
1969: a, v, i, na	rok	práce	lidu	Meré přes všechna DRAMATA
1970: a, v, i, to	země	rozvoje	roce	ZABEZPEČUJÍCÍ všestranný rozvoj
1971: a, i, v, se	socialistické	práce	rozvoje	KAZDĚHO z nás
1972: a, v, vám, i	rok	program	kommunistické	PŘEJI Vám všem
1973: v, a, se, i	roce	úspěchů	mezinárodní	USILUJEME o rozvoj
1974: a, se, v, pro	kommunistické	rozvoje	roku	díky UVĚDOMĚLEMU pochopení
<b>Gustav Husák (1913-1991)</b>				
1975: a, v, i, se	lidu	roce	hospodářství	všech MÍROVÝCH sil ve světě
1976: a, v, i, na	lidu	československa	národní	přičině se všichni
1977: a, v, i, na	lidu	práci	lidu	SVĚDOMÍM odpovědných hospodářů
1978: a, v, i, se	lidu	rok	státu	PUSOBÍ vnější dopady na naši ekonomiku
1979: a, v, se, i	lidu	roku	socialistické	V SOULADU se záměry plánu
1980: a, v, se, na	lidu	roku	socialistického	SOUČASNĚ jsou zdrojem jistoty
1981: a, v, i, se	lidu	práce	roce	ve ZKUŠENOSTI s budováním
1982: a, v, se, s	socialistického	lidu	práce	UROVENN a řízení a kvalitě veškeré práce
1983: a, v, na, i	lidu	státní	roce	K ČILÍ však může vést
1984: a, v, pro, se	lidu	světě	přítel	PŘIJALI jsme nezbytné závěry
1985: a, v, na, se	roku	národní	vlasti	ŠIROKÁ společenská a pracovní aktivita
1986: a, v, se, na	práce	lidu	roku	VYKONANOU práci přitom hodnotíme
1987: a, v, se, i	roku	národní	socialistické	hodnotíme VSAK realisticky
1988: a, v, je, na	práci	roku	národní	uvítali jsme UPRÍMNĚ výsledky
1989: a, v, na, k	společnosti	zasedání	práci	VYUŽÍVÁME to, co je možné uplatnit
<b>Václav Havel (1936-2011)</b>				
1990: a, se, že, v	republike	země	národy	BYLO by velmi nerozumné
1991: a, v, se, že	rok	věřím	systém	MILÍ spoluobčané
1992: a, se, v, že	rokem	státní	hledáme	TO všechno je samozřejmě dobré
<b>předseda vlády Václav Klaus (1941)</b>				
1993: a, se, v, na	české	republiky	stát	-
<b>předseda parlamentu Milan Uhde (1936)</b>				
1993: a, se, je, že	domov	stát	život	-
<b>Václav Havel (1936-2011)</b>				
1994: a, se, v, na	stát	občanské	společnosti	než ÚČTU ke všemu nestandardnímu
1995: a, se, v, je	stát	politiky	naděje	I NADEJE je práce, práce možná obtížná
1996: a, se, že, v	zemi	státu	společnosti	MYSLÍM, že i soudobý vývoj
1997: a, k, se, je	znamení	člověk	věci	což NENÍ zároveň bytostí
1998: a, se, to, v	politika	znamená	věci	MILÍ spoluobčané
1999: a, se, v, i	zdi	evropy	lety	MIRA našeho zápalu pro vízi
2000: a, se, i, že	život	světa	lidí	VÍM, jak to zní nepopulární
2001: a, v, je, se	zájmu	lidí	společnosti	VŠICHNI přece víme
2002: a, se, o, že	státem	evropské	země	POMOHLO by nám, uvědomit si
2003: a, se, v, je	státu	rok	soužití	MILÍ spoluobčané
<b>Václav Klaus (1941)</b>				
2004: a, se, v, i	rok	zemi	české	že TÍMTO okamžikem
2005: a, se, v, na	rok	stát	země	UČÍME se rozlišovat
2006: a, v, se, že	evropské	život	roku	po ODMĚTNUTÍ EVROPSKÉ ústavy
2007: a, se, v, že	roku	evropské	nie	NESMÍ nám brát hostejné
2008: a, se, v, i	rok	země	přicházejí	chtěl bych Vás UBEZPEČIT
2009: a, v, se, k	rok	unie	evropské	URČITÉ tímto kouzelným proutkem
2010: a, se, v, to	rok	pokles	občanů	URČITÉ lépe než bylo
2011: a, se, v, to	rok	vláda	hlasům	UVNITŘ společnosti
2012: a, se, v, že	rok	zemi	evropě	NAROKY bývají posilovány
2013: a, v, je, se	den	zemi	rok	UVNITŘ obcí
<b>Miloš Zeman (1944)</b>				
2014: a, se, že, jsem	roku	poděkovat	život	NEMLUVÍM teď o rozdělení společnosti
2015: a, se, že, jsem	státní	řídky	rok	MÁM-LI hodnotit vládu
2016: a, že, je, se	země	roce	republiky	DRUHOU dobrou zprávou

Figure 9: The most frequent words (Czech version). Source: own.

Tomáš Garrigue Masaryk (1850-1937)				
1935: and, with, on, in	nations	states	serenity	ECONOMIC and political
<b>Edvard Beneš (1884-1948)</b>				
1936: and, that, in, even	nation	president	state	-
1937: and, in, with, that	democracy	Europe	today	- we WILL
1938: and, in, is, what	year	state	good	political ARENA
1941: and, with, to, that	will	again	Christmas	-
1942: and, with, to, that	war	victory	today	-
1943: and, that, in, with	will	Germany	today	MAYBE IF so-called Slovak Government
1944: and, in, to, with	our	union	soviet	-
1945: and, in, on, that	will	today	year	-
1946: and, in, with, it	new	today	war	after WAR suffering at the hands of enemies
1947: and, in, on, about	apparitions	will	today	the birth of a new MAN
1948: and, even, on, with	live	human	year	AND IT will be beneficial to our lives
<b>Emil Hácha (1872-1945)</b>				
1939: -	-	-	-	CAN NOT
1940: and, from, even, in	heart	child	night	FULL family circle
1941: and, in, with, on	national	period	nations	WORLD rebirth
1942: and, in, on, the	Christmas	pleasures	today	beside the GLOWING Christmas tree
1943: with, in, and, all	empire	Christmas	plot	DO NOT HESITATE
1944: and, on, that, in	empire	nation	pleasures	-
1945: and, with, that, in	war	Christmas	joy	-
<b>Klement Gottwald (1896-1953)</b>				
1949: and, in, to, on	our	plan	year	-
1950: and, about, in, that	percent	our	more	-
1951: and, in, on, about	year	our	last	-
1952: and, in, on, with	year	more	industry	WITHSTOOD peace
1953: and, in, that, on	peace	soviet	war	-
<b>Antonín Zápotocký (1884-1957)</b>				
1954: and, in, even, on	economy	necessary	peace	CULTURAL level of our people
1955: and, even, in, with	peace	nation	republic	our peaceful BUILDING
1956: and, even, in, with	building	peace	republic	employees IN FOLK and state
1957: and, even, in, on	people	year	coal	-
<b>Antonín Novotný (1904-1975)</b>				
1958: and, in, with, on	party	people	communist	If we ENCLOSE hand
1959: and, in, on, about	results	work	today	the aim is to BUILD a socialist order
1960: and, in, with, on	agriculture	year	czechoslovakia	What ARE the existing results
1961: and, in, with, on	socialistic	year	society	STRENGTHEN our czechoslovac
1962: and, in, on, with	economy	year	socialistic	IMAGE of living standards growth
1963: and, on, in, with	congress	decree	socialistic	THEIR goal is to create
1964: and, on, even	economy	year	nation	WHAT TODAY can we say
1965: and, in, even, on	economy	society	year	We have BUILT socialism
1966: and, in, on, with	society	economy	progression	IF THEY WANT national committees
1967: and, in, on, that	economy	vital	year	THAW of international tensions
1968: and, in, even, on	republic	socialism	national	IF IT GOES this year economic development
<b>Ludvík Svoboda (1895-1979)</b>				
1969: and, in, even, on	year	work	people	that despite all the DRAMA
1970: and, in, even, it	country	advancement	year	ENSURING all-round development
1971: and, even, in, with	socialistic	work	advancement	EACH of us
1972: and, in, You, even	year	program	communist	I WISH you all
1973: in, and, with, even	year	achievements	international	We SEEK to develop
1974: and, with, in, for	communist	advancement	year	thanks CONSCIOUS understanding
<b>Gustav Husák (1913-1991)</b>				
1975: and, in, even, with	people	year	economy	All PEACEKEEPING force in the world
1976: and, in, even, on	people	czechoslovac	national	LET US ALL
1977: and, in, even, on	work	people	national	AWARE responsible landlords
1978: and, in, even, with	people	year	public	OPERATES external impacts on our economy
1979: and, in, with, even	people	year	socialistic	in ACCORDANCE with the intent Plan
1980: and, in, with, on	people	year	socialistic	CURRENT are a source of security
1981: and, in, even, with	people	work	year	in-building EXPERIENCE
1982: and, in, with, with	socialistic	people	work	LEVEL, quality, management of all labor
1983: and, in, on, even	people	states	year	however, it lead to the DESTINATION
1984: and, in, for, with	people	world	friends	we ADOPTED the necessary conclusions
1985: and, in, on, with	year	nation	motherland	BROAD social and work-related activity
1986: and, in, with, on	work	people	year	work DONE we simultaneously evaluate
1987: and, in, with, even	year	nation	socialistic	HOWEVER realistically evaluate
1988: and, in, is, on	work	year	national	We SINCERELY welcome the results
1989: and, in, on, about	society	meeting	work	We UTILIZES what may be exercised
<b>Václav Havel (1936-2011)</b>				
1990: and, with, that, in	republic	country	nations	It WOULD be very unwise
1991: and, in, with, that	year	believe	system	DEAR fellows
1992: and, with, in, that	year	state	search	THIS is all good, of course,
<b>Prime minister Václav Klaus (1941)</b>				
1993: and, with, in, on	czech	republic	state	-
<b>Chair of parliament Milan Uhde (1936)</b>				
1993: and, with, in, that	home	state	life	-
<b>Václav Havel (1936-2011)</b>				
1994: and, with, in, on	polity	civil	society	than RESPECT for all nonstandard
1995: and, with, in, is	polity	politics	hope	even HOPE is work, may be difficult
1996: and, with, that, in	country	state	society	I THINK that contemporary developments
1997: and, to, with, is	sign	human	things	This is NOT the same time being
1998: and, with, to, in	policy	means	things	DEAR fellows
1999: and, with, in, even	wall	Europe	years	RATE our passion for vision
2000: and, with, even, that	life	world	peoples	I KNOW how unpopular it sounds
2001: and, in, is, with	interest	people	society	We ALL know yet
2002: and, with, about, that	polity	European	country	It would help us realize
2003: and, with, in, is	polity	year	cohabitation	DEAR fellows
<b>Václav Klaus (1941)</b>				
2004: and, with, in, even	year	country	czech	that THIS moment
2005: and, with, in, on	year	state	country	are LEARNING to distinguish
2006: and, in, with, that	European	life	year	after the rejection of the European Constitution
2007: and, with, in, that	year	European	nothing	MUST NOT be indifferent to us
2008: and, with, in, even	year	country	coming	I would like to ASSURE you
2009: and, in, with, to	year	Union	europoan	CERTAINLY this magic wand
2010: and, with, in, it	year	decrease	citizen	CERTAINLY better than
2011: and, with, in, it	year	government	voices	INSIDE society
2012: and, with, in, that	year	country	Europe	CLAIMS are strengthening
2013: and, in, is, am	day	country	year	INSIDE municipalities
<b>Miloš Zeman (1944)</b>				
2014: and, with, that, am	year	thank	life	I AM NOT TALKING about the division of society
2015: and, with, that, with	public	never	year	IF I HAVE TO evaluate government
2016: and, that, is, with	country	year	republic	SECOND good reals

Figure 10: The most frequent words and words with the highest energy. Source: own.

Figures are divided into some subsections. As mentioned before, the Czechoslovak Republic had two presidents between 1940 and 1945. In 1993 the second anomaly appeared. The Czechoslovak republic ceased to exist. Since 1993 the country was divided into two smaller autonomous countries: The Czech Republic and the Slovak Republic. So, the president Václav Havel had no speech in 1993. Prime ministers were speaking to their nations instead of president.

Rows are sorted by years. Each row has its color depending on president or prime minister. Colors were chosen according to all figures. The first column contains first four most frequent conjunctions and prepositions. Then there are three columns containing the most frequent words sorted by order. The last column shows the word with the highest energy. Those words are written by uppercase.

### 3.3 Number of words

Scatter chart will be used to demonstrate a vocabulary richness. Coordinates on axis x means total number of words in speech and coordinates on axis y means total number of different words. Functional dependency can be modeled by Gompertz curve. Presidents with values above the curve have greater ratio of words than other presidents. Language richness of speeches under the curve can be considered lesser. In article [9] author mentioned that thematic concentration of president Havel is surprisingly low. But this claim doesn't seem to correspond with language richness. According to the Figure 10, ratio of number of different words and total number of words is greater as for Havel. This could be caused by choosing different methods of evaluating language richness.

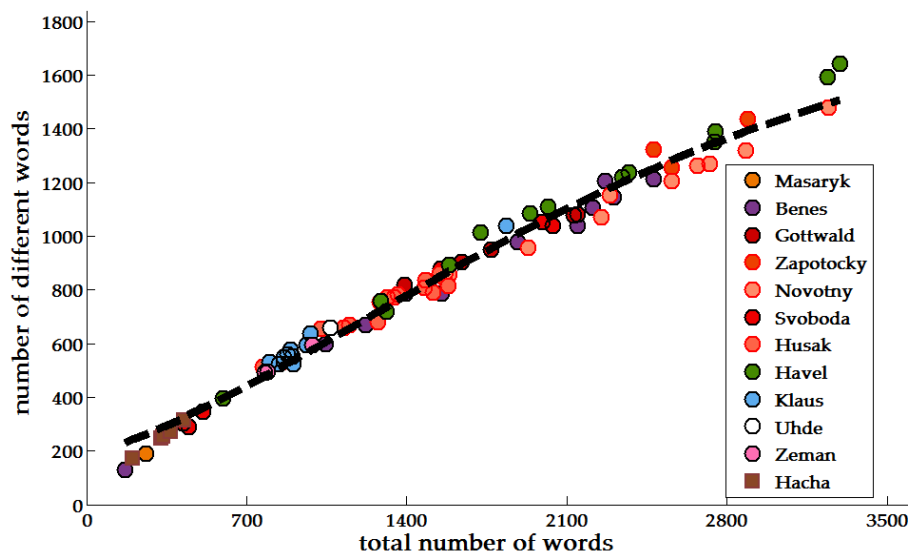


Figure 11: Number of different words. Source: own.

## 4 Conclusions

This article's goal is to present results of our research and show that data we already had can be processed in different way. The extraction of information is much discussed nowadays. Main purpose of research is finding the words with the greatest energy. Because they have historical importance, they can be used as keywords and they even characterize the speaker.

Scale of publication doesn't allow to detail comment and the description of used algorithms. Many hours of machine time have been needed during the calculations of phonetic parameters of speeches. Archive [13] contains 74 speeches. So, this is more than 19 hours of recordings to be analyzed. Before calculation of mean values of ZCR and Log energy, there was an extensive table for each speech. Presented parameters were created by reducing the table containing millions of values (each frame parameter values) into one mean value. Unreduced data may be used for further analysis.

Comparing the table of most frequent thematic words with table containing the words pronounced with the greatest energy brings almost no match. The speaker didn't emphasize the most frequent words. But he chose to highlight other words. For example, Masaryk talked about economy. Beneš emphasized the war and human kind. Novotný insisted on hard work and improving the communistic country. Havel emphasized the very first words: "Dear fellow citizens." It can provide some information without listening to the whole speech. It even characterizes the president himself and an era of each president (the most important events, standard of living, relationship between president and citizens).

## 5 References

1. LIANG, Bai, et al. Feature analysis and extraction for audio automatic classification. In: *Systems, Man and Cybernetics, 2005 IEEE International Conference on*. IEEE, 2005. pp. 767-772.
2. LIU, Zhu; WANG, Yao; CHEN, Tsuhan. Audio feature extraction and analysis for scene segmentation and classification. *The Journal of VLSI Signal Processing*, 1998, 20.1: 61-79.
3. Short Term Time Domain Processing of Speech (Theory): Speech Signal Processing Laboratory. *IIT GUWAHATI Virtual Lab* [online]. Guwahati: Bhattacharjee, 2016. URL: <http://iitg.vlab.co.in/?sub=59&brch=164&sim=857&cnt=1>. [accessed 2016-12-05]
4. KANDA, Naoyuki, et al. Open-vocabulary keyword detection from super-large scale speech database. In: *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*. IEEE, 2008. pp. 939-944.
5. NOUZA, Jan, et al. Very large vocabulary speech recognition system for automatic transcription of Czech broadcast programs. In: *INTERSPEECH*. 2004.
6. WASSON, D.; DONALDSON, R. Speech amplitude and zero crossings for automated identification of human speakers. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1975, 23.4: 390-392.
7. KUBÁT, M., ČECH, R. Quantitative Analysis of US Presidential Inaugural Addresses. *Glottometrics* (34) 2016, pp. 14-27.

8. TUZZI, Arjuna; POPESCU, Ioan-Iovitz; ALTMANN, Gabriel. *Quantitative analysis of Italian texts*. Lüdenscheid: RAM-Verlag, 2010.
9. ČECH, R. 2014. Language and ideology: Quantitative thematic analysis of New Year speeches given by Czechoslovak and Czech presidents (1949-2011). *Quality & Quantity*, 48, pp. 899–910.
10. ŠLAHORA, Jan. *Matematická lingvistika a překlady básně E.A. Poea Havran* [online]. Pardubice, 2015. Master thesis. University Pardubice, Faculty of electrical engineering and informatics. URL: [https://dk.upce.cz/bitstream/handle/10195/60403/SlahoraJ\\_MatematickaLingvistika\\_JM\\_2015.zip?sequence=1&isAllowed=y](https://dk.upce.cz/bitstream/handle/10195/60403/SlahoraJ_MatematickaLingvistika_JM_2015.zip?sequence=1&isAllowed=y). [accessed. 2016-12-05].
11. R. Bachu, S. Kopparthi, B. Adapa, and B. Barkana, “Separation of voiced and unvoiced using zero crossing rate and energy of the speech signal” in American Society for Engineering Education (ASEE) Zone Conference Proceedings, 2008, pp. 1–7.
12. MAREK, J. - ŠLAHORA, J. Měření podobnosti překladů básně Havran. *Forum Statisticum Slovacum*, 2013, vol. X (2013), no. 5, pp. 90-95. ISSN: 1336-7420
13. URL: [http://www.rozhlas.cz/zpravy/data/\\_zprava/od-tgm-k-zemanovi-poslechnete-si-vanocni-a-novorocni-projevy-vsech-prezidentu--1436738](http://www.rozhlas.cz/zpravy/data/_zprava/od-tgm-k-zemanovi-poslechnete-si-vanocni-a-novorocni-projevy-vsech-prezidentu--1436738). [accessed. 2016-12-17].