

Čeština pro 21. století

Jak s ní nakládají prezidenti?

Radek Čech

# Struktura přednášky

1. možnosti analýzy (nejen politických) projevů a textů
  - aneb čtème a počítejme...
2. novoroční a vánoční projevy československých a českých prezidentů
  - aneb klady a zápory jednoho specifického žánru...
3. kvantitativní charakteristiky prezidentských projevů
  - aneb co a proč počítat a jak to interpretovat...
4. QuitaUp
  - aneb jak jednoduše měřit vybrané charakteristiky textů...

# 1. Možnosti analýzy (nejen politických) projevů a textů

- aneb čtème a počítejme...

# Možnosti analýzy (nejen politických) projevů a textů

- jak na analýzu textů?

# Možnosti analýzy (nejen politických) projevů a textů

- jak na analýzu textů?
- přečíst si ho....

# Možnosti analýzy (nejen politických) projevů a textů

- jak na analýzu textů?
- přečíst si ho....
- **kvalitativní** analýza
  - obsahové i formální analýzy

# Možnosti analýzy (nejen politických) projevů a textů

- jak na analýzu textů?
- přečíst si ho....
- **kvalitativní** analýza
  - obsahové i formální analýzy
  - limity?

# Možnosti analýzy (nejen politických) projevů a textů

- jak na analýzu textů?
- přečíst si ho....
- **kvalitativní** analýza
  - obsahové i formální analýzy
  - limity?





# Možnosti analýzy (nejen politických) projevů a textů

- jak na analýzu textů?
- využít početní nástroje....

# Možnosti analýzy (nejen politických) projevů a textů

- jak na analýzu textů?
- využít početní nástroje....
- **kvantitativní** analýza
  - obsahové i formální analýzy

```
95
96 #vypocet RR
97 N = len(tokens)
98 RR = suma_fr2 / N ** 2
99 print(RR)
100
101 #a ted vse dohromady na jednom miste
102 frekvence = list(frekv_slovnik.values())
103 umocnene_frekvence = []
104 for hodnota in frekvence:
105     umocnene_frekvence.append(hodnota ** 2)
106 suma_fr2 = sum(umocnene_frekvence)
107 N = len(tokens)
108 RR = suma_fr2 / N ** 2
109 print(round(RR, 4))
110
111
112 #pro snadnejsi interpretaci se RR relativizuje zde
```

# Možnosti analýzy (nejen politických) projevů a textů

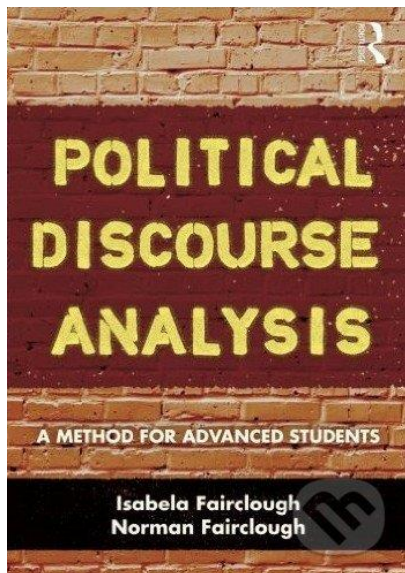
- jak na analýzu textů?
- využít početní nástroje....
- **kvantitativní** analýza
  - obsahové i formální analýzy
  - limity?

```
95
96 #vypocet RR
97 N = len(tokens)
98 RR = suma_fr2 / N ** 2
99 print(RR)
100
101 #a ted vse dohromady na jednom miste
102 frekvence = list(frekv_slovnik.values())
103 umocnene_frekvence = []
104 for hodnota in frekvence:
105     umocnene_frekvence.append(hodnota ** 2)
106 suma_fr2 = sum(umocnene_frekvence)
107 N = len(tokens)
108 RR = suma_fr2 / N ** 2
109 print(round(RR, 4))
110
111
112 #pro snadnejsi interpretaci se RR relativizuje zde
```

# Možnosti analýzy (nejen politických) projevů a textů

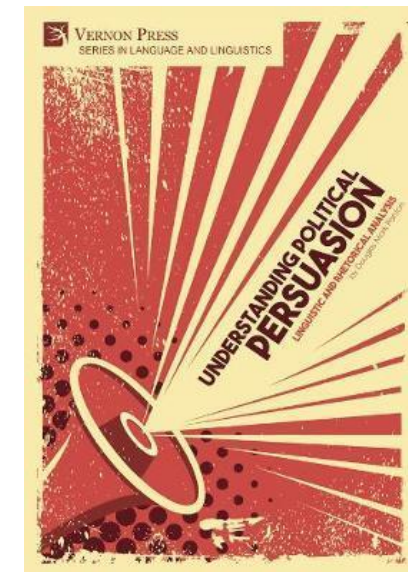
kvalitativní & kvantitativní

# Kvalitativní analýzy



Fairclough, I., & Fairclough, N. (2013). *Political discourse analysis*. Routledge.

Svobodová, J. (2016). *Manipulace a argumentace v politickém a mediálním diskurzu*. Univerzita Palackého v Olomouci.



Ponton, D. M. (2020). *Understanding Political Persuasion: Linguistic and Rhetorical Analysis*. Vernon Press.

# Kvalitativní analýzy - ukázky

- Hrušková B., *Analýza vybraných vánočních a novoročních projevů československých a českých prezidentů*. Olomouc 2018.

- [https://theses.cz/id/aj0yoh/hruskova\\_analyza\\_vanocnich\\_a\\_novorocnich\\_projevu.pdf](https://theses.cz/id/aj0yoh/hruskova_analyza_vanocnich_a_novorocnich_projevu.pdf)

# Kvalitativní analýzy - ukázky

- Gottwlad 1949

## „3.4.2.1 Analýza textu

Prezident v tomto projevu hovořil k lidu jako „rodič k dítěti“, jak je patrné z častého použití modálních sloves, kterými většinou upozorňoval na nutnost zlepšení pracovního procesu (*budou muset naši zemědělci dohánět, musíme si uvědomit, i když s ním nemůžeme být ještě zdaleka spokojeni*). Dále se objevuje značné množství hodnotících slov (*z významných opatření, se plně osvědčil, s velkými obavami*), a právě hodnocení je typické pro transakci rodič–dítě. Ani argumentace, kterou využíval, neodpovídá dospělému, není totiž věcná. Uváděl výroky, z nichž nelze vyvodit ověřitelné závěry, neboť jsou příliš obecné. Problémy detailně nevysvětloval a argumentaci stavěl na kritice protistrany:

*Odstraněním všech škůdců z našeho hospodářství co nejrychleji dosáhneme lepší budoucnosti.*

# Kvalitativní analýzy - ukázky

- Gottwlad 1949

„Z celého textu je nejzřetelnější zdůrazňování a časté použití přivlastňovacího zájmena „náš“ (*našim národům, naše zemědělství, našeho lidově demokratického zřízení*), které je rysem blízkosti.“

„Ve velkém množství se objevuje i nutnost (*jdeme po správné cestě, nutně pro zdárný postup, jedině svornou spoluprací*), na kterou navazuje správnost (*velké úspěchy (...) potvrzují, zůstanou mezníkem v dějinách, historickým únorovým vítězstvím*)



# Kvalitativní analýzy - ukázky

- Svoboda 1969

„Způsob, jímž Svoboda k lidu hovořil, odpovídá transakci rodič–dítě. Snažil se navodit pocit důvěry a bezpečí. Cítil na emoce a vyvolával dojem, že ví, co je pro recipienty dobré. Spíše radil a hodnotil, věcně problém neanalyzoval. Zmiňoval problémy, nedokázal však ani přesně pojmenovat jejich příčiny (*příčiny našich obtíží jsou především ve vážných chybách minulých let*), ani uvést konkrétní pokyny, jak je vyřešit (*záleží na odhodlání a poctivé práci*).“

# Kvalitativní analýzy - ukázky

- Svoboda 1969

- <https://interaktivni.rozhlas.cz/prezidentske-projevy/www/>

- Svoboda 1971

- <https://interaktivni.rozhlas.cz/prezidentske-projevy/www/#1971-svoboda>

# Kvalitativní analýzy - ukázky

- Svoboda 1969

„Patrná je snaha vyvolat iluzi dialogu, který samozřejmě v monologickém projevu není možný. Sám svou řeč pojmenoval jako rozhovor, navozoval tedy pocit, že lidé stojí přímo před ním a mají možnost reakce, a zodpovídal nevyřčené otázky, předjímal možné protesty a výtky, které by mohly posluchače napadnout.

*Hned na počátku našeho rozhovoru vám chci říci, (...)*

*Hovořím k vám v prvních hodinách roku nového a je mi, jako byste vy všichni, kdo mě posloucháte, byli přede mnou. Ve vašich očích vidím plno zvědavosti, zájmu i nadějí. V některých i obavy. “*

# Kvantitativní analýza politických projevů

- H. D. Lasswell, N. Leites:  
*Language of Politics*, New York 1949

## **LANGUAGE OF POLITICS**

---

*studies in quantitative semantics*

by **Harold D. Lasswell,**  
**Nathan Leites**  
and **Associates**

*Raymond Fadner*  
*Joseph M. Goldsen*  
*Alan Grey*  
*Irving L. Janis*  
*Abraham Kaplan*  
*David Kaplan*  
*Alexander Mintz*  
*I. De Sola Pool*  
*Sergius Yakobson*

*George W. Stewart, Publisher, Inc., New York*



# Kvantitativní analýza politických projevů

- H. D. Lasswell, N. Leites:  
*Language of Politics*, New York 1949

## CONTENTS

CHAPTER	I. INTRODUCTION	PAGE
I.	THE LANGUAGE OF POWER— <i>Harold D. Lasswell</i> .	3
II.	STYLE IN THE LANGUAGE OF POLITICS— <i>Harold D. Lasswell</i> . . . . .	20
III.	WHY BE QUANTITATIVE?— <i>Harold D. Lasswell</i> .	40

## LANGUAGE OF POLITICS

*studies in quantitative semantics*

by Harold D. Lasswell,  
Nathan Leites  
and Associates

*Raymond Fadner*  
*Joseph M. Goldsen*  
*Alan Grey*  
*Irving L. Janis*  
*Abraham Kaplan*  
*David Kaplan*  
*Alexander Mintz*  
*I. De Sola Pool*  
*Sergius Yakobson*

*George W. Stewart, Publisher, Inc., New York*



# Kvantitativní analýza politických projevů

- H. D. Lasswell, N. Leites:  
*Language of Politics*, New York 1949

## CONTENTS

CHAPTER		PAGE
	I. INTRODUCTION	
	I. THE LANGUAGE OF POWER— <i>Harold D. Lasswell</i> .	3
	II. STYLE IN THE LANGUAGE OF POLITICS— <i>Harold D. Lasswell</i> . . . . .	20
	III. WHY BE QUANTITATIVE?— <i>Harold D. Lasswell</i> .	40

## LANGUAGE OF POLITICS

*studies in quantitative semantics*

by **Harold D. Lasswell,**  
**Nathan Leites**  
and **Associates**

*Raymond Fadner*  
*Joseph M. Goldsen*  
*Alan Grey*  
*Irving L. Janis*  
*Abraham Kaplan*  
*David Kaplan*  
*Alexander Mintz*  
*I. De Sola Pool*  
*Sergius Yakobson*

*George W. Stewart, Publisher, Inc., New York*



# H. D. Lasswell: Why Be Quantitative?

- **pozornost**

- Can we assume that a scholar read his sources with the same degree of care throughout his research?

# H. D. Lasswell: Why Be Quantitative?

- pozornost
- **jasně vymezená vlastnost vzorku**
  - Did he allow his eye to travel over the thousands upon thousands of pages of parliamentary debates, newspapers, magazines and other sources listed in his bibliography or notes?
  - Was the sampling system for the *Frankfurter Zeitung*, if one was employed, comparable with the one for the *Manchester Guardian*?

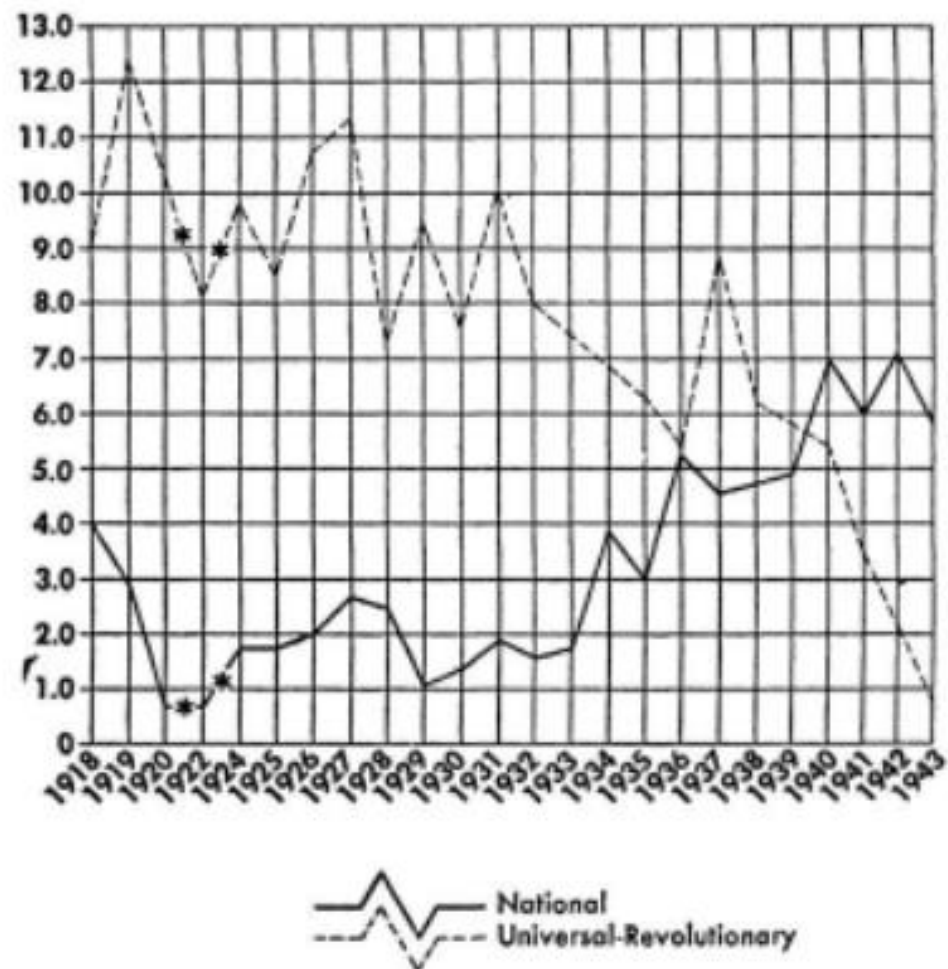


# H. D. Lasswell: Why Be Quantitative?

- pozornost
- jasně vymezená vlastnost vzorku
- **jasně vymezený (kvantifikovaný) podklad pro interpretaci**
  - evidence-based interpretation

**'NATIONAL' AND 'UNIVERSAL-REVOLUTIONARY'  
SYMBOLS IN MAY-DAY SLOGANS OF  
THE COMMUNIST PARTY (S.U.)**

Ratio of  
frequency to total  
Word-count



\*No slogans issued in 1921 and 1923

# Vývoj kvantitativních analýz – faktory

# Vývoj kvantitativních analýz – faktory

- **technické aspekty**
  - výpočetní technika
  - dostupnost dat

# Vývoj kvantitativních analýz – faktory

- technické aspekty
  - výpočetní technika
  - dostupnost dat
- **tradice humanitního vzdělávání**

# Vývoj kvantitativních analýz – faktory

- technické aspekty
  - výpočetní technika
  - dostupnost dat
- tradice humanitního vzdělávání
- **proměny lingvistiky**
  - empirizace
  - metodologie

# Kvantitativní analýza

- **klady**

- replikovatelnost
- porovnatelnost
  - možnost aplikace na jiné vzorky
- „robustnost“
  - interpretace
- intersubjektivita

# Kvantitativní analýza

- klady

- replikovatelnost
- porovnatelnost
  - možnost aplikace na jiné vzorky
- „robustnost“
  - interpretace
- intersubjektivita

- **zápory**

- redukcionismus
- metodologické limity



# Kvantitativní analýza politických projevů

- obsahová/tematická analýza

# Kvantitativní analýza politických projevů

- obsahová/tematická analýza
- analýza (na první pohled) „skrytých“ vlastností
  - míra zaměřenosti na hlavní témata
  - aktivita/deskriptivita
  - syntaktická komplexita

# Kvantitativní analýza politických projevů

- ukázky

# Kvantitativní analýza politických projevů

- ukázky
- jazykový materiál
  - novoroční a vánoční projevy československých a českých prezidentů
  - 1935-dosud
  - texty dnes dostupné v *Korpusu prezidentských projevů Speeches* (ČNK)
    - <https://wiki.korpus.cz/doku.php/cnk:speeches>
  - ukázka: zjištění relativní frekvence modálních sloves

# Kvalitativní analýzy - ukázky

- Gottwlad 1949

## „3.4.2.1 Analýza textu

Prezident v tomto projevu hovořil k lidu jako „rodič k dítěti“, jak je patrné z častého použití modálních sloves, kterými většinou upozorňoval na nutnost zlepšení pracovního procesu (*budou muset naši zemědělci dohánět, musíme si uvědomit, i když s ním nemůžeme být ještě zdaleka spokojeni*). Dále se objevuje značné množství hodnotících slov (*z významných opatření, se plně osvědčil, s velkými obavami*), a právě hodnocení je typické pro transakci rodič–dítě. Ani argumentace, kterou využíval, neodpovídá dospělému, není totiž věcná. Uváděl výroky, z nichž nelze vyvodit ověřitelné závěry, neboť jsou příliš obecné. Problémy detailně nevysvětloval a argumentaci stavěl na kritice protistrany:

*Odstraněním všech škůdců z našeho hospodářství co nejrychleji dosáhneme lepší budoucnosti.*

# Frekvenční analýza slov (lemmat)

pořadí	Gottwald (1952)	Zápotocký (1954)	Havel (1999)	Klaus (2006)
1.	rok	rok	zed'	život
2.	americký	výroba	dnes	rok
3.	nový	hospodářství	různý	volba
4.	průmysl	práce	lidský	politika
5.	výroba	zemědělský	nový	země
6.	průmyslový	nutný	vlastní	evropský
7.	hodně	národní	právo	občan
8.	sovětský	lid	rok	přát
9.	válečný	plán	dobrý	velký
10.	potravina	průmysl	občanský	člověk

# Frekvenční analýza slov (lemmat)

pořadí	Gottwald (1952)	Zápotocký (1954)	Havel (1999)	Klaus (2006)
1.	rok	rok	zed'	život
2.	americký	výroba	dnes	rok
3.	nový	hospodářství	různý	volba
4.	průmysl	práce	lidský	politika
5.	výroba	zemědělský	nový	země
6.	průmyslový	nutný	vlastní	evropský
7.	hodně	národní	právo	občan
8.	sovětský	lid	rok	přát
9.	válečný	plán	dobry	velký
10.	potravina	průmysl	občanský	člověk

# Frekvenční analýza slov (lemmat)

pořadí	Gottwald (1952)	Zápotocký (1954)	Havel (1999)	Klaus (2006)
1.	rok	rok	zed'	život
2.	americký	výroba	dnes	rok
3.	nový	hospodářství	různý	volba
4.	průmysl	práce	lidský	politika
5.	výroba	zemědělský	nový	země
6.	průmyslový	nutný	vlastní	evropský
7.	hodně	národní	právo	občan
8.	sovětský	lid	rok	přát
9.	válečný	plán	dobry	velký
10.	potravina	průmysl	občanský	člověk



# Frekvenční analýza slov (lemmat)

pořadí	Gottwald (1952)	Zápotocký (1954)	Havel (1999)	Klaus (2006)
1.	rok	rok	zed'	život
2.	americký	výroba	dnes	rok
3.	nový	hospodářství	různý	volba
4.	průmysl	práce	lidský	politika
5.	výroba	zemědělský	nový	země
6.	průmyslový	nutný	vlastní	evropský
7.	hodně	národní	právo	občan
8.	sovětský	lid	rok	přát
9.	válečný	plán	dobrý	velký
10.	potravina	průmysl	občanský	člověk

# Frekvenční analýza slov (lemmat)

pořadí	Gottwald (1952)	frekvence	relativní frekvence (%)	Zápotocký (1954)	frekvence	relativní frekvence (%)
1.	rok	40	2,18	rok	25	1,36
2.	americký	18	0,98	výroba	19	1,04
3.	nový	18	0,98	hospodářství	15	0,82
4.	průmysl	12	0,65	práce	15	0,82
5.	výroba	12	0,65	zemědělský	15	0,82
6.	průmyslový	11	0,60	nutný	11	0,60
7.	hodně	10	0,55	národní	11	0,60
8.	sovětský	8	0,44	lid	10	0,55
9.	válečný	8	0,44	plán	10	0,55
10	potravina	7	0,33	průmysl	10	0,55

# Frekvenční analýza slov (lemmat)

- Gottwald = 1,9 %
- Zápotocký = 4,6 %

pořadí	Gottwald (1952)	frekvence	relativní frekvence (%)	Zápotocký (1954)	frekvence	relativní frekvence (%)
1.	rok	40	2,18	rok	25	1,36
2.	americký	18	0,98	výroba	19	1,04
3.	nový	18	0,98	hospodářství	15	0,82
4.	průmysl	12	0,65	práce	15	0,82
5.	výroba	12	0,65	zemědělský	15	0,82
6.	průmyslový	11	0,60	nutný	11	0,60
7.	hodně	10	0,55	národní	11	0,60
8.	sovětský	8	0,44	lid	10	0,55
9.	válečný	8	0,44	plán	10	0,55
10.	potravina	7	0,33	průmysl	10	0,55

# Analýza klíčových slov (lemmat)

- klíčové slovo
- slovo, které se ve daném textu objeví významně častěji než v referenčním korpusu

# Analýza klíčových slov (lemmat)

- klíčové slovo
- slovo, které se ve daném textu objeví významně častěji než v referenčním korpusu
- vyhodnocení → skóre, statistické testy
  - např. log-likelihood ( $LL$ )

$$LL = 2 \left( f_{slovo\_text} \cdot \log \frac{f_{slovo\_text}}{f(o)_{slovo\_text}} + f_{slovo\_korpus} \cdot \log \frac{f_{slovo\_korpus}}{f(o)_{slovo\_korpus}} \right)$$

# Analýza klíčových slov (lemmat)

min. frekvence slova v textu = 3

nejfrekventovanější slova		klíčová slova			
Klaus (2006)	f	Klaus (2006)	f	$f_{\text{SYN2010}}$	log likelihood
život	8	volba	7	23 529	36,41
rok	7	politika	6	18 866	31,99
volba	7	spoluobčan	3	717	31,29
politika	6	občan	5	14 679	27,33
země	6	přát	5	16 608	26,13
evropský	5	vážený	3	2 373	24,15
občan	5	život	8	92 237	23,00
přát	5	evropský	5	34 290	19,17
velký	5	volit	3	5 757	18,89
člověk	5	odpovědnost	3	6 066	18,59

# Analýza klíčových slov (lemmat)

min. frekvence slova v textu = 3

nejfrekventovanější slova		klíčová slova			
Klaus (2006)	f	Klaus (2006)	f	f <sub>SYN2010</sub>	log likelihood
život	8	volba	7	23 529	36,41
rok	7	politika	6	18 866	31,99
volba	7	spoluobčan	3	717	31,29
politika	6	občan	5	14 679	27,33
země	6	přát	5	16 608	26,13
evropský	5	vážený	3	2 373	24,15
občan	5	život	8	92 237	23,00
přát	5	evropský	5	34 290	19,17
velký	5	volit	3	5 757	18,89
člověk	5	odpovědnost	3	6 066	18,59

# Analýza klíčových slov (lemmat)

min. frekvence slova v textu = 3

nejfrekventovanější slova		klíčová slova			
Klaus (2006)	f	Klaus (2006)	f	$f_{\text{SYN2010}}$	log likelihood
život	8	volba	7	23 529	36,41
rok	7	politika	6	18 866	31,99
volba	7	spoluobčan	3	717	31,29
politika	6	občan	5	14 679	27,33
země	6	přát	5	16 608	26,13
evropský	5	vážený	3	2 373	24,15
občan	5	život	8	92 237	23,00
přát	5	evropský	5	34 290	19,17
velký	5	volit	3	5 757	18,89
člověk	5	odpovědnost	3	6 066	18,59



# Analýza klíčových slov (lemmat)

min. frekvence slova v textu = 3

nejfrekventovanější slova		klíčová slova			
Klaus (2006)	f	Klaus (2006)	f	f <sub>SYN2010</sub>	log likelihood
život	8	volba	7	23 529	36,41
rok	7	politika	6	18 866	31,99
volba	7	spoluobčan	3	717	31,29
politika	6	občan	5	14 679	27,33
země	6	přát	5	16 608	26,13
evropský	5	vážený	3	2 373	24,15
občan	5	život	8	92 237	23,00
přát	5	evropský	5	34 290	19,17
velký	5	volit	3	5 757	18,89
člověk	5	odpovědnost	3	6 066	18,59

# Analýza klíčových slov (lemmat)

- KWords
  - <https://kwords.korpus.cz/>

Analýza (na první pohled) „skrytých“ vlastností

# Tematická koncentrace textu

# Předpoklady měření tematické koncentrace (TK)

- v různých textech se autor na dané téma či témata může zaměřovat s různou intenzitou

# Předpoklady měření tematické koncentrace (TK)

- v různých textech se autor na dané téma či témata může zaměřovat s různou intenzitou
- lze identifikovat jazykové jednotky, které je možné chápat jako nositele určitého tématu či témat

# Předpoklady měření tematické koncentrace (TK)

- v různých textech se autor na dané téma či témata může zaměřovat s různou intenzitou
- lze identifikovat jazykové jednotky, které je možné chápat jako nositele určitého tématu či témat
- míru zaměření se na dané téma či témata lze detekovat analýzou frekvenčních charakteristik textu

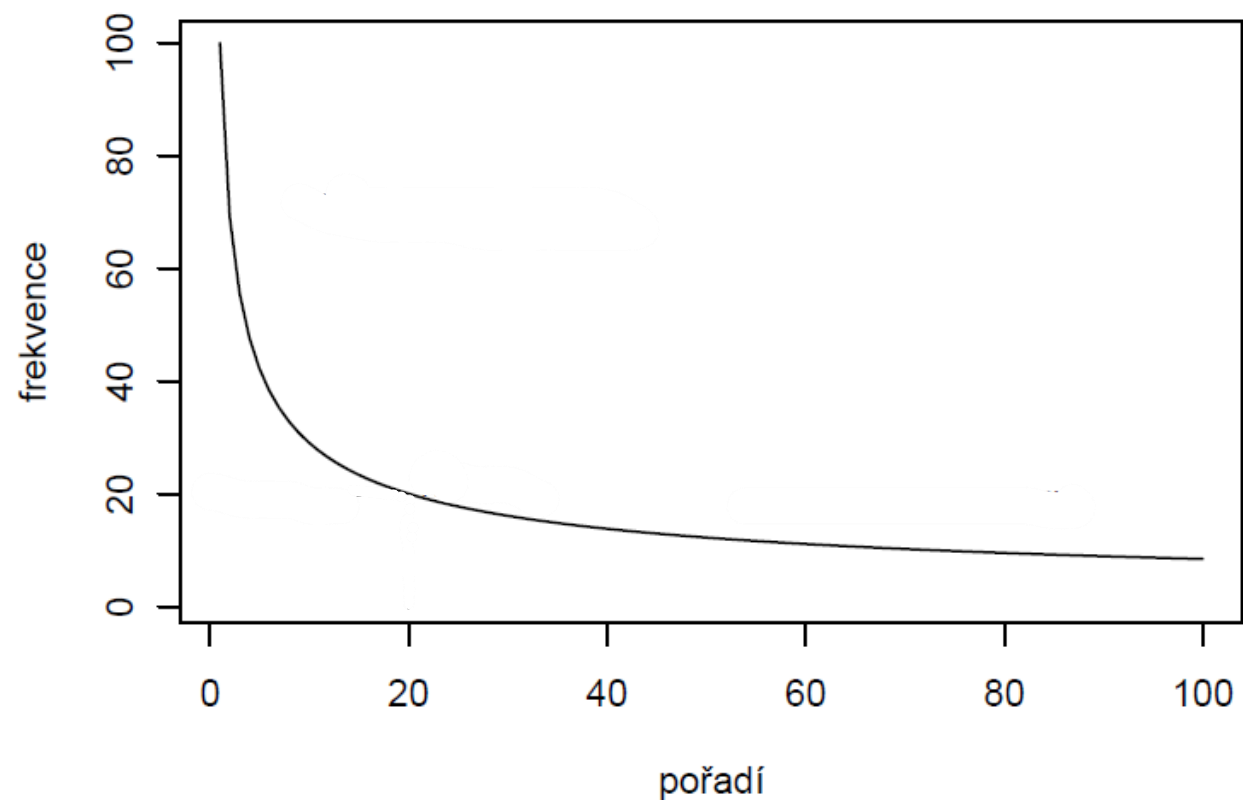
# Předpoklady měření tematické koncentrace (TK)

- v různých textech se autor na dané téma či témata může zaměřovat s různou intenzitou
- lze identifikovat jazykové jednotky, které je možné chápat jako nositele určitého tématu či témat
- míru zaměření se na dané téma či témata lze detekovat analýzou frekvenčních charakteristik textu
- míra zaměření se na dané téma či témata není náhodná, tj. předpokládá se její systematické chování vzhledem jak k jiným vlastnostem textu, tak k faktorům pragmatickým



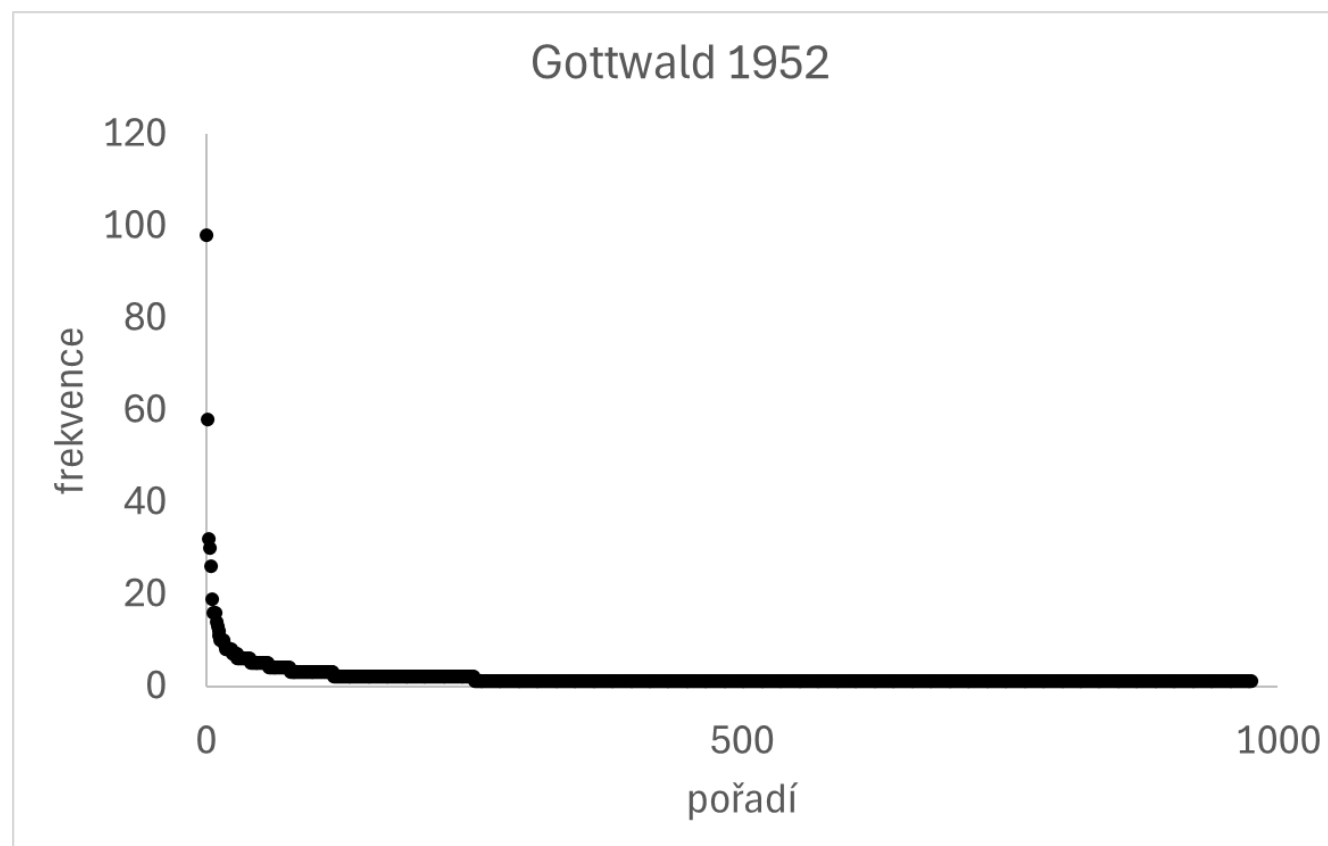
# Frekvenční struktura textu

- uspořádání slov podle frekvence



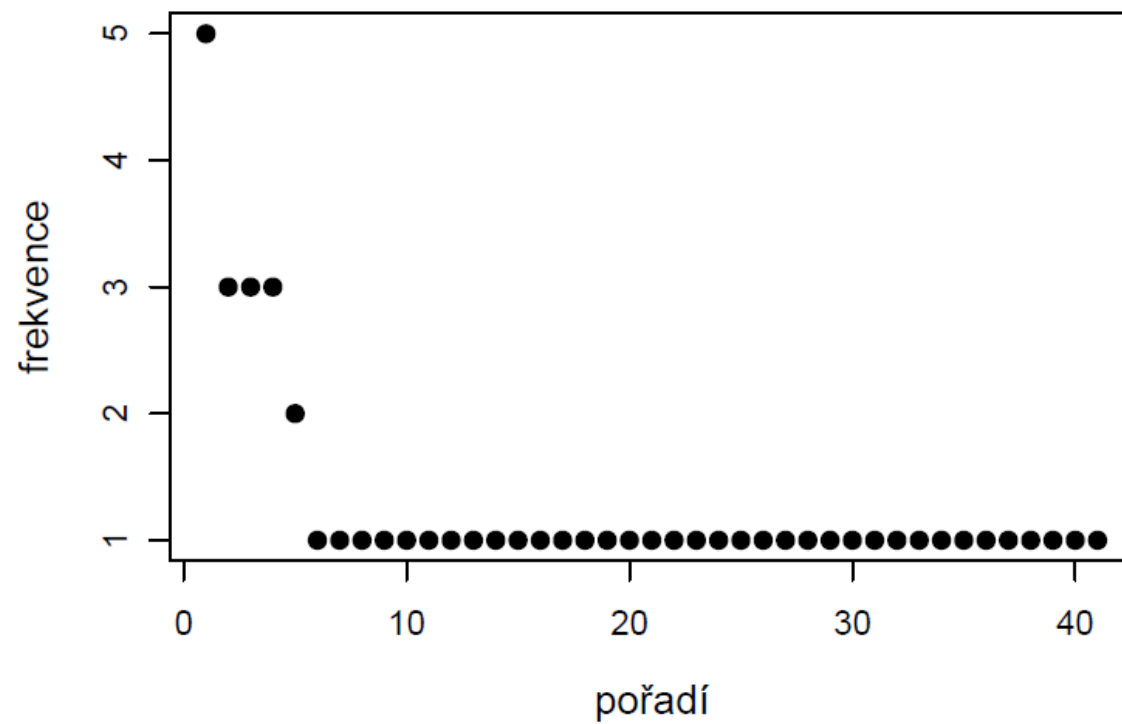
# Tematická koncentrace textu

- frekvenční struktura textu

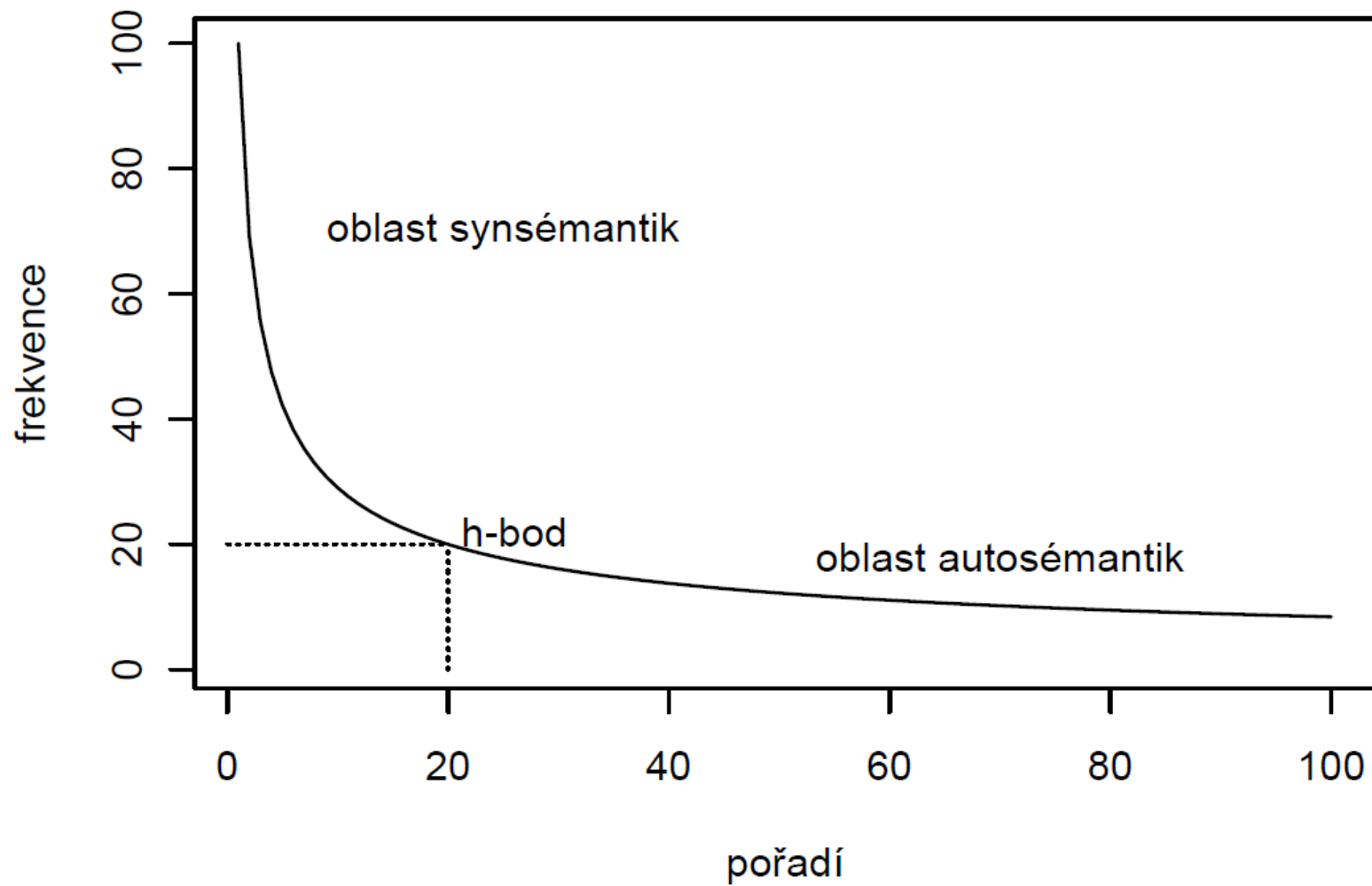


# Tematická koncentrace textu

- J. Skácel: Odvaha k tomu



# Frekvenční struktura textu



# Frekvenční struktura textu

- J. Škvorecký: Neurčité kontury

pořadí	slovo	frekvence
1	a	388
2	jsem	310
3	se	198
4	na	179
5	v	126
6	to	101
7	s	75
8	do	66
9	už	62
10	mi	61

# Frekvenční struktura textu

- J. Skácel: Odvaha k tomu

pořadí	slovní tvar	frekvence
1	<i>a</i>	5
2	<i>tam</i>	3
3	<i>byl</i>	3
4	<i>v</i>	3
5	<i>jsem</i>	2
6	<i>tím</i>	1
7	<i>prázdný</i>	1
8	<i>dílem</i>	1
9	<i>rovným</i>	1
(...)	(...)	(...)
41	<i>zděšený</i>	1

# Frekvenční struktura textu

- J. Skácel: Smuténka

*Smuténka*

*To až se v září stmívá,  
už bez sametu, drsně naholo,  
po poli chodí smuténka  
a zpívá,  
smuténka chodí kolem hrud  
šedých jak skřivani a zpívá,  
(je příběh starší nežli já,  
než moje smrt,  
než smutek ze mne, odpusť)  
zpívá si na poli smuténka  
a chodí  
po konopných cestách podzimu.*

# Frekvenční struktura textu

- J. Skácel: Smuténka

## *Smuténka*

*To až se v září stmívá,  
už bez sametu, drsně naholo,  
po poli chodí smuténka  
a zpívá,  
smuténka chodí kolem hrud  
šedých jak skřivani a zpívá,  
(je příběh starší nežli já,  
než moje smrt,  
než smutek ze mne, odpusť)  
zpívá si na poli smuténka  
a chodí  
po konopných cestách podzimu.*

pořadí	slovní tvar	frekvence
1	<i>smuténka</i>	4
2	<i>a</i>	3
3	<i>chodí</i>	3
4	<i>zpívá</i>	3
5	<i>poli</i>	2
6	<i>po</i>	2
7	<i>než</i>	2
8	<i>starší</i>	1
9	<i>příběh</i>	1
(...)	(...)	(...)
39	<i>září</i>	1

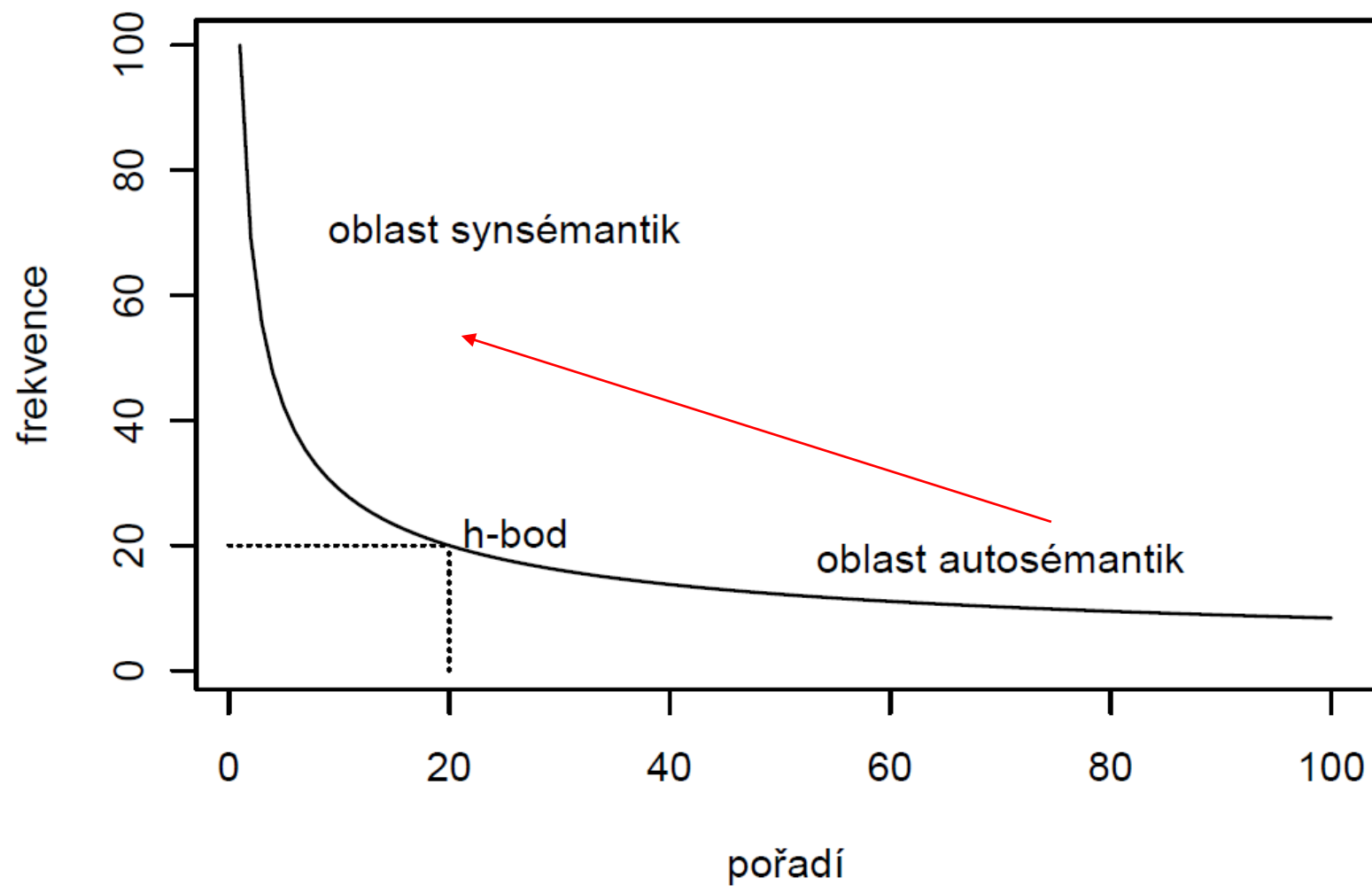


# Frekvenční struktura textu

- ČTK: V Beskydech blesk zapálil chatu, vítr lámal stromy

pořadí	slovní tvar	frekvence
1	<i>v</i>	18
2	<i>a</i>	15
3	<i>na</i>	13
4	<i>hasiči</i>	10
5	<i>voda</i>	7
6	<i>do</i>	6
7	<i>i</i>	6
8	<i>zaplavila</i>	5
9	<i>ze</i>	4
(...)	(...)	(...)
340	<i>zdravotníkům</i>	1

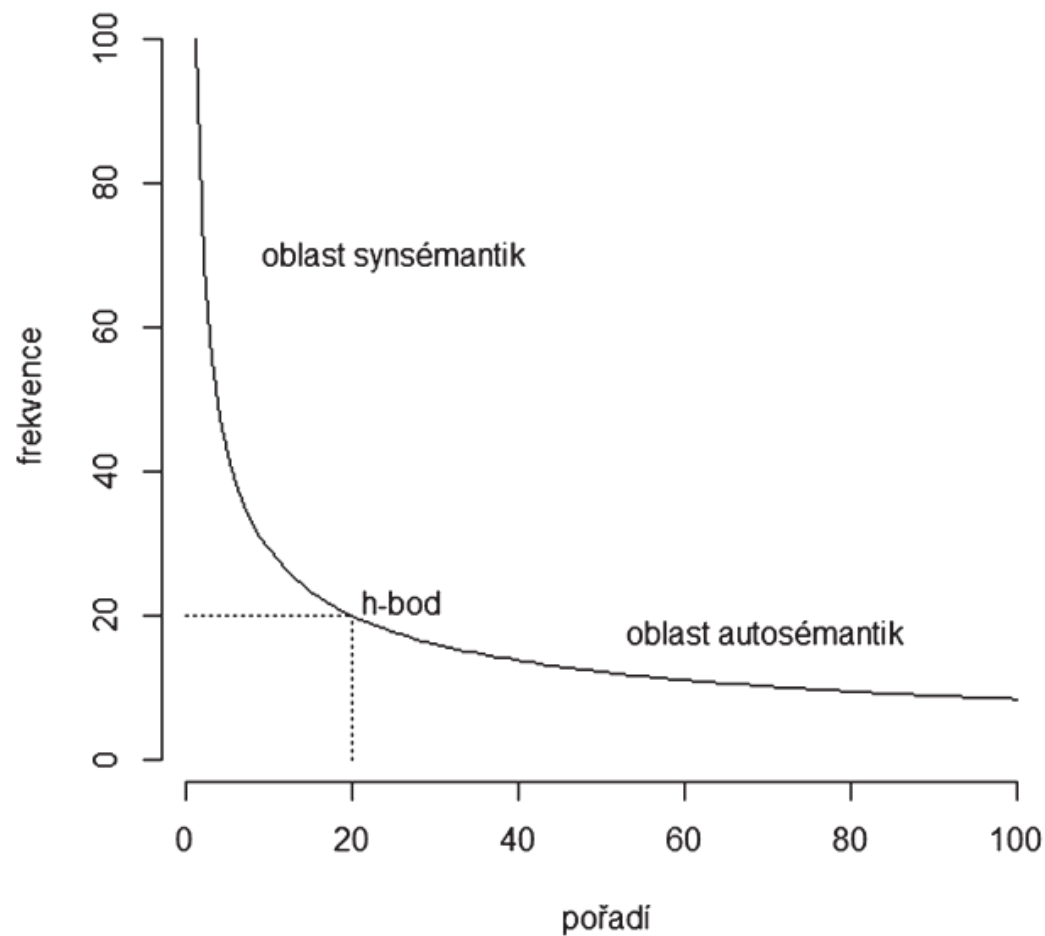
# Frekvenční struktura textu



# Tematická koncentrace textu

- tematická váha slova

$$TV_{\text{slovo}} = 2 \frac{(h - r')f(r')}{h(h - 1)f(1)}$$



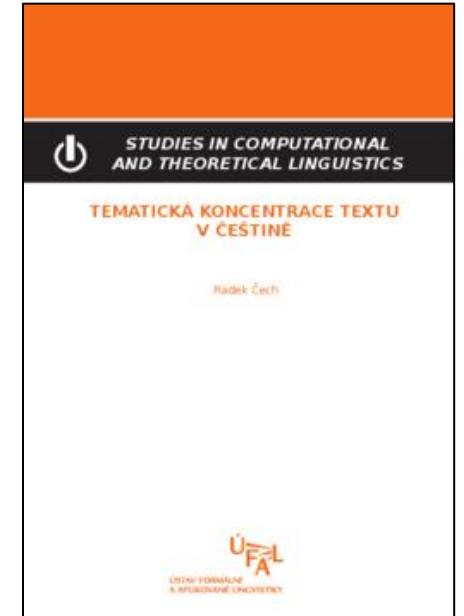
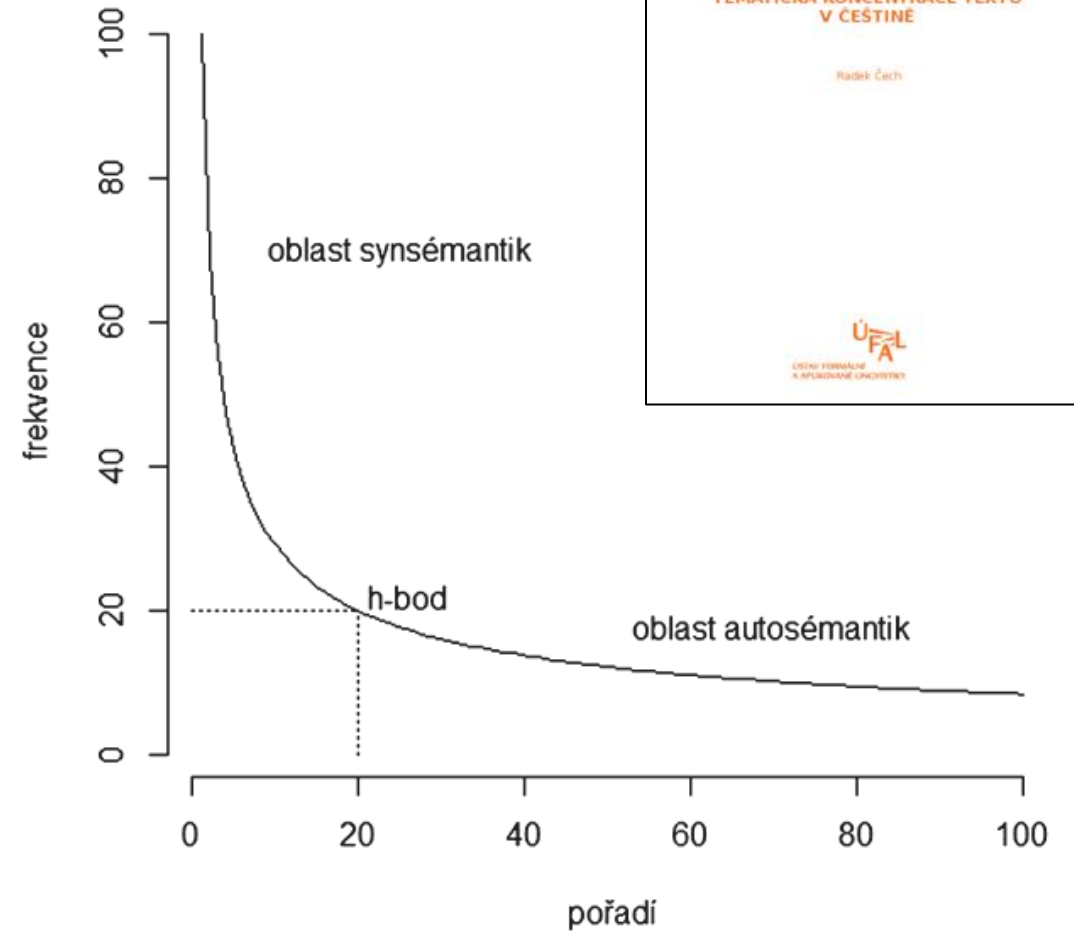
# Tematická koncentrace textu

- tematická váha slova

$$TV_{\text{slovo}} = 2 \frac{(h - r')f(r')}{h(h - 1)f(1)}$$

- tematická koncentrace textu

$$TK_{\text{text}} = \sum TV_{\text{slovo}} = \sum_{j=1}^T 2 \frac{(h - r'_{(j)})f(r'_{(j)})}{h(h - 1)f(1)}$$



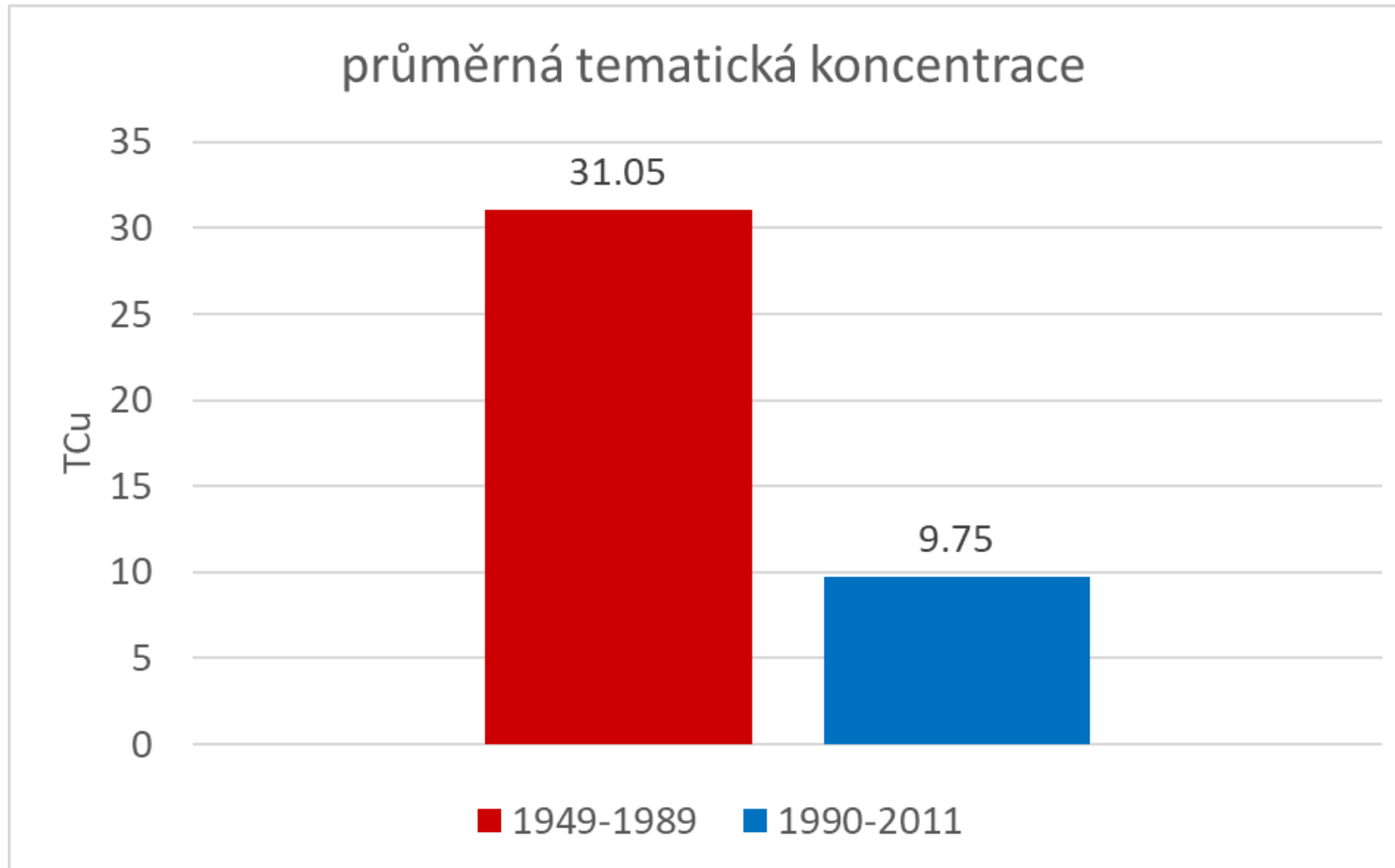
# Tematická koncentrace textu

- „We hypothesize (a) that the levels of **thematic concentration** in the texts of **totalitarian presidents** will be (significantly) **higher** than the levels of the democratic presidents **due to the influence of totalitarian ideology**“

Čech, R. (2014). Language and ideology: Quantitative thematic analysis of New Year speeches given by Czechoslovak and Czech presidents (1949-2011). *Quality & Quantity*, 48(2), 899-910.



# Tematická koncentrace textu



## 2. Novoroční a vánoční projevy československých a českých prezidentů

- aneb klady a zápory jednoho specifického žánru...

# Novoroční a vánoční projevy československých a českých prezidentů

- 1935–dosud

---

Tomáš Garrigue Masaryk	1918–35
Edvard Beneš	1935–38, 1940–48
Emil Hácha	1939–45
Klement Gottwald	1948–53
Antonín Zápotocký	1953–57
Antonín Novotný	1957–68
Ludvík Svoboda	1968–75
Gustáv Husák	1975–89
Václav Havel	1989–2003
Václav Klaus	2003–13
Miloš Zeman	2013ff

---



# Novoroční a vánoční projevy československých a českých prezidentů

- specifický žánr
  - slavností charakter
  - shrnutí událostí předchozího roku
  - výhled do budoucna
- homogenní žánr
  - význam pro kvantitativní analýzy
  - srov. vztah délky slova a žánru

# Novoroční a vánoční projevy československých a českých prezidentů

- autorství
  - mnohdy nejasné
    - tajemníci, úpravy
  - Havel, Klaus
    - nepřímé potvrzení vlastního autorství
  - Husák
    - Slovák, ale projevy česky
  - Svoboda
    - 1974 – mozkové příhody
- autorství jako projev politické odpovědnosti

# Novoroční a vánoční projevy československých a českých prezidentů

- autorství
  - mnohdy nejasné
    - tajemníci, úpravy
  - Havel, Klaus
    - nepřímé potvrzení vlastního autorství
  - Husák
    - Slovák, ale projevy česky
  - Svoboda
    - 1974 – mozkové příhody
- autorství jako **projev politické odpovědnosti**

# Novoroční a vánoční projevy československých a českých prezidentů

- <http://interaktivni.rozhlas.cz.s3-website.eu-central-1.amazonaws.com/prezidentske-projevy-2017/www/#1935-masaryk>

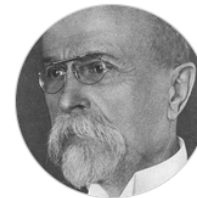
Zprávy



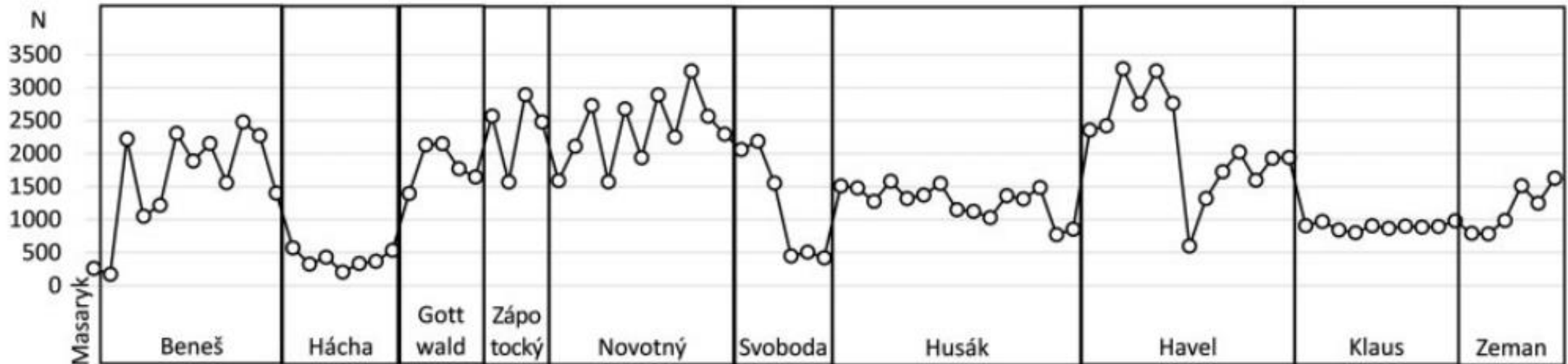
## 1935: Tomáš G. Masaryk

„Nerozčilujme se otázkami hospodářskými a politickými –  
problém dneška není jen hospodářský a politický, nýbrž  
především mravní“

Naše děti se již připravují na svátky vánoční, i vzpomínám svého  
dětství, když jsem se spolužáky v naší vesnici koledoval od domu  
k domu: Sláva na výsostech Bohu a na zemi pokoj lidem dobré vůle.



# Novoroční a vánoční projevy československých a českých prezidentů



**Fig. 1** Chronologically ordered text lengths ( $N$ ) of presidential annual addresses

# 3. kvantitativní charakteristiky prezidentských projevů

- aneb *co a proč počítat a jak to interpretovat...*

# Kvantitativní analýzy novoroční projevů

- tematická slova
- slovní bohatství
- průměrná délka slova (tokenu)
- aktivita textu
- vzdálenost mezi slovesy
- proporce nejfrekventovanějších slov
  
- Kubát, M., Mačutek, J., Čech, R. (2021). Communists spoke differently. An analysis of Czechoslovak and Czech annual presidential speeches. *Digital Scholarship in the Humanities*, 36, 138-152.
  - [Kubát, M., Mačutek, J., Čech, R. \(2021\). Communists spoke differently. An analysis of Czechoslovak and Czech annual presidential speeches. Digital Scholarship in the Humanities, 36, 138-152.](#)

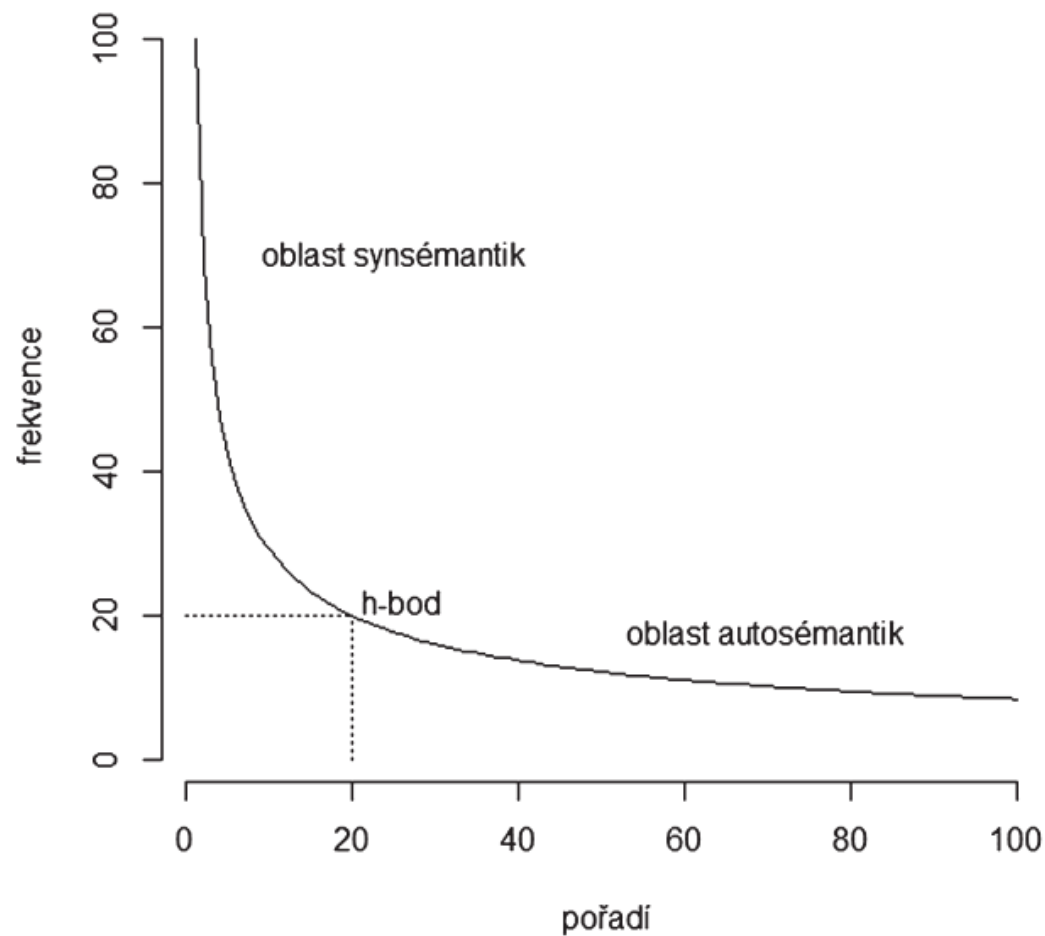
# Tematická slova



# Tematická slova

- tematická váha slova

$$TV_{\text{slovo}} = 2 \frac{(h - r')f(r')}{h(h - 1)f(1)}$$

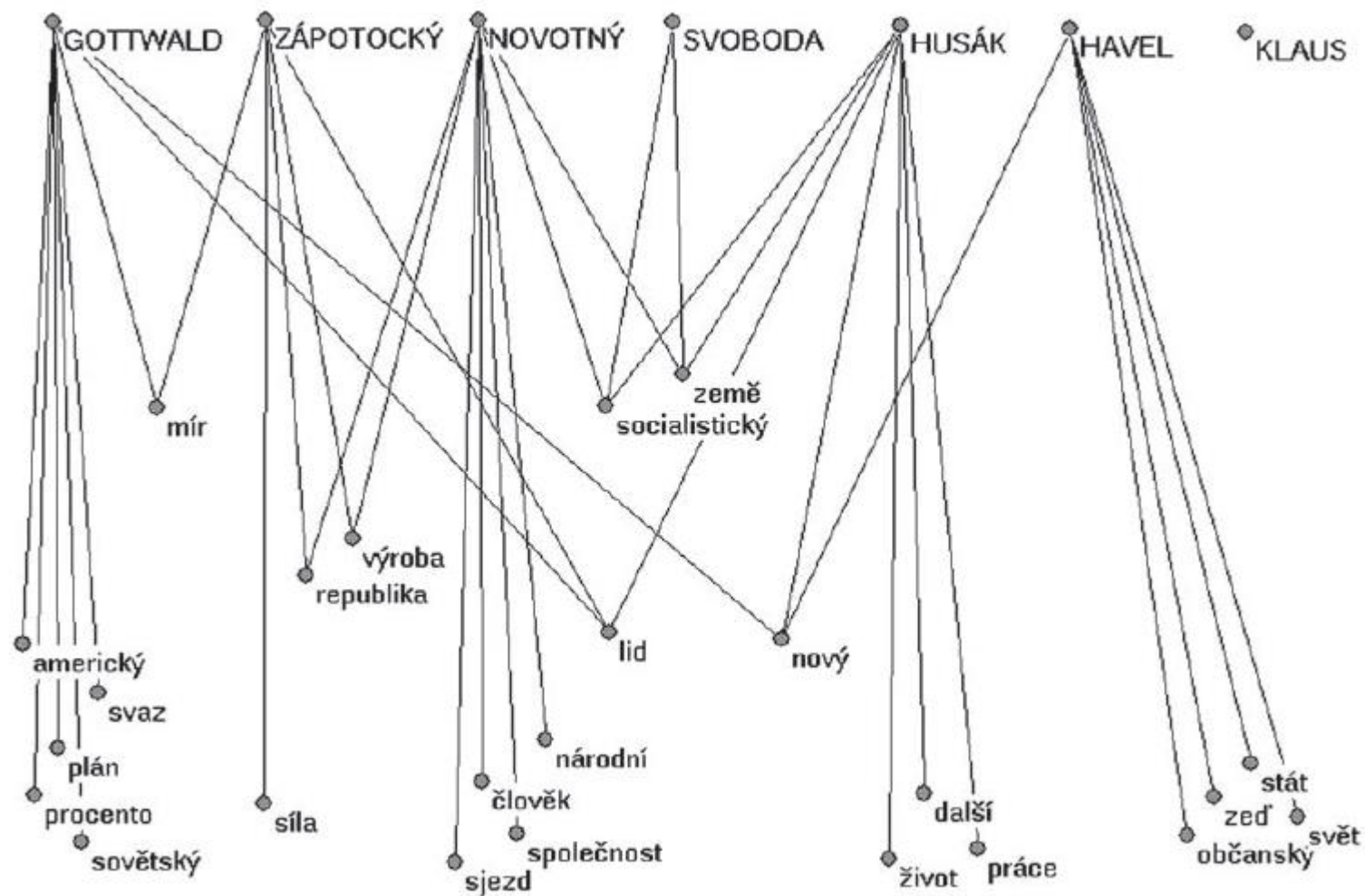


# Tematická slova

prezident	rok	f(1)	h-bod	tematická slova (r/f)	TK	tcu
Gottwald	1949	65	13	<i>rok</i> (5/25); <i>plán</i> (7/20)	0,063116	63,12
Gottwald	1950	93	16,6	<i>rok</i> (5,5/53); <i>procento</i> (7/37); <i>lid</i> (14/21)	0,082887	82,89
Gottwald	1951	87	14,66667	<i>rok</i> (5/51); <i>nový</i> (13,5/16)	0,058682	58,68
Gottwald	1952	98	14,66667	<i>rok</i> (4/40); <i>nový</i> (11,5/18); <i>americký</i> (11,5/18)	0,055048	55,05
Gottwald	1953	80	14	<i>rok</i> (6,5/22); <i>mír</i> (9/21); <i>sovětský</i> (9/21); <i>svaz</i> (13/16)	0,053709	53,71
Zápotocký	1954	133	17,33333	<i>rok</i> (11/25); <i>výroba</i> (16/19)	0,009756	9,76
Zápotocký	1955	99	14	<i>mír</i> (7/20); <i>rok</i> (9/17); <i>lid</i> (12,5/15);	0,027473	27,47
Zápotocký	1956	141	19	<i>rok</i> (10/31); <i>republika</i> (17/21)	0,013313	13,31
Zápotocký	1957	101	16	<i>rok</i> (12/19); <i>síla</i> (12/19); <i>lid</i> (14,5/17)	0,014645	14,65



# Tematická slova



# Tematická slova

$$S(A, B) = \frac{|A \cap B|^2}{|A| \cdot |B|}$$

<b>dvojice prezidentů</b>	<b>S</b>
Svoboda — Husák	0,2857
Novotný — Svoboda	0,25
Gottwald — Zápotocký	0,1
Zápotocký — Novotný	0,1
Gottwald — Husák	0,0714
Novotný — Husák	0,0714
Zápotocký — Husák	0,0286
Gottwald — Havel	0,0251
Husák — Havel	0,0204

# Slovní bohatství / diverzifikovanosti slovníku

T1

„Byl jsem doma a doma jsem jen ležel a ležel“

T2

„Byl jsem doma a tam jsem jen ležel nebo spal“

# Slovní bohatství / diverzifikovanosti slovníku

T1

„Byl **jsem** doma a doma **jsem** jen ležel a ležel“

T2

„Byl **jsem** doma a tam **jsem** jen ležel nebo spal“

# Slovní bohatství / diverzifikovanosti slovníku

T1

„Byl **jsem doma a doma jsem** jen **ležel a ležel**“

- N = 10 tokenů
- V = 5 typů {byl, jsem, doma, a, jen}

T2

„Byl **jsem** doma a tam **jsem** jen ležel nebo spal“

- N = 10 tokenů
- V = 9 typů {byl, jsem, doma, a, tam, jen, ležel, nebo, spal}

# Slovní bohatství / diverzifikovanosti slovníku

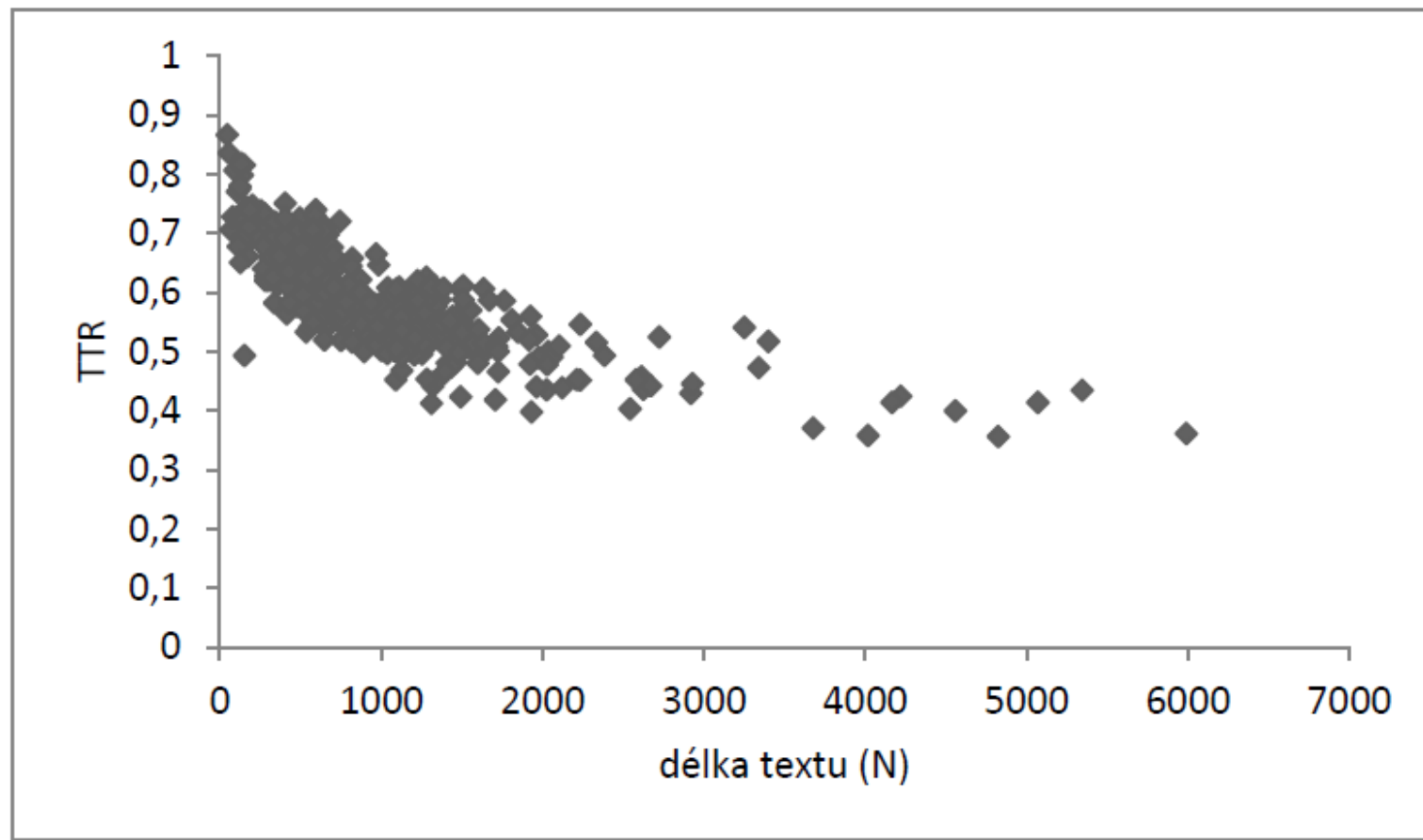
$$\text{TTR} = V / N$$

$$\text{TTR1} = 5 / 10 = 0,5$$

$$\text{TTR2} = 9 / 10 = 0,9$$



# Poměr typů a tokenů



Obr. 16. *TTR* v 760 textech Karla Čapka

# Poměr typů a tokenů

- nevhodný pro texty různé délky
- upravené způsob měření
  - standardizovaný TTR -> STTR
    - TTR na blocích textu
    - např. o velikosti 100, 1000,...  $n$  slov
  - klouzavý průměr TTR -> MATTR

# Klouzavý průměr TTR – MATTR

Marie miluje Petra. Petra miluje taky Jana. Petr miluje Emu.

# Klouzavý průměr TTR – MATTR

**Marie miluje Petra.** Petra miluje taky Jana. Petr miluje Emu.

- tokeny = 5
- typy = 3
- TTR =  $3/5 = 0.6$

# Klouzavý průměr TTR – MATTR

Marie **miluje Petra**. Petra miluje **taky** Jana. Petr miluje Emu.

- tokeny = 5
- typy = 3
- TTR =  $3/5 = 0.6$

# Klouzavý průměr TTR – MATTR

Marie miluje **Petra**. Petra **miluje taky Jana**. Petr miluje Emu.

- tokeny = 5
- typy = 4
- TTR =  $4/5 = 0.8$

# Klouzavý průměr TTR – MATTR

Marie miluje Petra. **Petra miluje taky Jana. Petr** miluje Emu.

- tokeny = 5
- typy = 5
- TTR =  $5/5 = 1$

# Klouzavý průměr TTR – MATTR

Marie miluje Petra. Petra **miluje taky Jana. Petr** miluje Emu.

- tokeny = 5
- typy = 5
- TTR =  $4/5 = 0.8$



# Klouzavý průměr TTR – MATTR

Marie miluje Petra. Petra miluje **taky Jana. Petr miluje Emu.**

- tokeny = 5
- typy = 5
- TTR =  $5/5 = 1$

# Klouzavý průměr TTR – MATTR

Marie miluje Petra. Petra miluje taky Jana. Petr miluje Emu.

$$MATTR = \frac{0.6 + 0.6 + 0.8 + 1 + 0.8 + 1}{6} = \frac{4.8}{6} = 0.8$$

# Klouzavý průměr TTR – MATTR

Marie miluje Petra. Petra miluje taky Jana. Petr miluje Emu.

$$MATTR = \frac{\sum_{i=1}^{N-L} V_i}{L(N - L + 1)}$$

$L$ ... velikost okna

$V_i$ ... počet typů v daném okně

$N$ ... délka textu

# Klouzavý průměr TTR – MATTR

Marie miluje Petra. Petra miluje taky Jana. Petr miluje Emu.

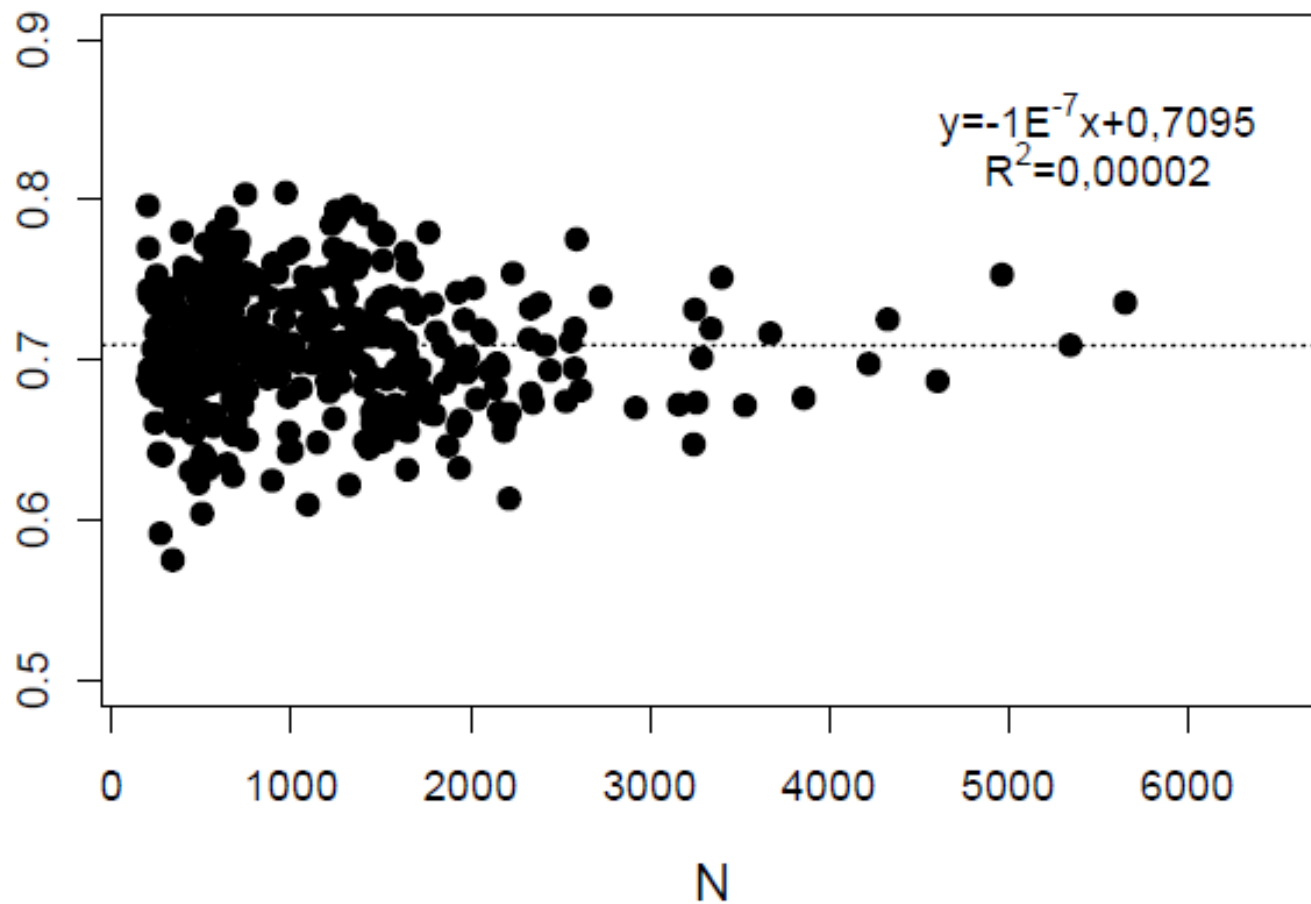
$$MATTR = \frac{\sum_{i=1}^{N-L} V_i}{L(N-L+1)} = \frac{3 + 3 + 4 + 5 + 4 + 5}{5(10 - 5 + 1)} = \frac{24}{30} = 0.8$$

$L$ ... velikost okna

$V_i$ ... počet typů v daném okně

$N$ ... délka textu

# Klouzavý průměr TTR – MATTR



# Klouzavý průměr TTR – MATTR

- velikost okna obvykle

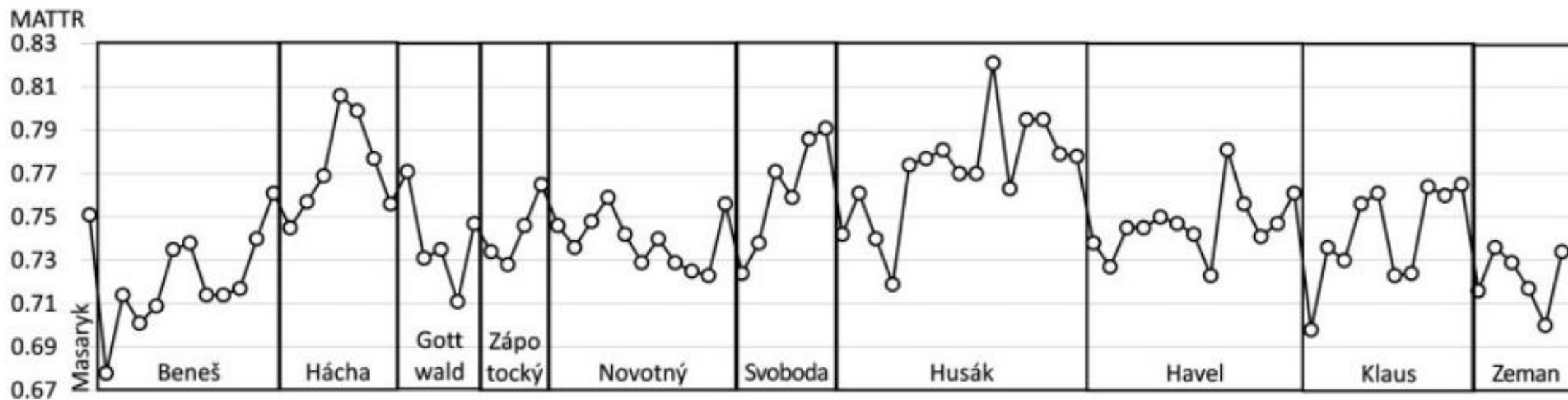
- $L = 100$
- $L = 500$

- nástroje

- QuitaUp
  - <https://korpus.cz/quitaup/>

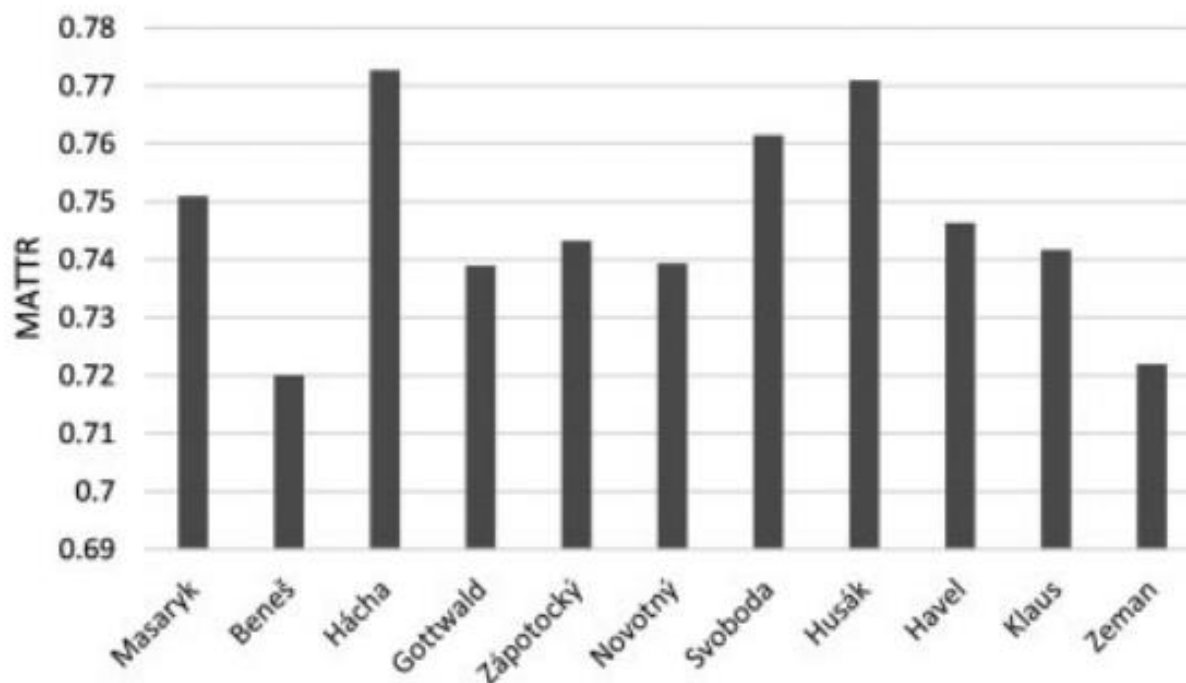
# Slovní bohatství / diverzifikovanosti slovníku

$$\text{MATTR}(L) = \frac{\sum_{i=1}^{N-L} V_i}{L(N-L+1)}$$



# Slovní bohatství / diverzifikovanosti slovníku

$$\text{MATTR}(L) = \frac{\sum_{i=1}^{N-L} V_i}{L(N-L+1)}$$

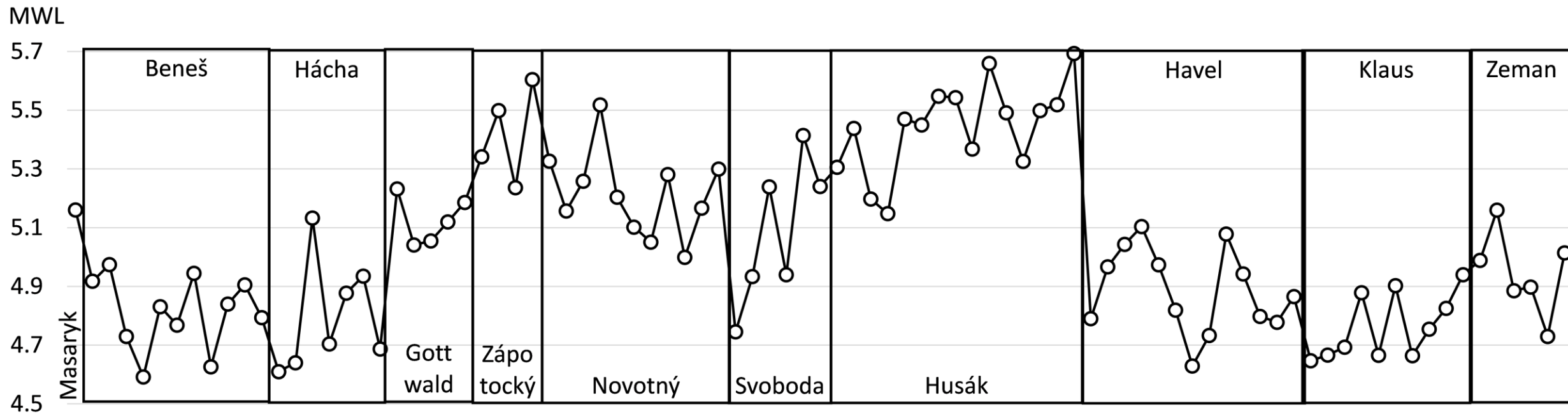




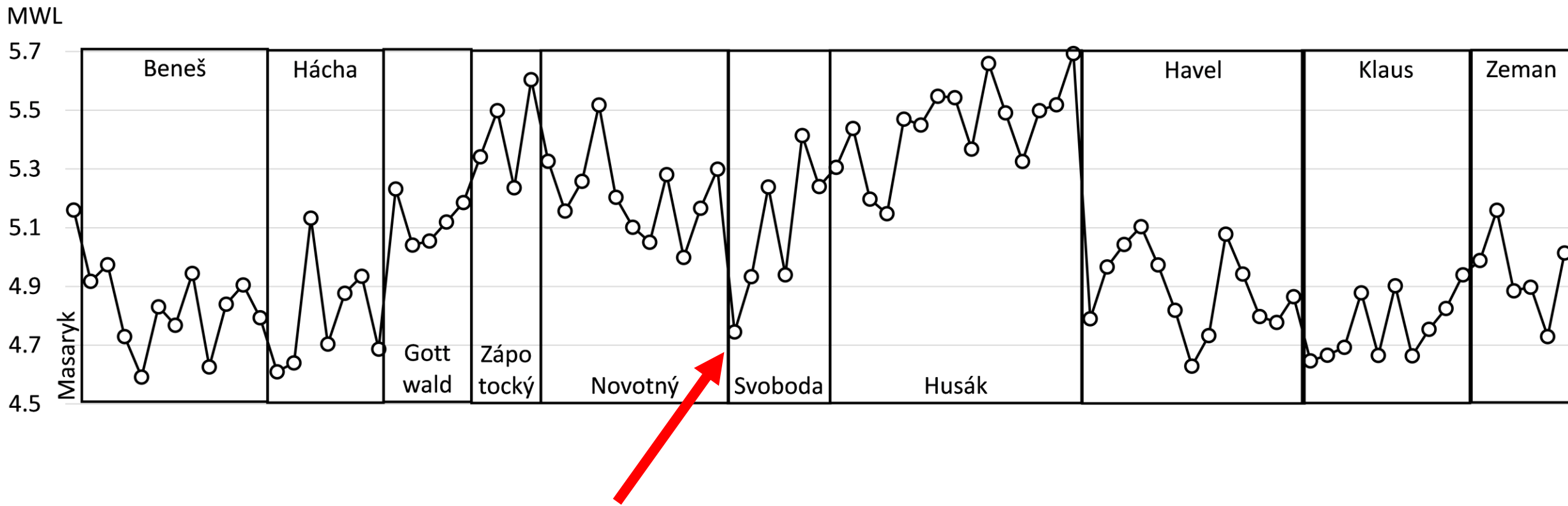
# Průměrná délka slova

- délka slova koreluje s frekvencí
- čím je slovo frekventovanější, tím je kratší
  - tendence

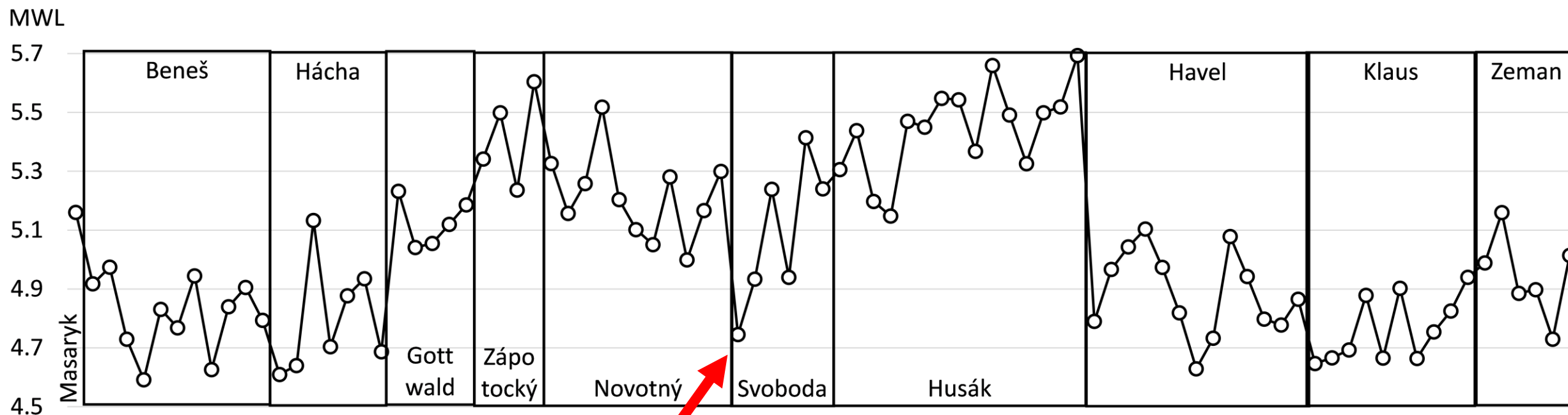
# Průměrná délka slova



# Průměrná délka slova

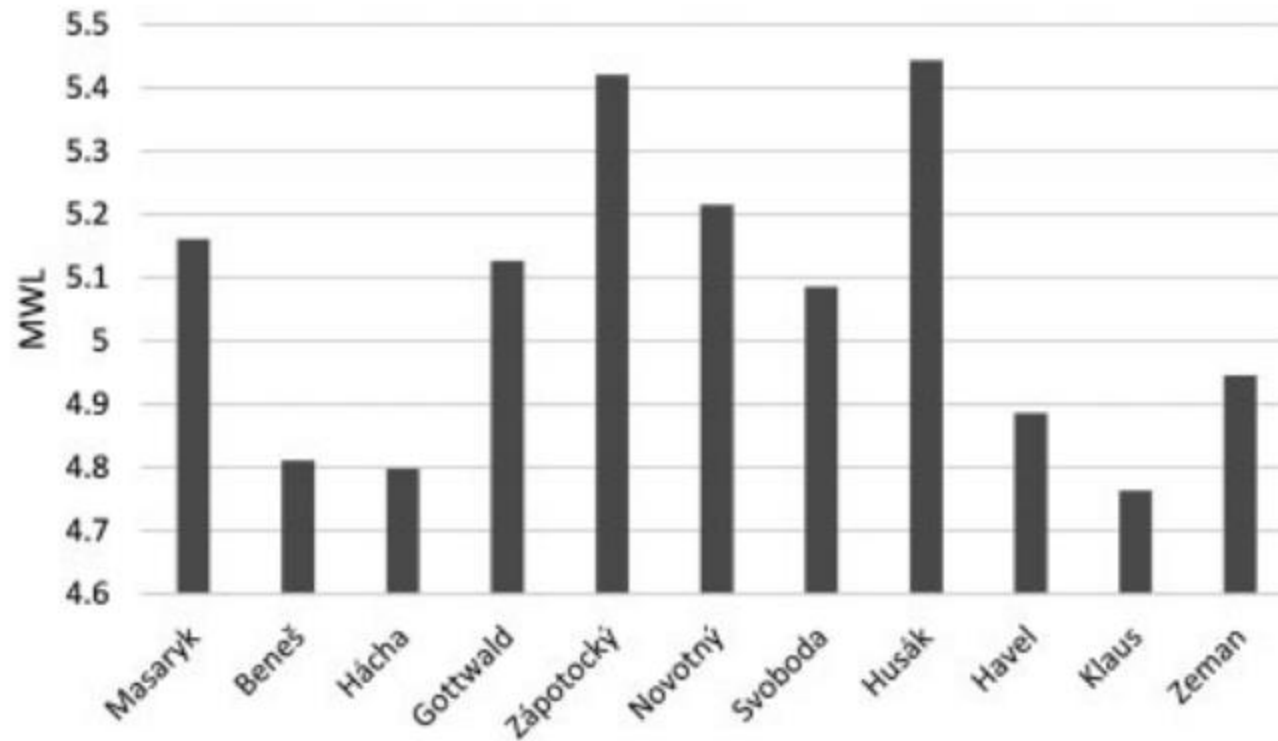


# Průměrná délka slova



<http://interaktivni.rozhlas.cz.s3-website.eu-central-1.amazonaws.com/prezidentske-projevy-2017/www/#1969-svoboda>

# Průměrná délka slova



# Aktivita / deskriptivita textu

T1

„Běžel domů, a když uviděl tu spoušť, vůbec neváhal, zahnal hladové psy a začal konat.“

T2

„Viděl dlouhé zelené stráně plné krásné zvěře, která se téměř nehýbala.“

# Aktivita / deskriptivita textu

T1

„**Běžel** domů, a když **uviděl** tu spoušť, vůbec **neváhal**, **zahnal** **hladové** psy a **začal konat**.“

T2

„**Viděl** **dlouhé zelené** stráně **plné krásné** zvěře, která se téměř **nehýbala**.“

# Aktivita / deskriptivita textu

$$Q = \frac{V}{V + A}$$

$$T1: V = 6, A = 1$$

$$Q1 = 6 / 7 = 0,86$$

$$T2: V = 2, A = 4$$

$$Q2 = 2 / 6 = 0,33$$





# Analýza novoroční projevů – vzdálenost mezi slovesy

T1

„**Běžel** domů, a když **uviděl** tu spoušť, vůbec **neváhal**, **zahnal** hladové psy a **začal konat**.“

T2

„**Viděl** dlouhé zelené stráně plné krásné zvěře, která se téměř **nehýbala**.“

# Analýza novoroční projevů – vzdálenost mezi slovesy

T1

„**Běžel** domů, a když **uviděl** tu spoušť, vůbec **neváhal**, **zahnal** hladové psy a **začal konat**.“

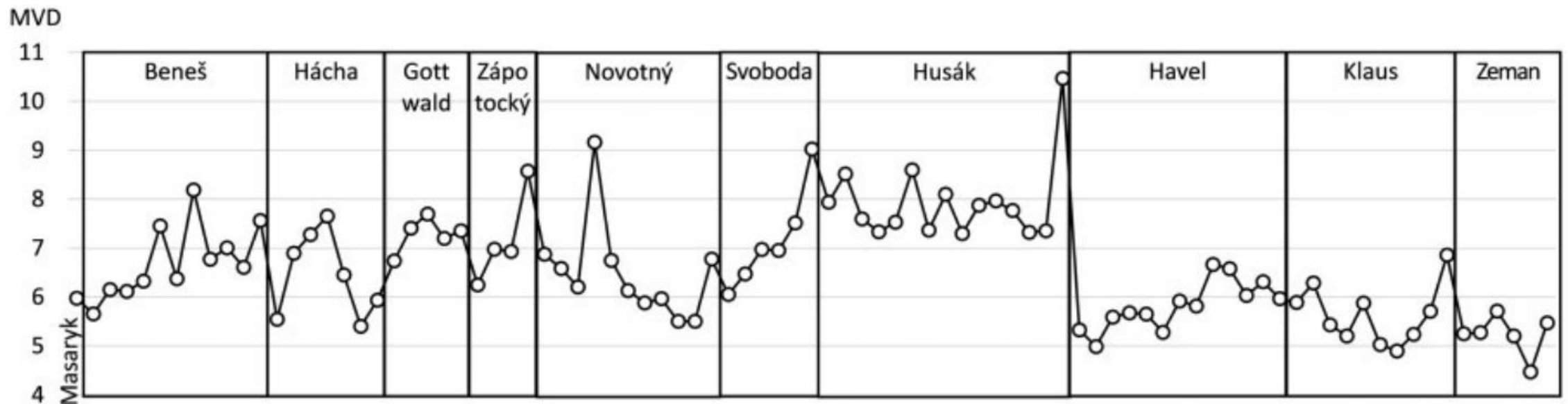
$$VD1 = (3 + 3 + 3) / 3 = 9 / 3 = 3$$

T2

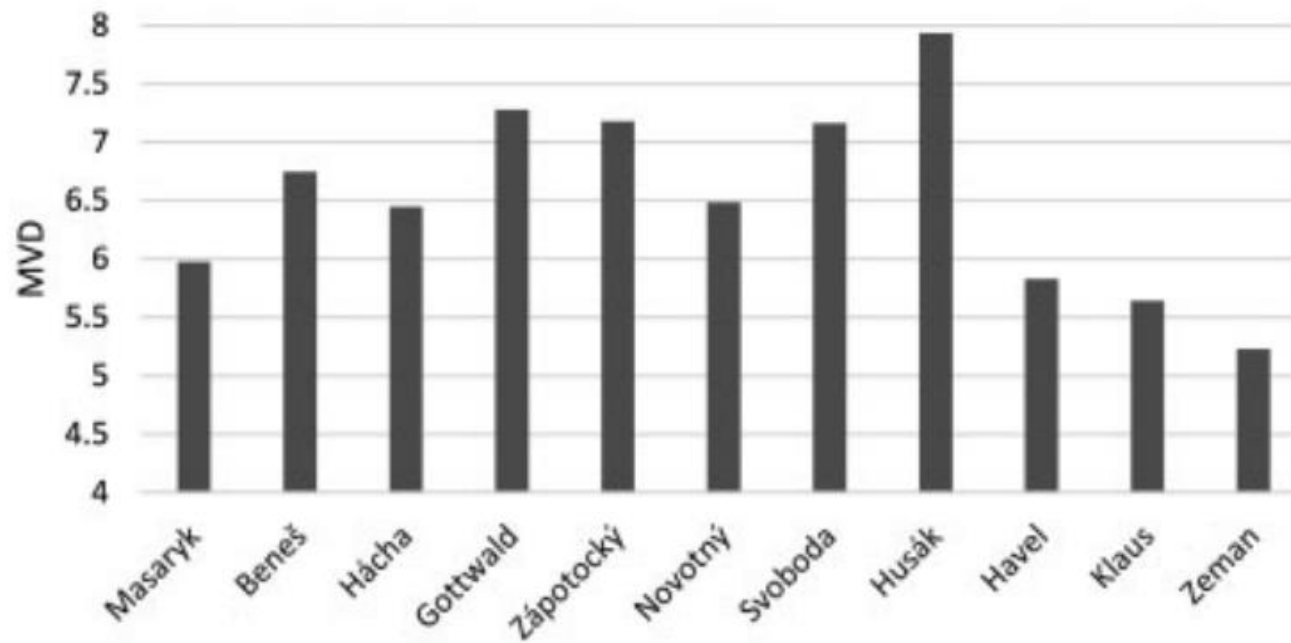
„**Viděl** dlouhé zelené stráně plné krásné zvěře, která se téměř **nehýbala**.“

$$VD2 = 9$$

# Analýza novoroční projevů – vzdálenost mezi slovesy



# Analýza novoroční projevů – vzdálenost mezi slovesy



# Analýza novoroční projevů – proporce nejfrekventovanějších slov

- nejfrekventovanější slova → zpravidla synsémantika
- relativní frekvence
- nezávislost na tématu
- více Eder (2017)

# Analýza novoroční projevů – proporce nejfrekventovanějších slov

- nejfrekventovanější slova → zpravidla synsémantika
- relativní frekvence
- nezávislost na tématu
- více Eder (2017)

# Analýza novoroční projevů – proporce nejfrekventovanějších slov

$$\Delta_{(AB)} = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - \mu_i}{\sigma_i} - \frac{B_i - \mu_i}{\sigma_i} \right|$$

$n$  ... vybraný počet nejfrekventovanějších slov

$A, B$  ... texty

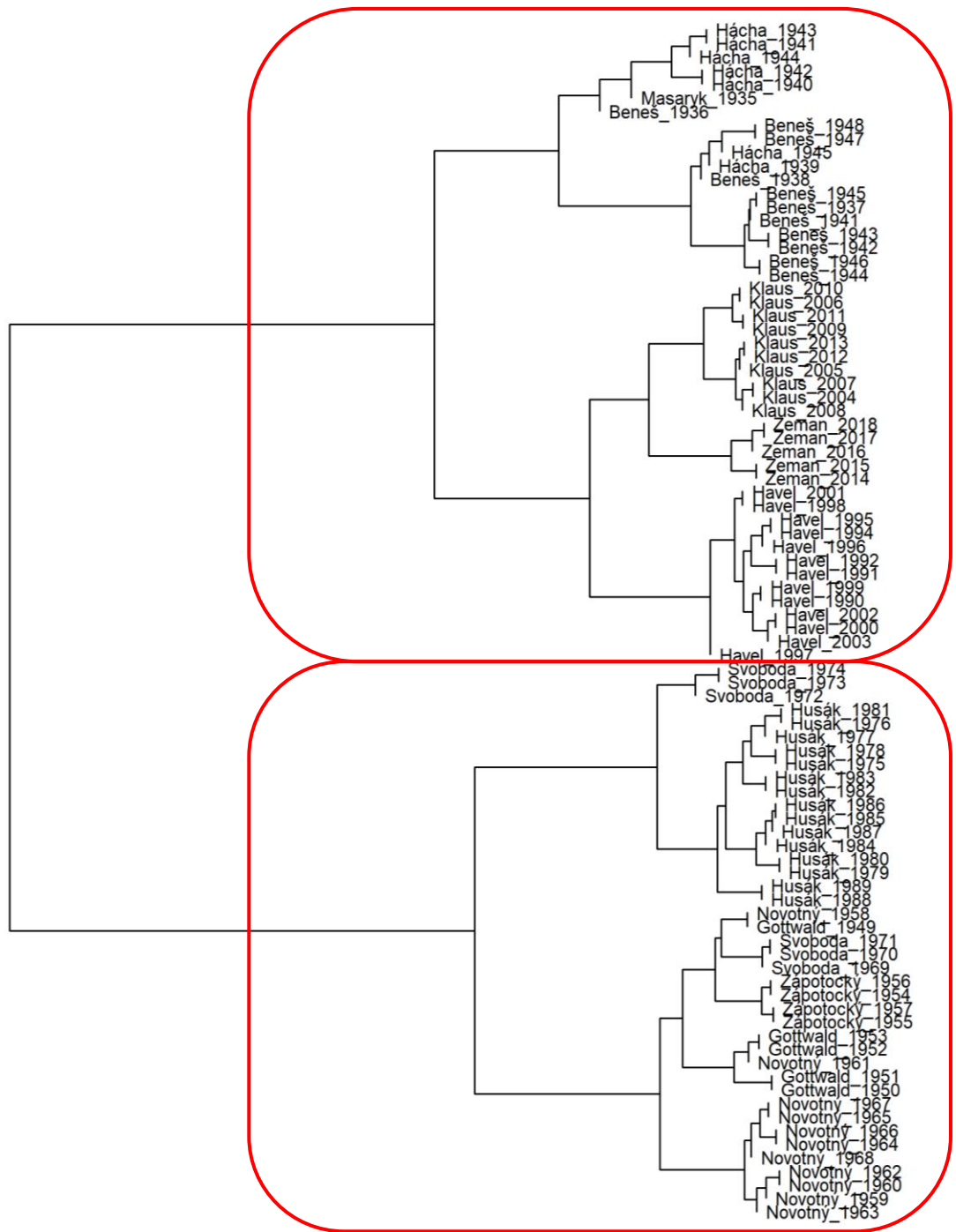
$A_i$  ... frekvence daného slova v textu  $A$

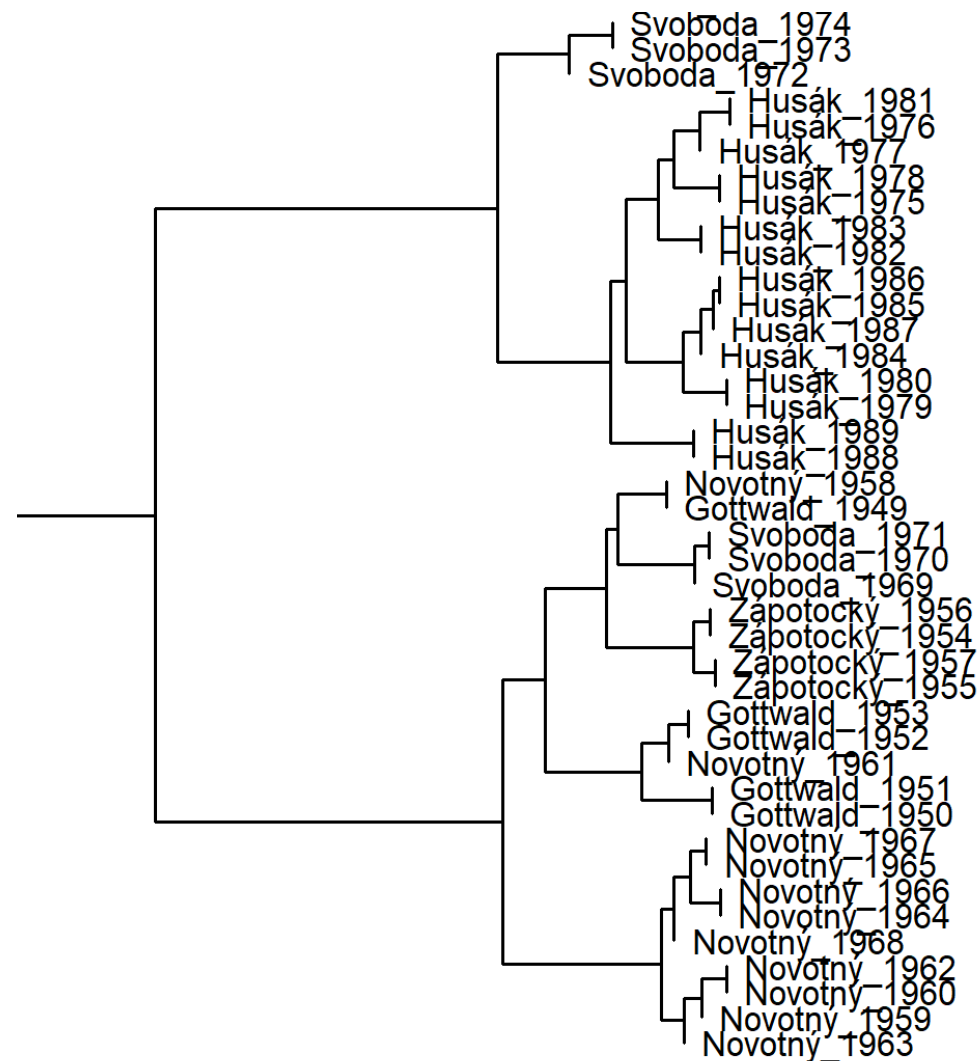
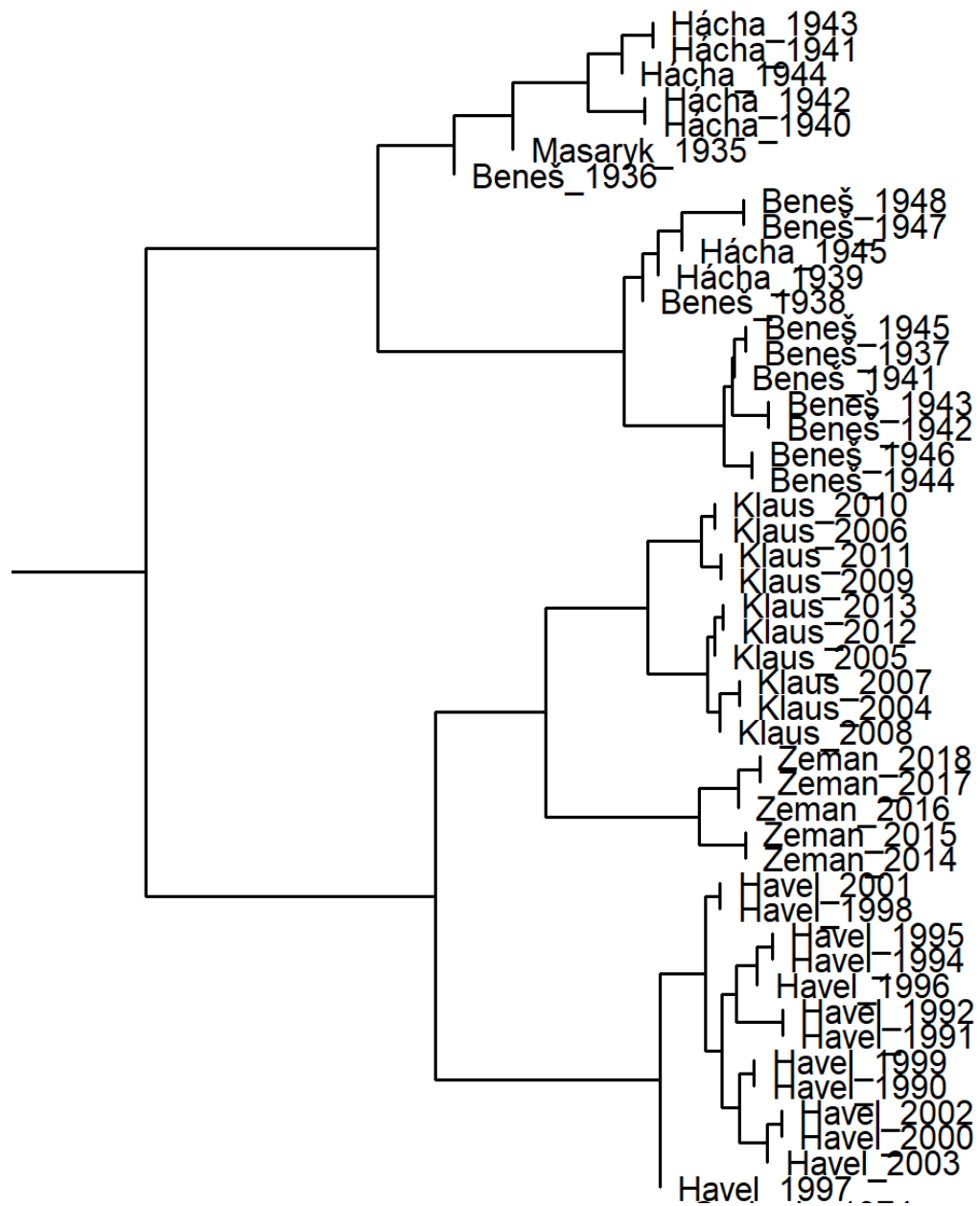
$B_i$  ... frekvence daného slova v textu  $B$

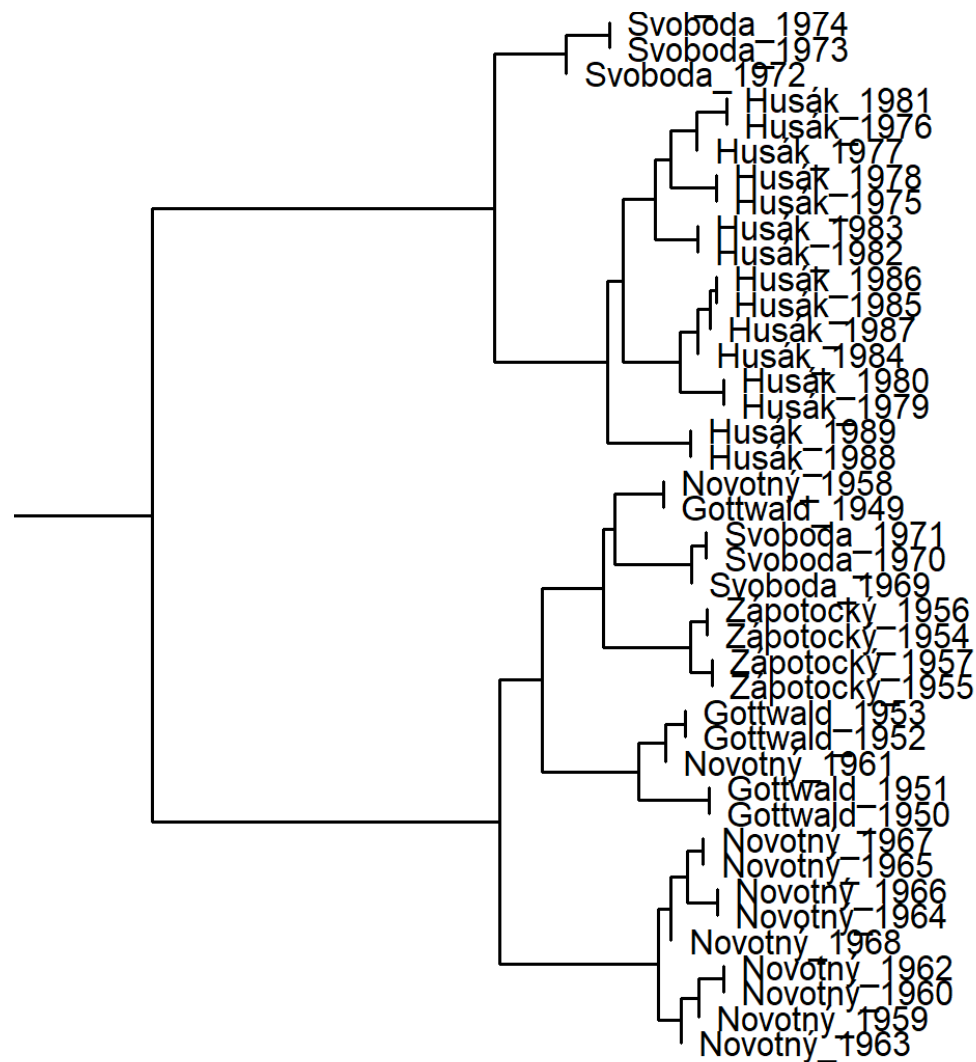
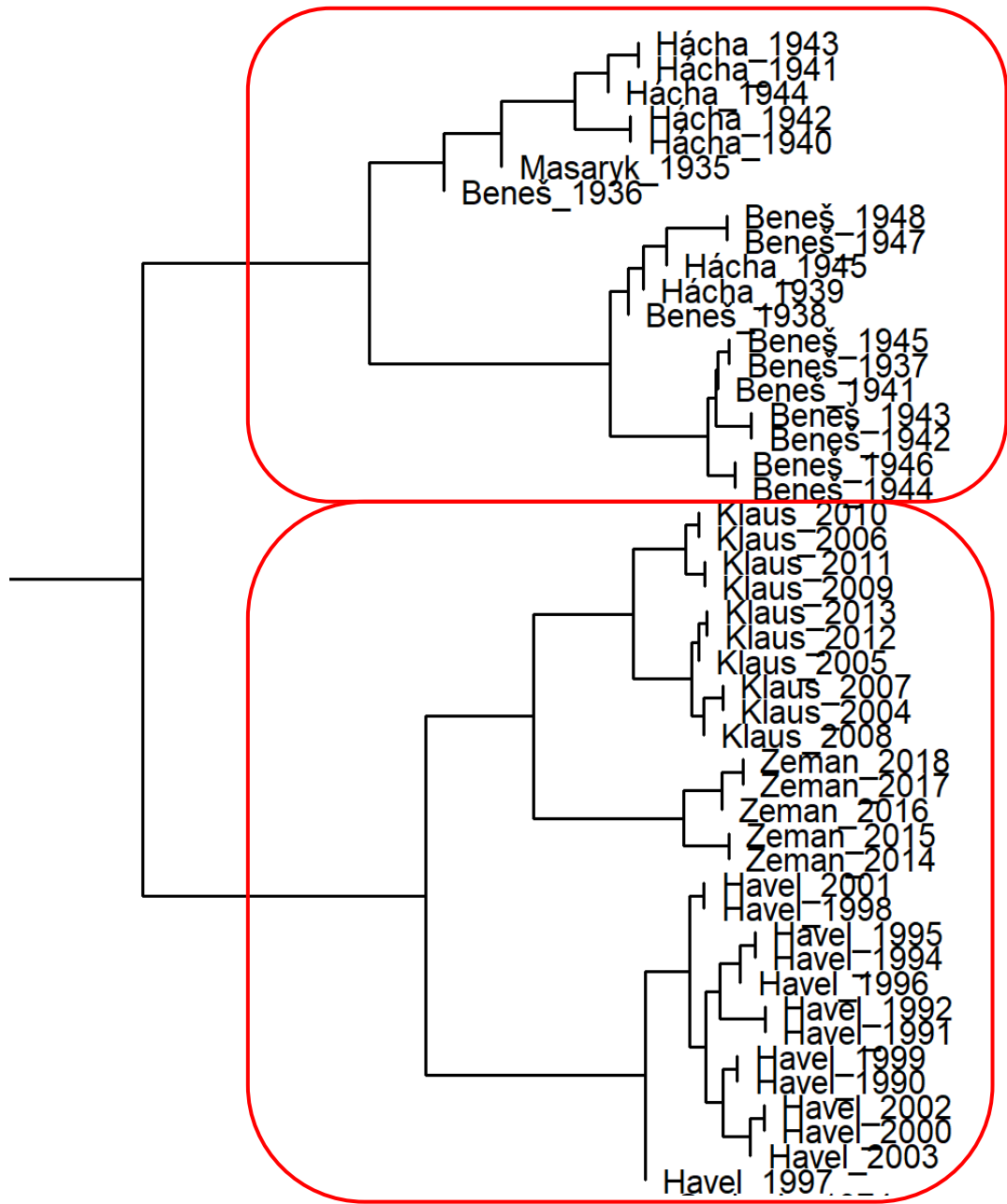
$\mu_i$  ... průměrná frekvence daného slova ve všech textech (v korpusu)

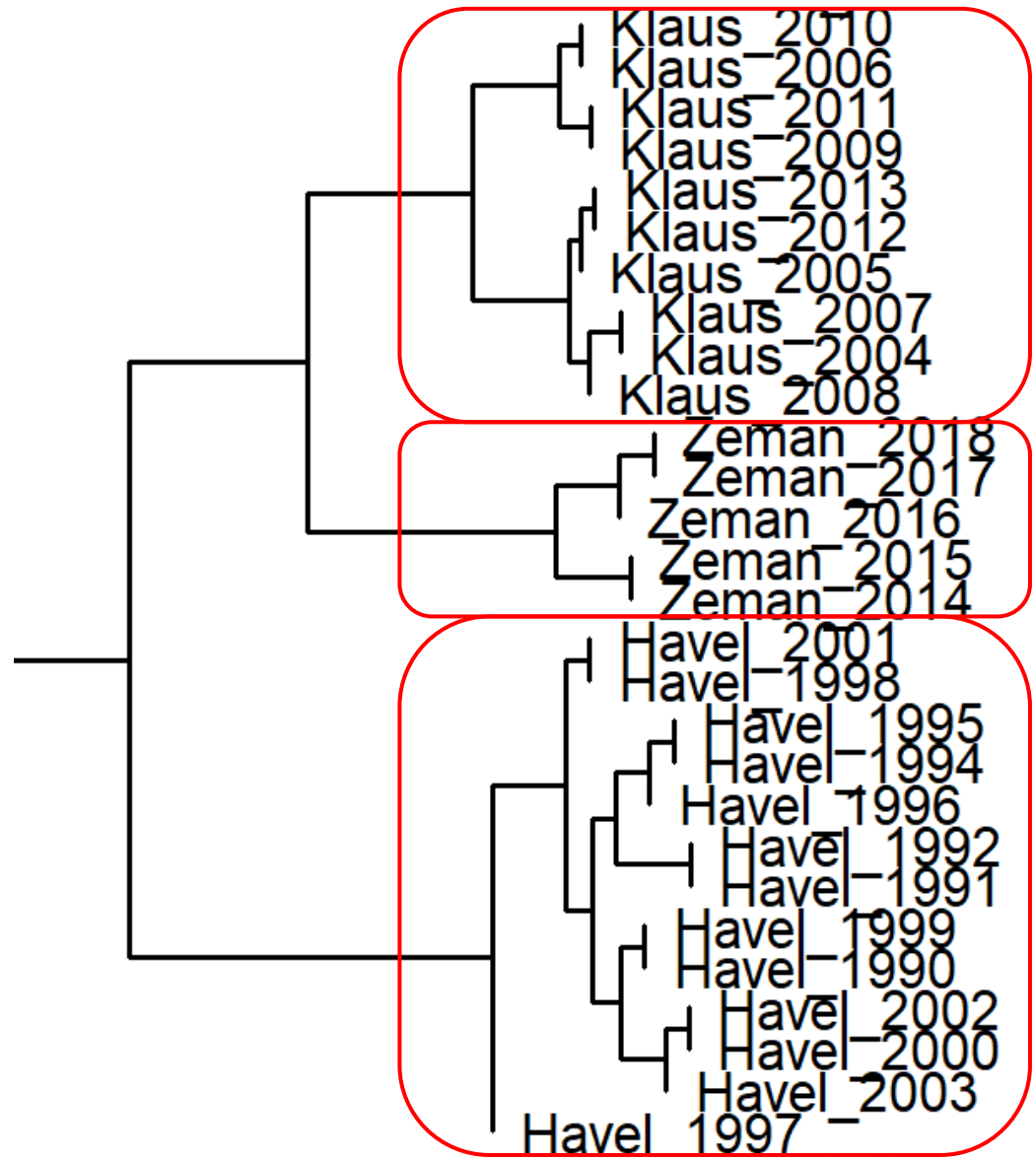
$\sigma_i$  ... směrodatná odchylka frekvence daného slova ze všech textů











## 4. QuitaUp

- aneb jak jednoduše měřit vybrané charakteristiky textů...

## 4. QuitaUp



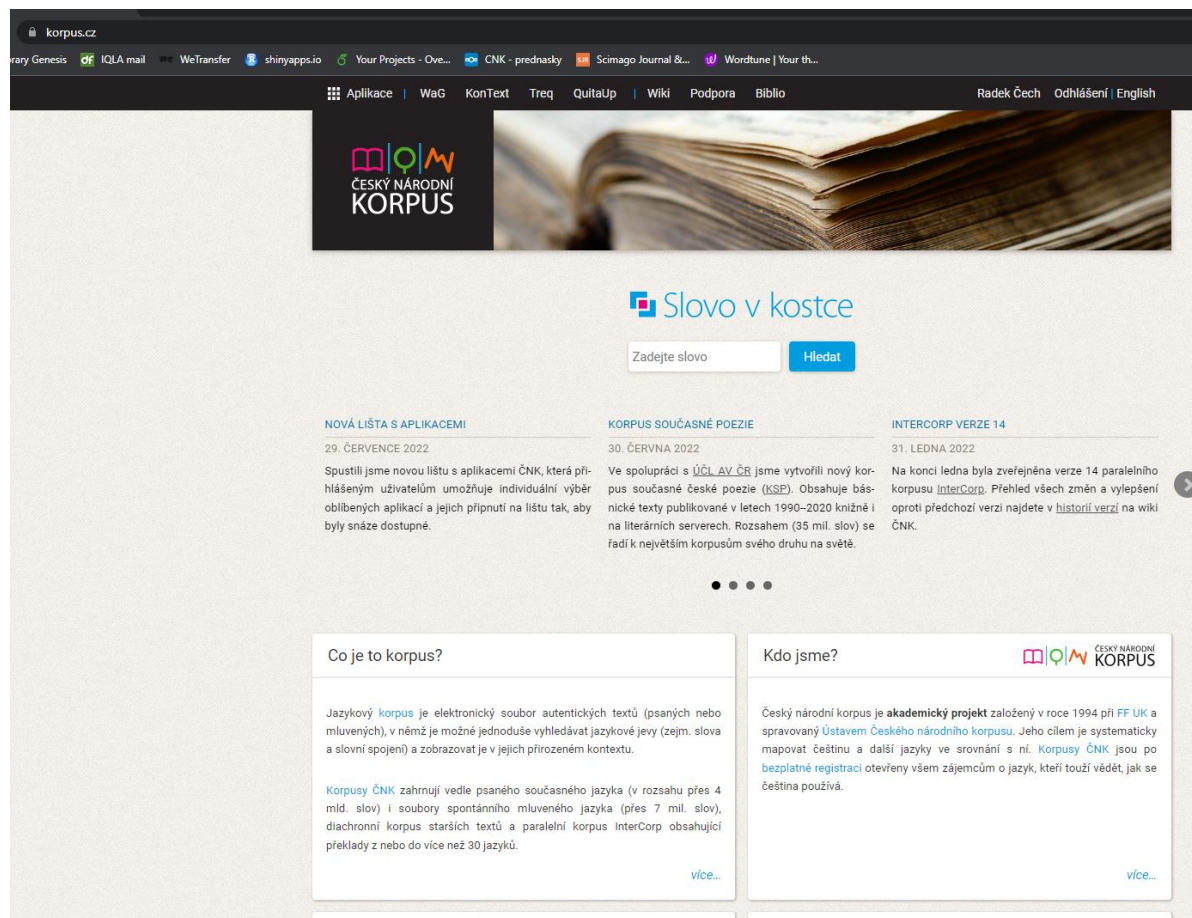
---

Katedra českého jazyka

Filozofická fakulta Ostravské univerzity

# 4. QuitaUp

- <https://www.korpus.cz/>



The screenshot shows the homepage of the website korpus.cz. The browser's address bar displays 'korpus.cz'. The navigation menu includes 'Applikace', 'WaG', 'KonText', 'Treq', 'QuitaUp', 'Wiki', 'Podpora', and 'Biblio'. The main header features the logo for 'ČESKÝ NÁRODNÍ KORPUS' and a banner image of an open book. Below the banner is a search section titled 'Slovo v kostce' with a search input field labeled 'Zadejte slovo' and a 'Hledat' button. Three news items are displayed: 'NOVÁ LIŠTA S APLIKACEMI' (29. ČERVENCE 2022), 'KORPUS SOUČASNĚ POEZIE' (30. ČERVNA 2022), and 'INTERCORP VERZE 14' (31. LEDNA 2022). At the bottom, there are two informational boxes: 'Co je to korpus?' and 'Kdo jsme?'. The 'Co je to korpus?' box explains that a corpus is an electronic collection of authentic texts and provides details about the 'Korpusy ČNK'. The 'Kdo jsme?' box states that the project was founded in 1994 and is managed by the 'Ústavem Českého národního korpusu'.

# 4. QuitaUp

- <https://korpus.cz/quitaup/>

QuitaUp

Vyberte soubor

Prohlízet txt, rtf (odt, doc, pdf)

Jazyk

čeština

Jednotky

Slovní tvary (case insensitive)

Ignorovat interpunkci

Náhled Výsledky O aplikaci

Vítejte!

Aplikace *QuitaUp* byla vytvořena za účelem poskytnout lingvistům i širšímu okruhu zájemců jednoduchý nástroj pro výpočet vybraných stylometrických indikátorů, které kvantitativně vyjadřují některé vlastnosti textu. Patří sem například výpočet slovního bohatství, tematické koncentrace či aktivity textu (viz dále).

Vytvoření softwaru *QuitaUp* bylo podpořeno grantovým projektem SGS č. 02/FF/2020–2021 *Reflexe jazykové a jazykovědné problematiky v nelingvistických textech* (poskytovatel Ostravská univerzita, Filozofická fakulta) a projektem *Kreativita a adaptabilita jako předpoklad úspěchu Evropy v propojeném světě*, reg. č.: CZ.02.1.01/0.0/0.0/16\_019/0000734, financovaným z Evropského fondu pro regionální rozvoj.



EVROPSKÁ UNIE  
Evropské strukturální a investiční fondy  
Operační program Výzkum, vývoj a vzdělávání



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



Děkuji za pozornost!

<https://cechradek.cz/>