Čeština pro 21. století

Jak s ní nakládají prezidenti?

Radek Čech

Struktura přednášky

1. možnosti analýzy (nejen politických) projevů a textů
   - aneb čtěme 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 čtěme 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, zd
```

Možnosti analýzy (nejen politických) projevů a textů

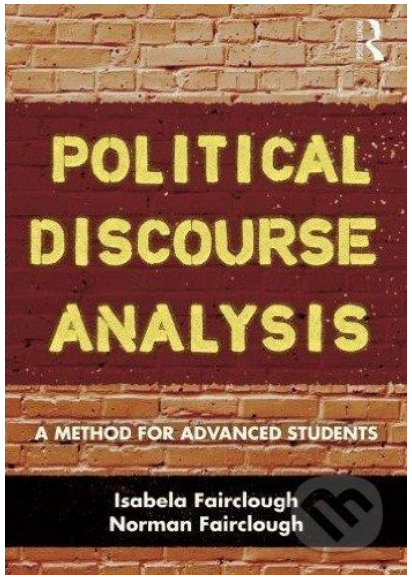
- jak na analýzu textů?
- využít početní nástroje….
- kvantitativní analýza
  - obsahové i formální analýzy
  - limity?

```
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, zd
```

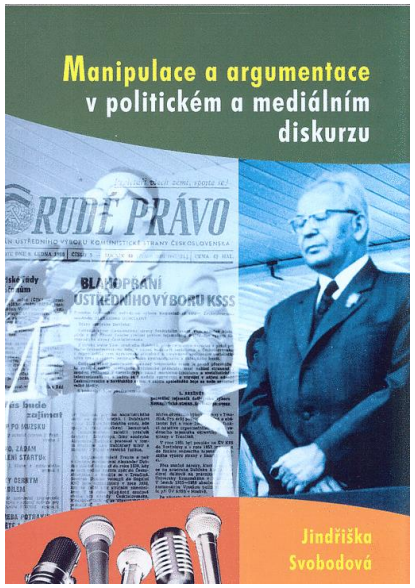
Možnosti analýzy (nejen politických) projevů a textů

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

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. Cílil 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

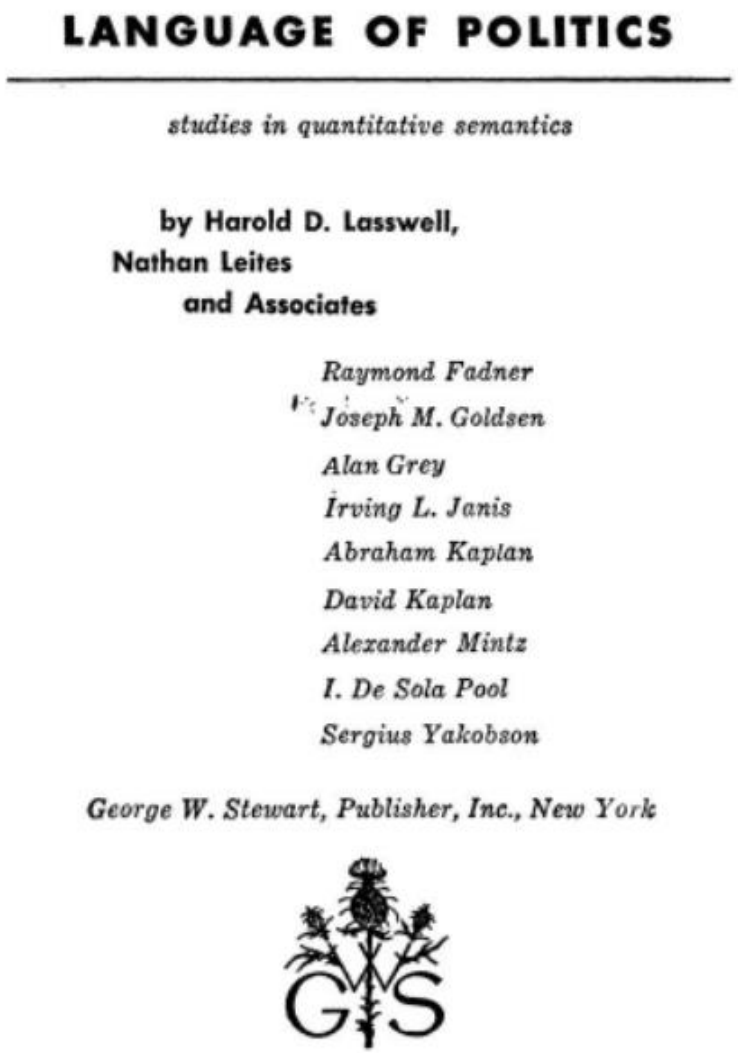
„Patrná je snaha vyvolat iluzi dialogu, který samozřejmě v monologickém projevu není možný. Sám svou řeč pojmenoval jako rozhovor, navozoval tedy pocit, že lidé stojí přímo před ním a mají možnost reakce, a zodpovídal nevyřčené otázky, předjímal možné protesty a výtky, které by mohly posluchače napadnout.

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 zvídavosti, 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

GWS

Kvantitativní analýza politických projevů

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

CONTENTS

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

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?

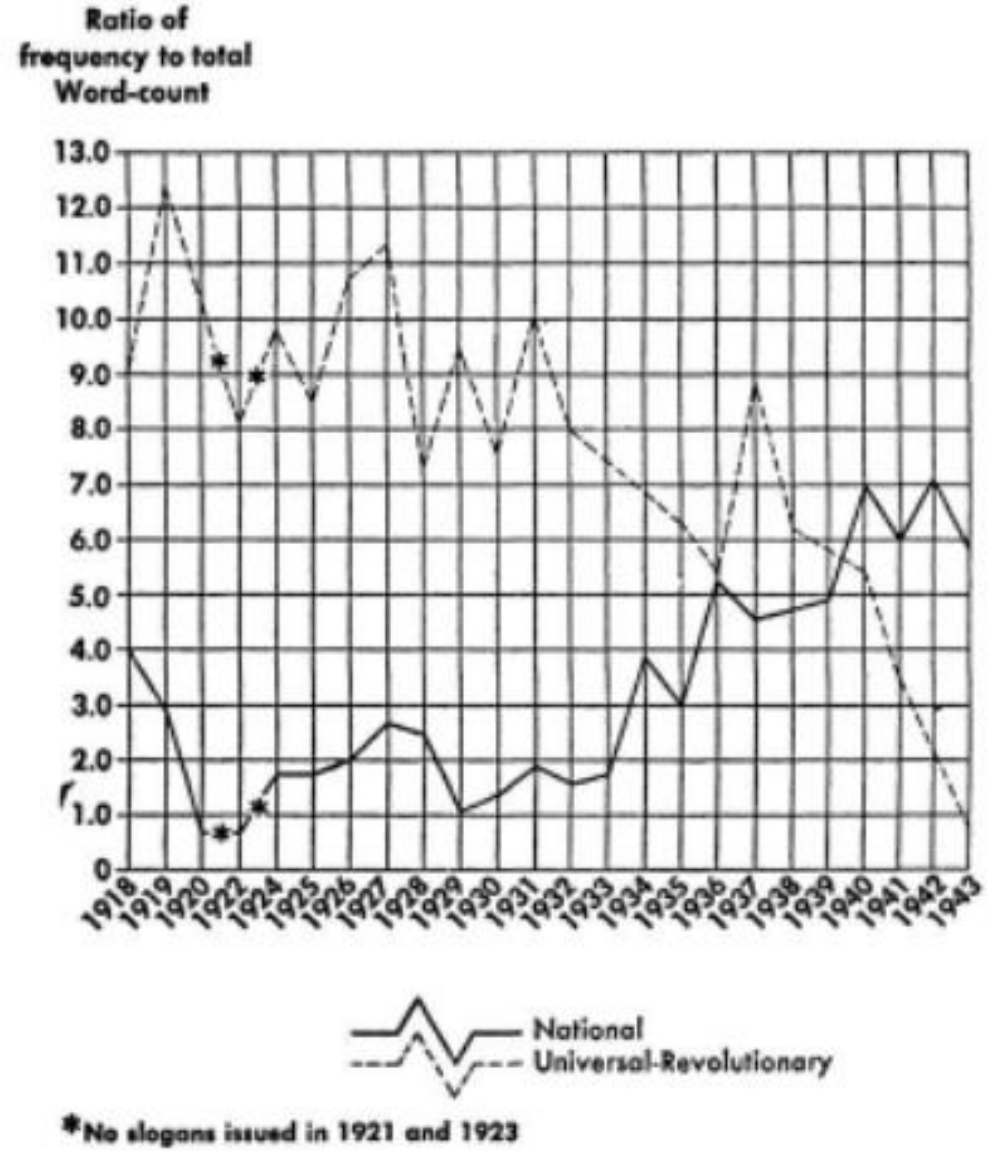
- pozornost
- **jasně vymezená vlastnost vzorku**

  - 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.)



*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

Frekvenční analýza slov (lemmat)

pořadí	Gottwald (1952)	Zápotocký (1954)	Havel (1999)	Klaus (2006)
1.	rok	rok	zeď	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	zeď	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	zeď	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	zeď	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	$f_{SYN2010}$	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	$f_{SYN2010}$	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	$f_{SYN2010}$	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

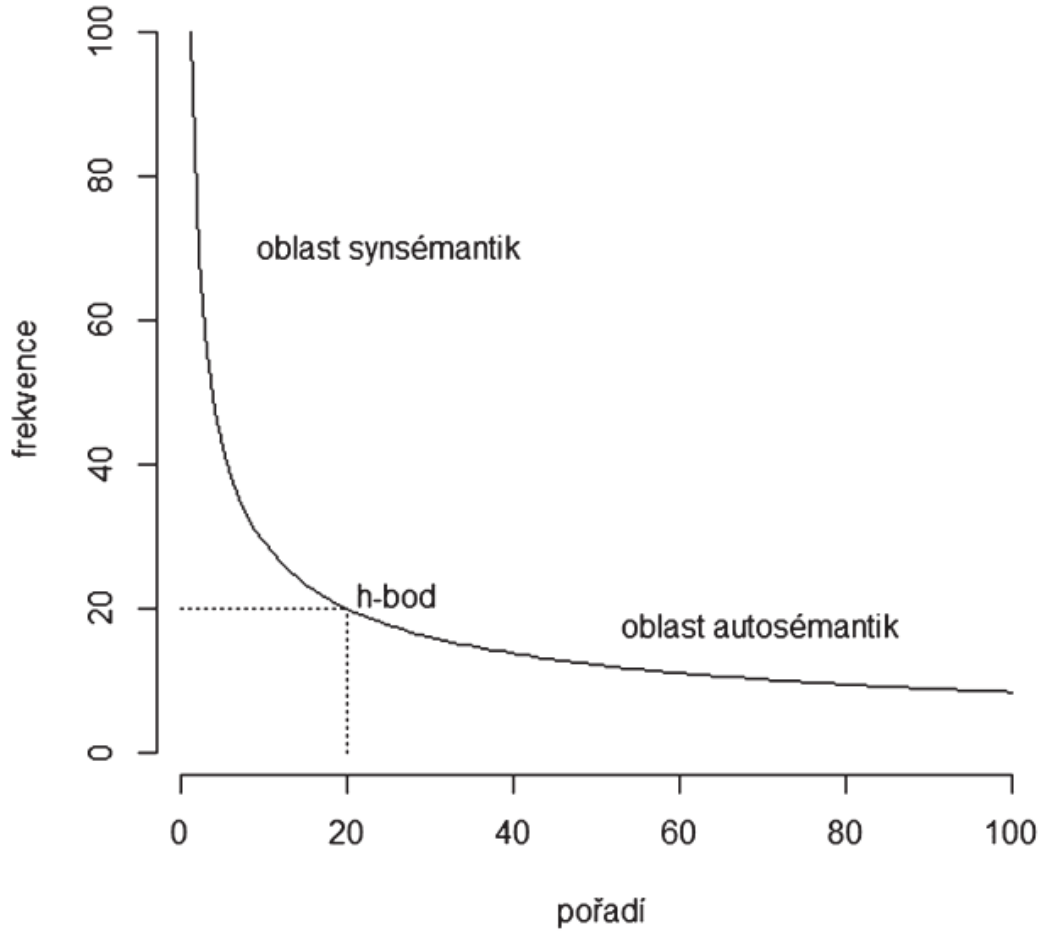
nejfrekventovanější slova		**klíčová slova**			
Klaus (2006)	f	Klaus (2006)	f	$f_{SYN2010}$	log likelihood
život	8	volba	7	23 529	36,41
rok	7	politika	6	18 866	31,99
volba	7	spoluobčan	3	717	31,29
politika	6	občan	5	14 679	27,33
země	6	přát	5	16 608	26,13
evropský	5	vážený	3	2 373	24,15
občan	5	život	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 (na první pohled) „skrytých“ vlastností

Tematická koncentrace textu

- tematická váha slova

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



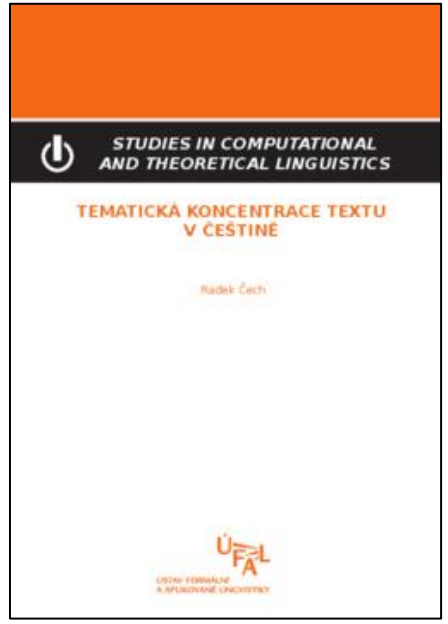
Tematická koncentrace textu

- tematická váha slova

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

- tematická koncentrace textu

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

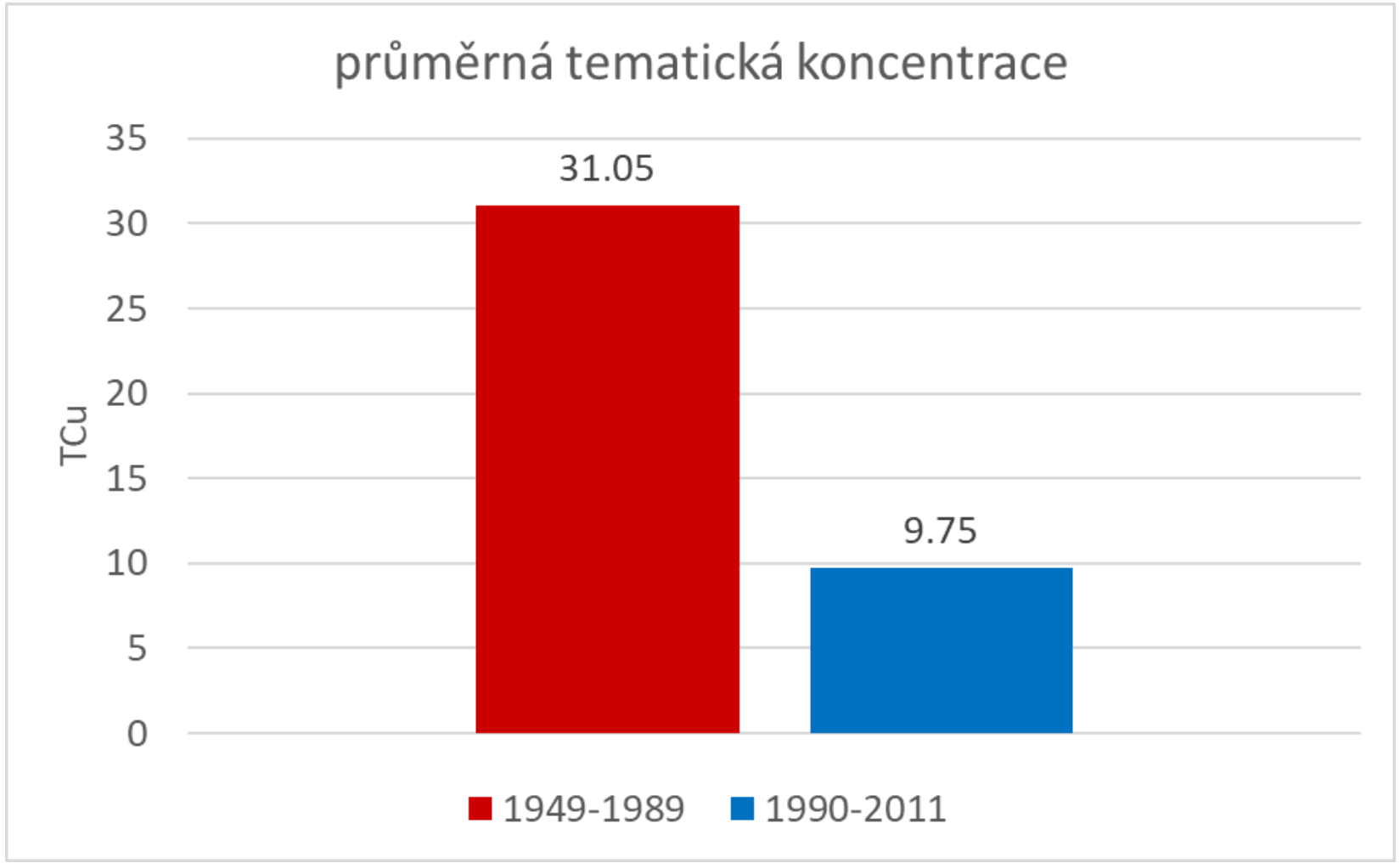


Tematická koncentrace textu

- „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

Novoroční a vánoční projevy československých a českých prezidentů

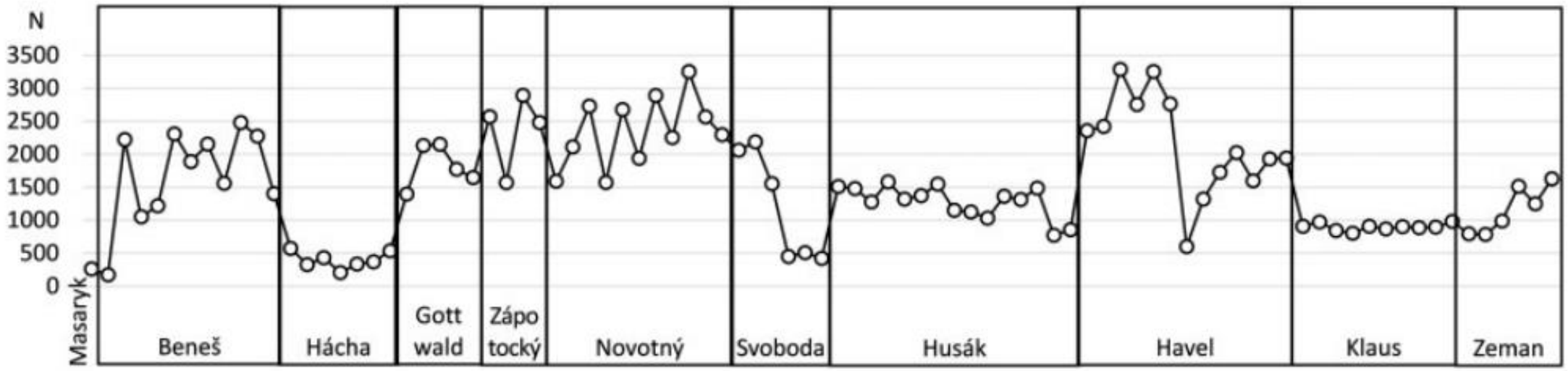


Fig. 1 Chronologically ordered text lengths (*N*) of presidential annual addresses

3. kvantitativní charakteristiky prezidentských projevů

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

Kvantitativní analýzy novoroční projevů

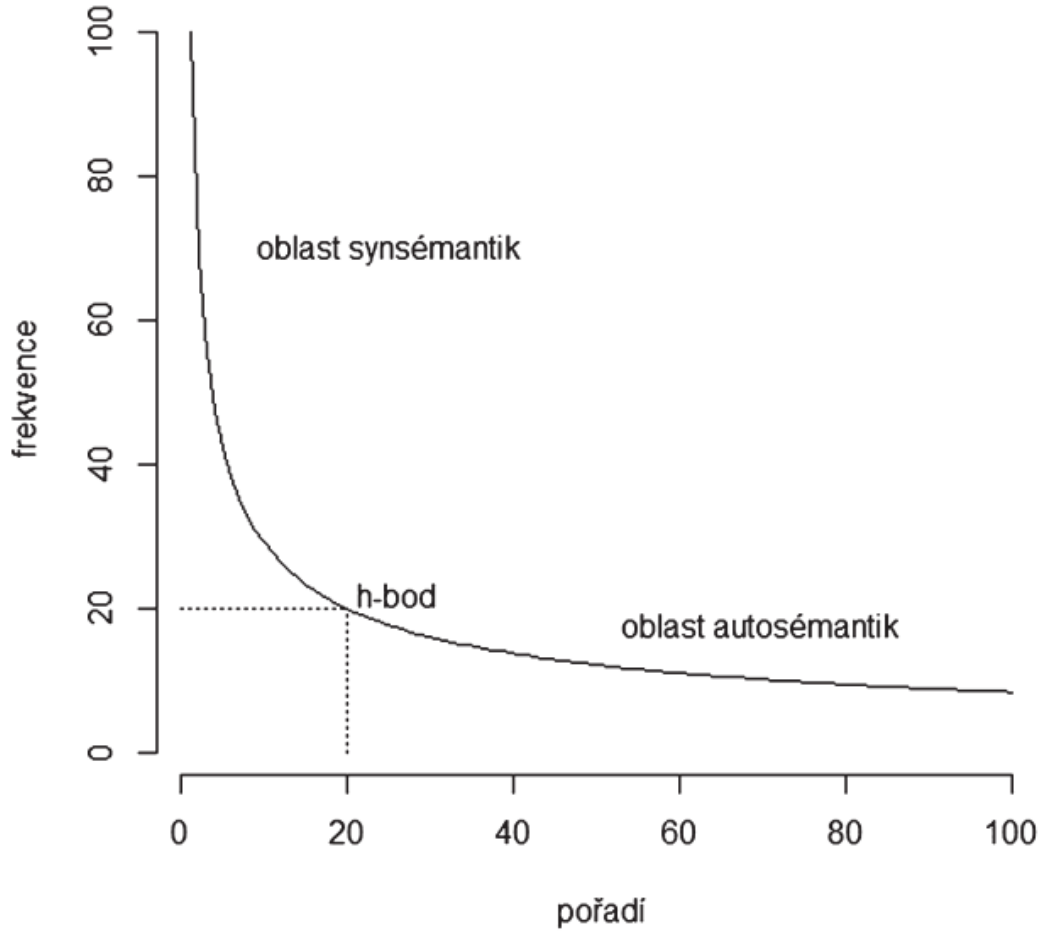
- tematická slova
- slovní bohatství
- průměrná délka slova (tokenu)
- aktivita textu
- vzdálenost mezi slovesy
- proporce nejfrekventovanějších slov

Tematická slova

Tematická slova

- tematická váha slova

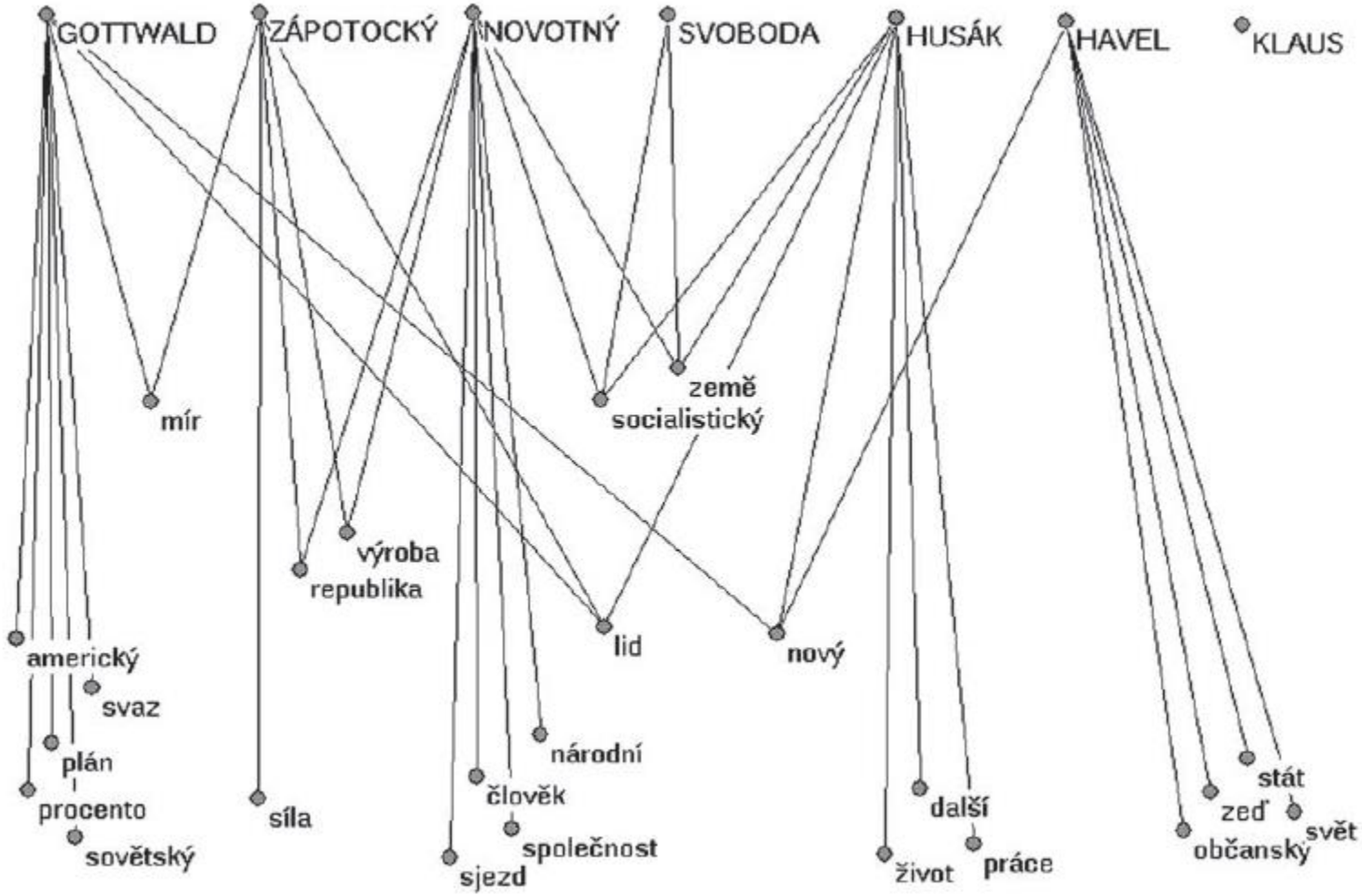
$$TV_{slovo} = 2\frac{(h - r')f(r')}{h(h-1)f(1)}.$$



Tematická slova

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

Tematická slova



Tematická slova

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

dvojice prezidentů	S
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

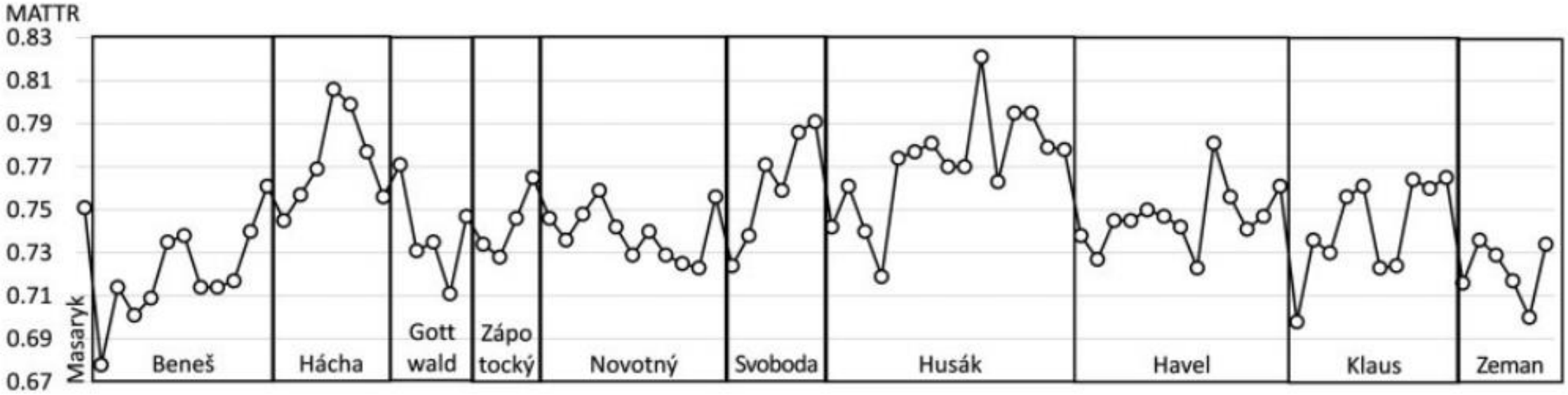
TTR = V / N

TTR1 = 5 / 10 = 0,5

TTR2 = 9 / 10 = 0,9

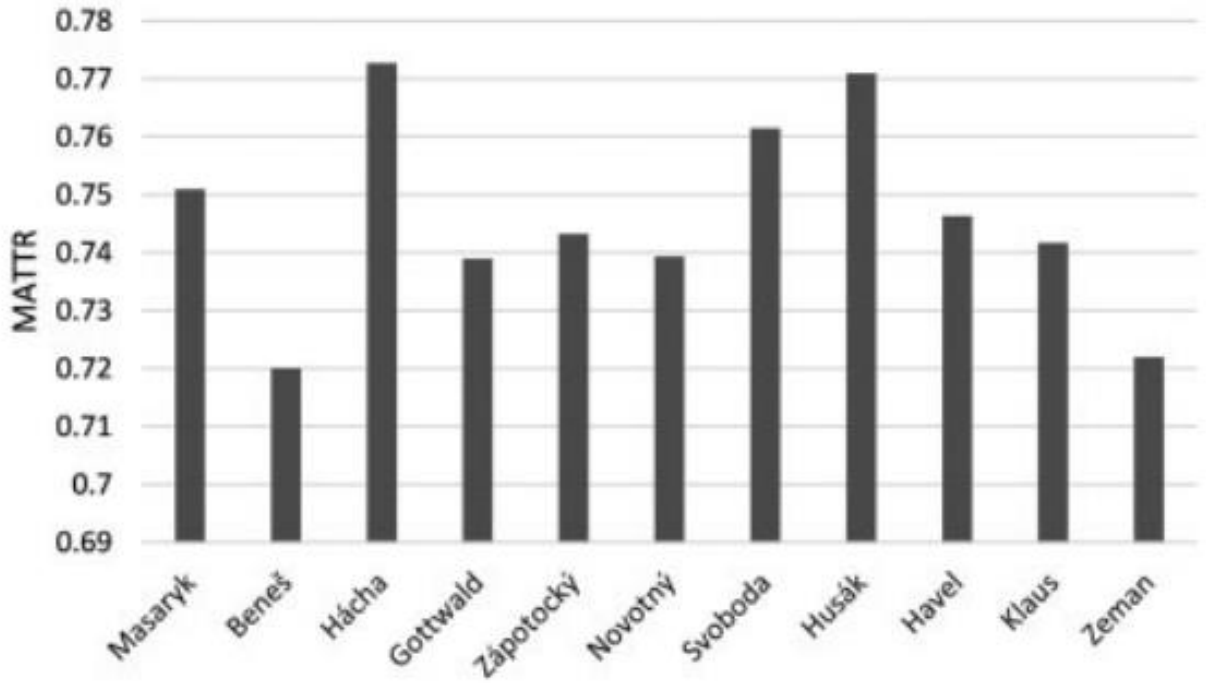
Slovní bohatství / diverzifikovanosti slovníku

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



Slovní bohatství / diverzifikovanosti slovníku

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

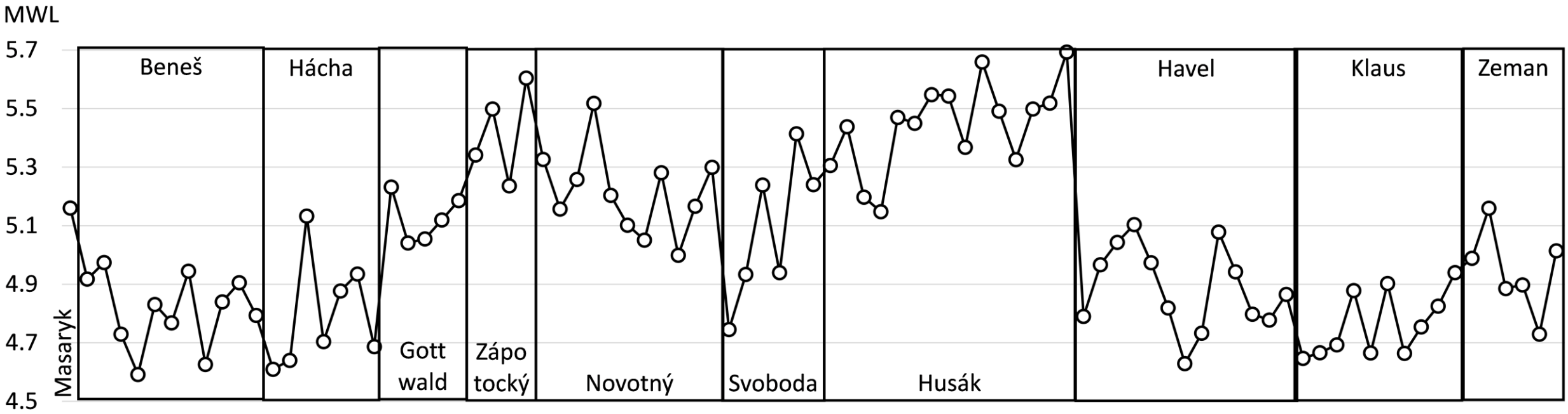


Průměrná délka slova

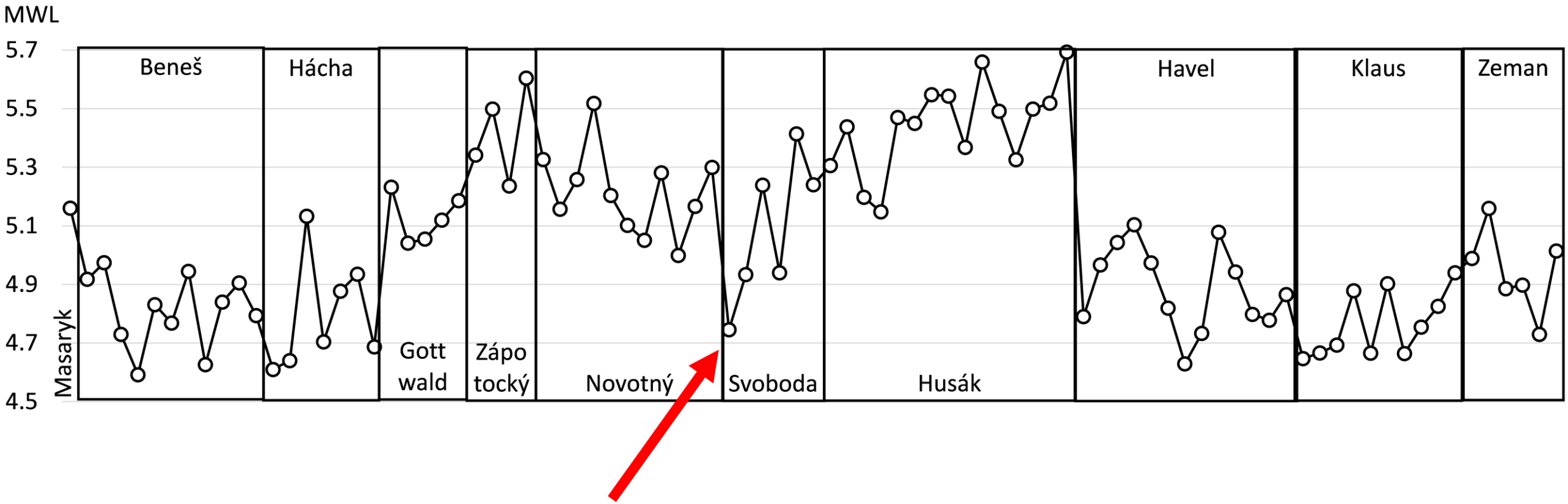
- délka slova koreluje s frekvencí

- čím je slov frekventovanější, tím je kratší
  - tendence

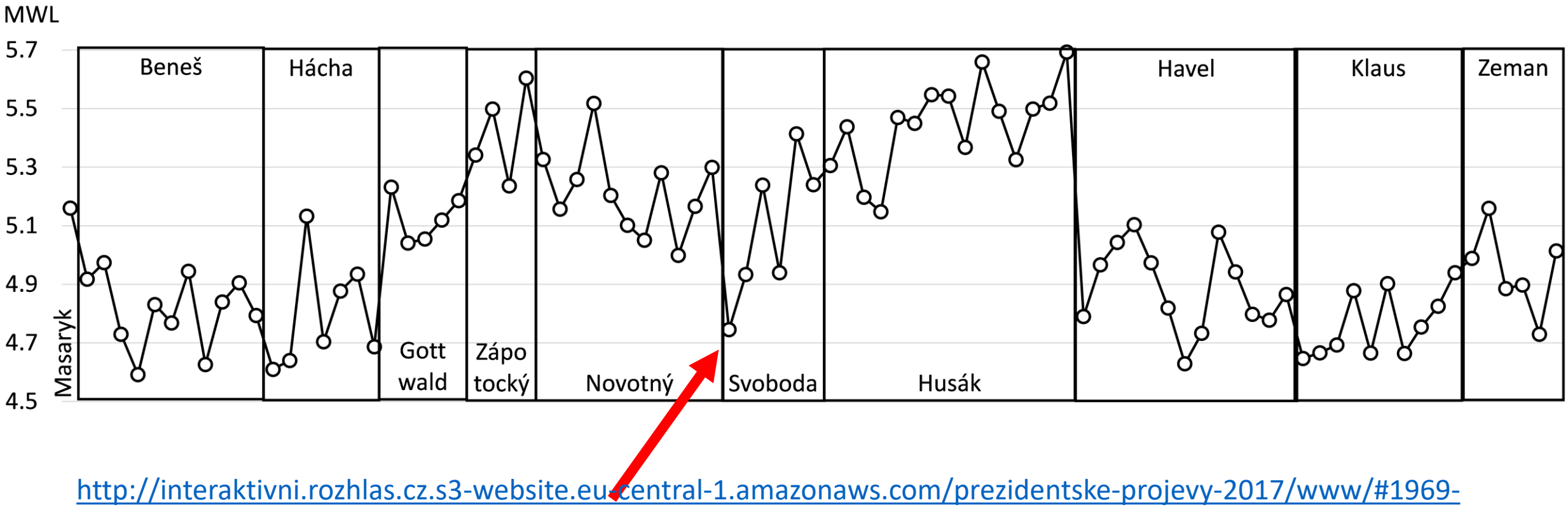
Průměrná délka slova



Průměrná délka slova

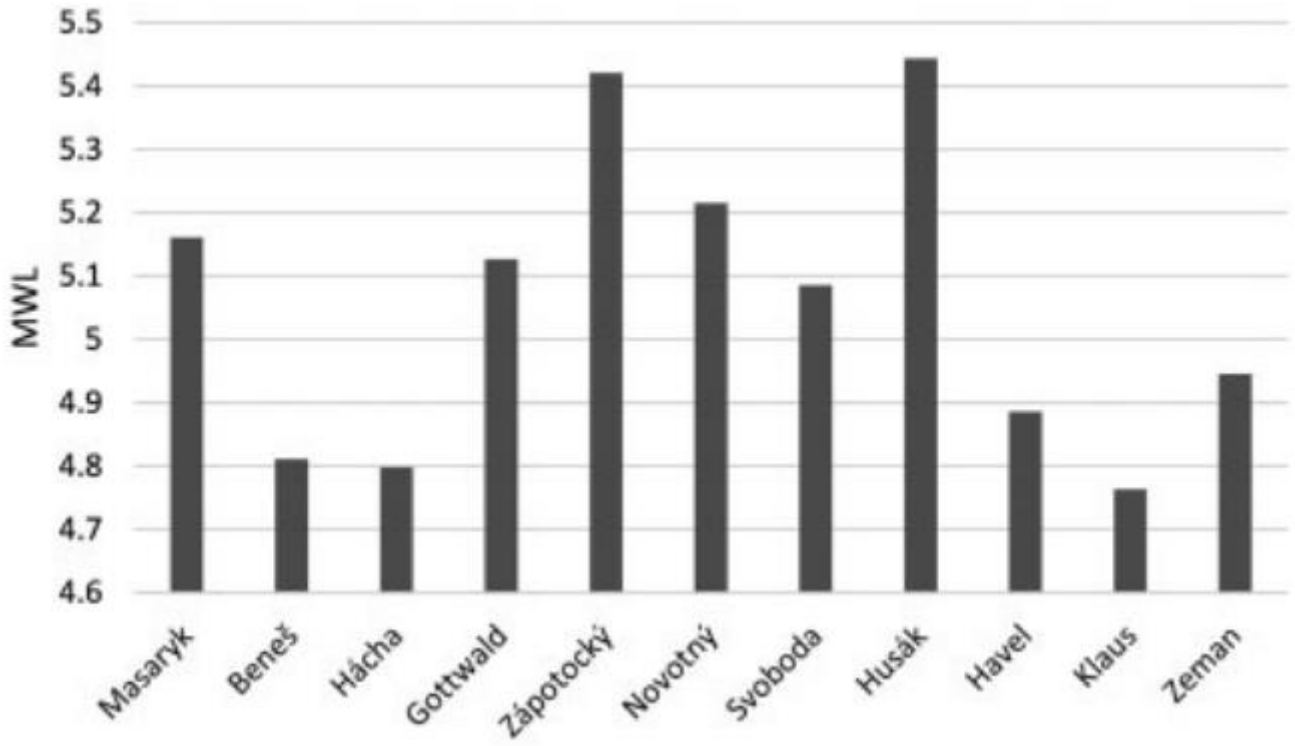


Průměrná délka slova



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.“

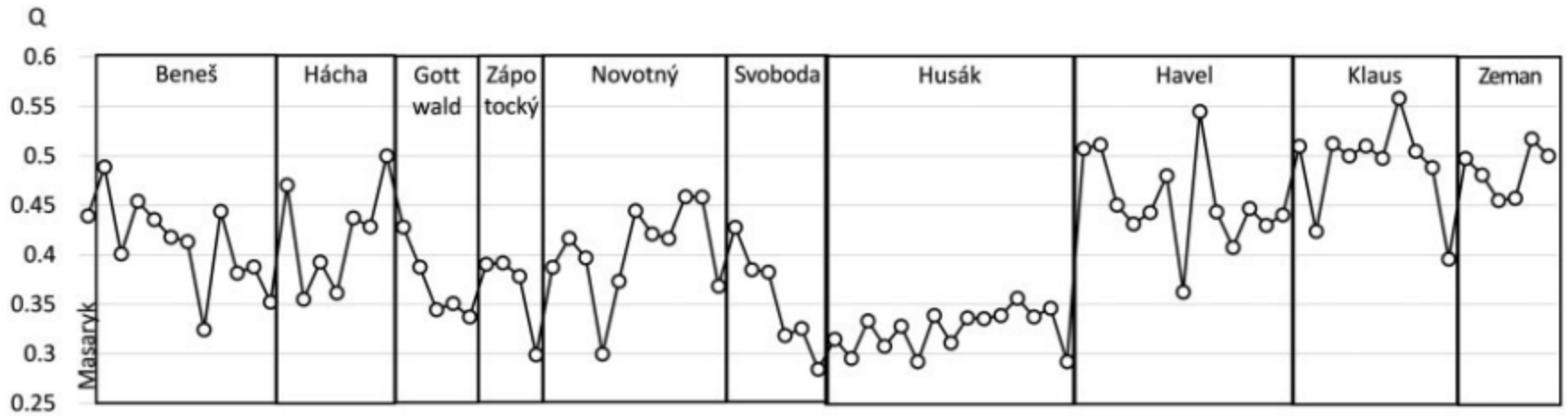
Aktivita / deskriptivita textu

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

T1: $V = 6$, $A = 1$ Q1 = 6 / 7 = 0,86

T2: $V = 2$, $A = 4$ Q2 = 2 / 6 = 0,33

Aktivita / deskriptivita textu



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.“

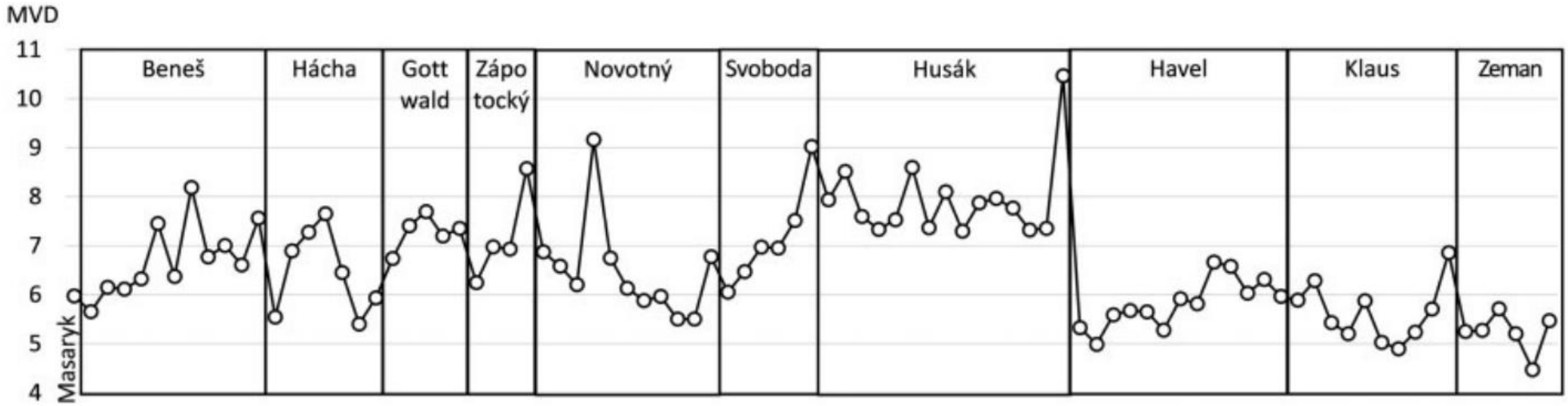
VD1 = (3 + 3 + 3) / 3 = 9 / 3 = 3

T2

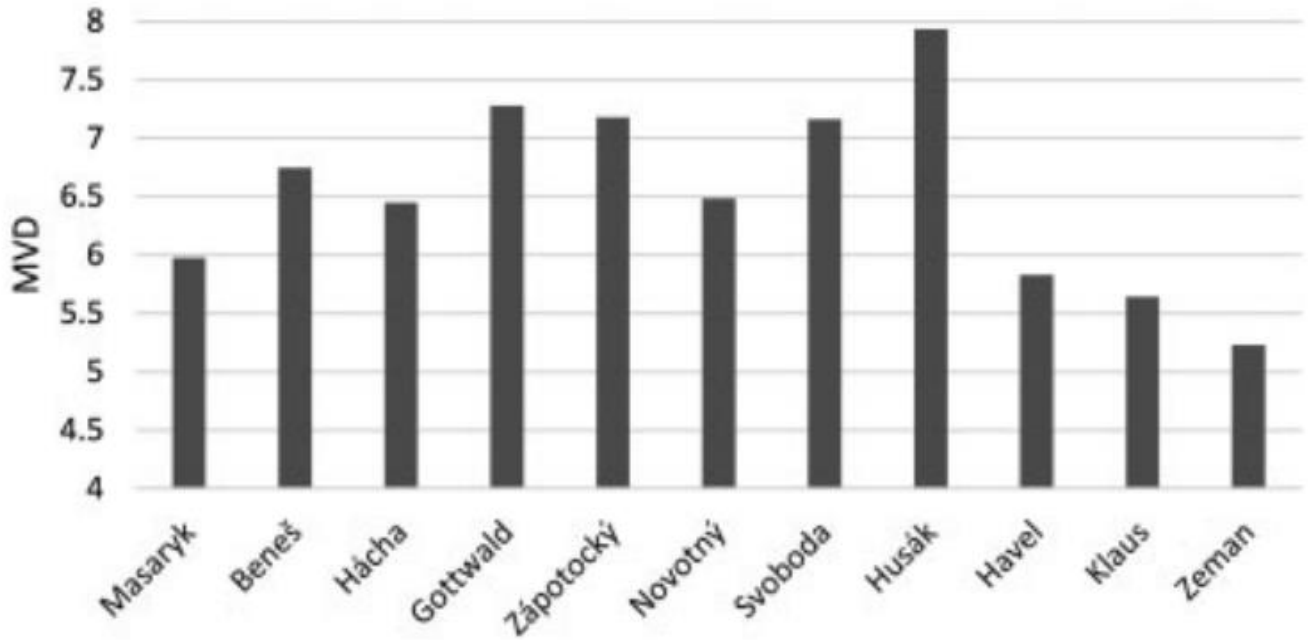
„Viděl dlouhé zelené stráně plné krásné zvěře, která se téměř nehýbala.“

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

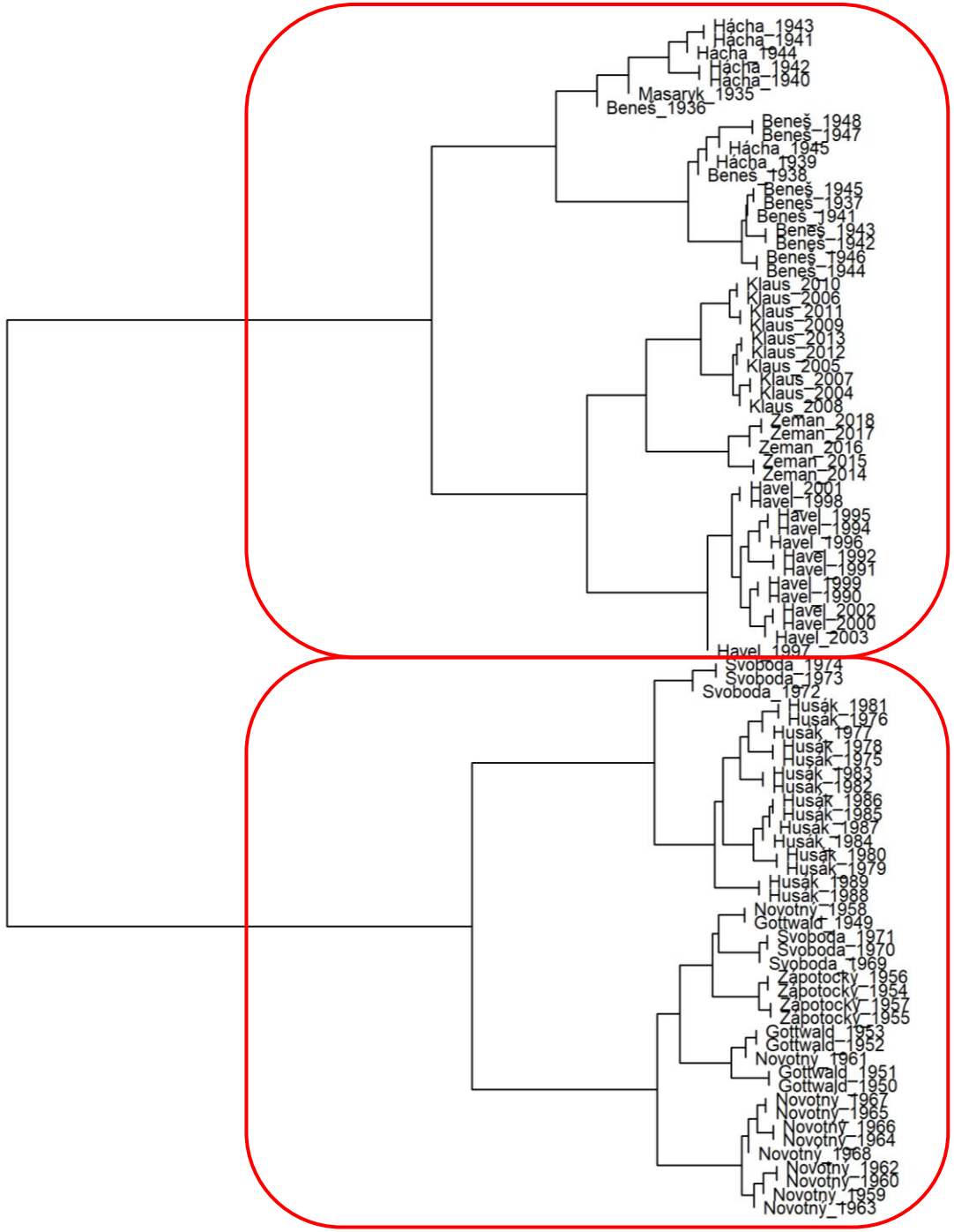
A, B … texty

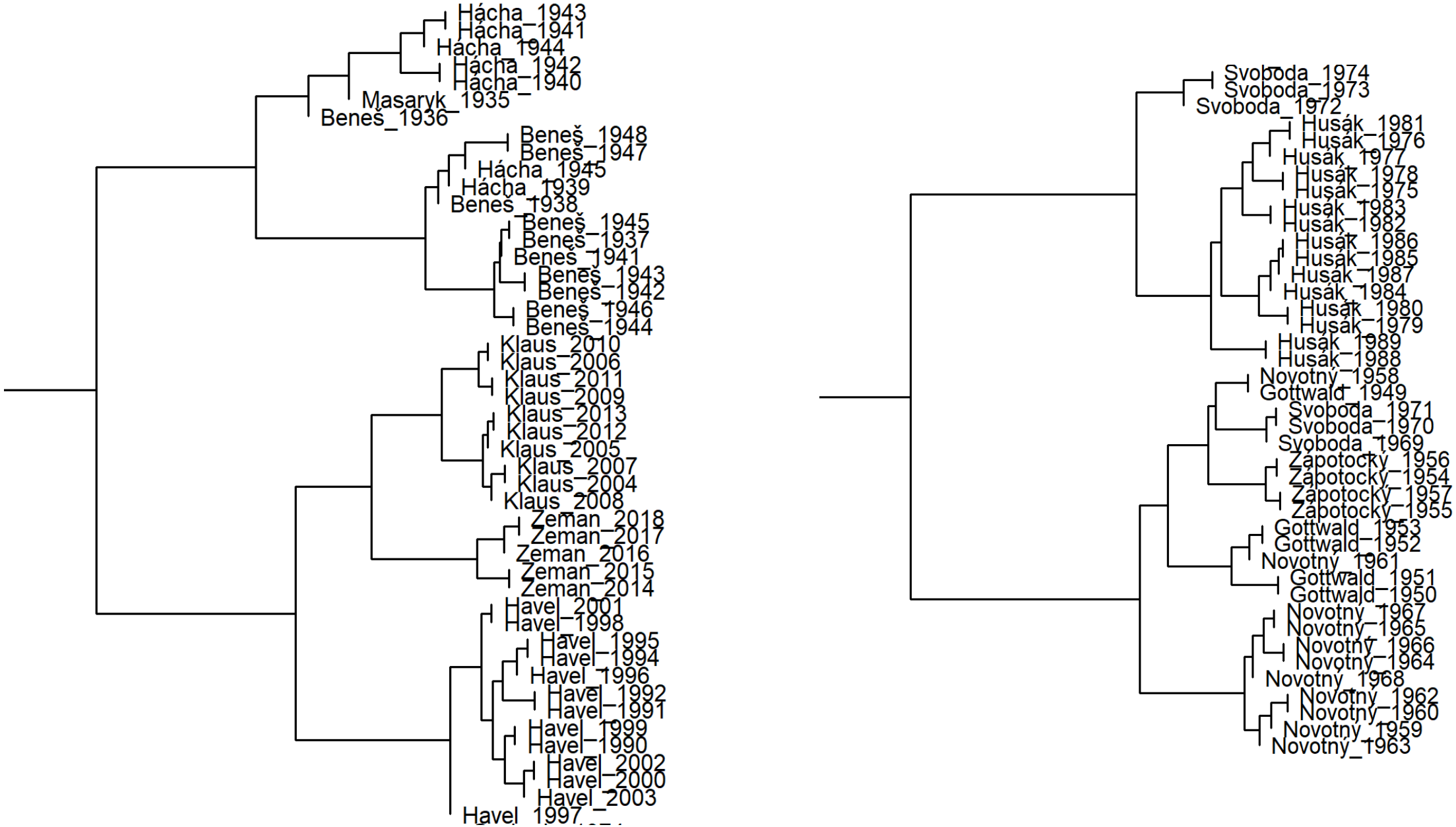
A_i … frekvence daného slova v textu A

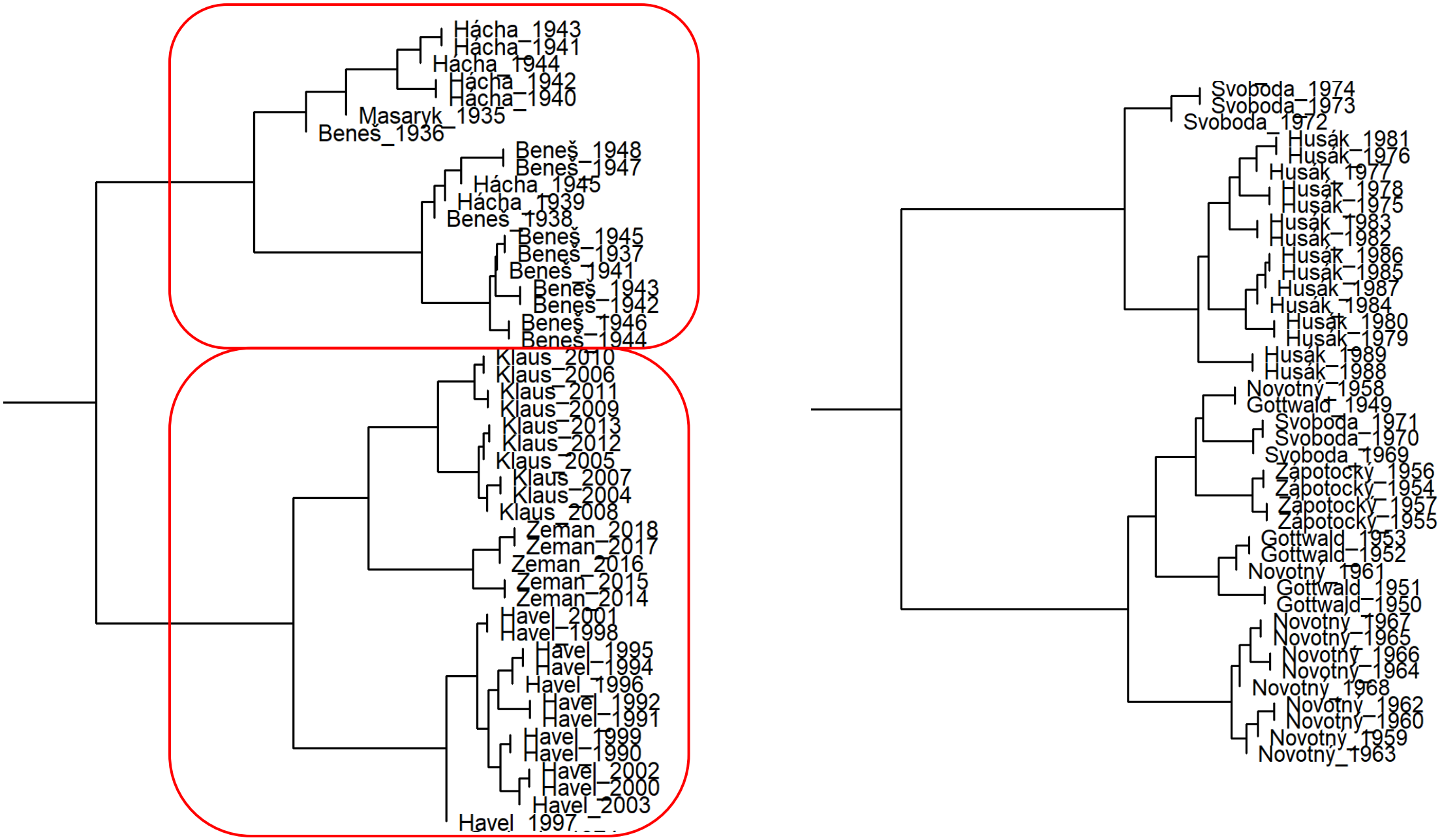
B_i … frekvence daného slova v textu B

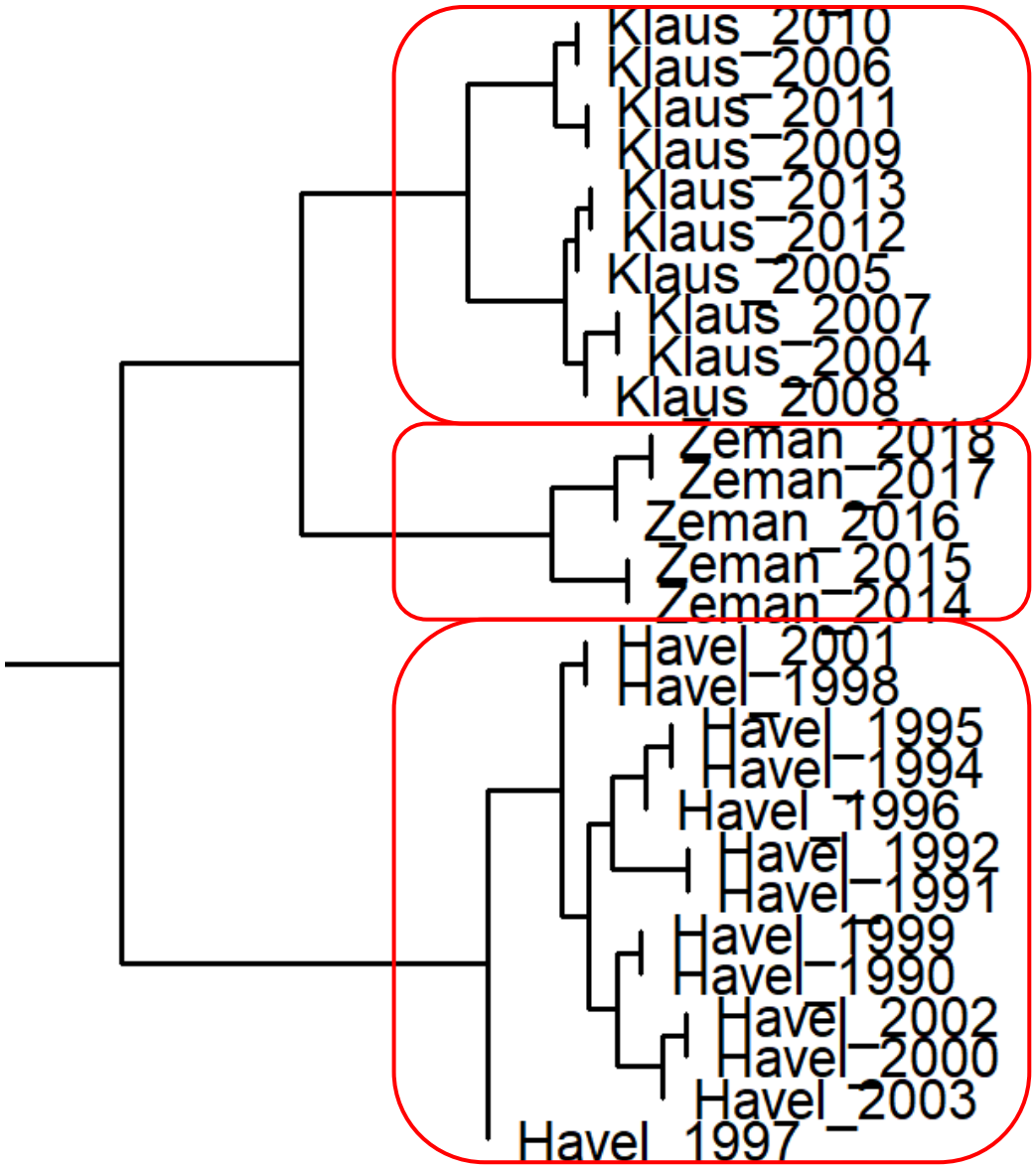
μ_i … průměrná frekvence daného slova ve všech textech (v korpusu)

σ_i … směrodatná odchylka frekvence daného slova ze všech textů


Hácha_1943
Hácha_1941
Hácha_1944
Hácha_1942
Hácha_1940
Masaryk_1935
Beneš_1936
Beneš_1948
Beneš_1947
Hácha_1945
Hácha_1939
Beneš_1938
Beneš_1945
Beneš_1937
Beneš_1941
Beneš_1943
Beneš_1942
Beneš_1946
Beneš_1944
Klaus_2010
Klaus_2006
Klaus_2011
Klaus_2009
Klaus_2013
Klaus_2012
Klaus_2005
Klaus_2007
Klaus_2004
Klaus_2008
Zeman_2018
Zeman_2017
Zeman_2016
Zeman_2015
Zeman_2014
Havel_2001
Havel_1998
Havel_1995
Havel_1994
Havel_1996
Havel_1992
Havel_1991
Havel_1999
Havel_1990
Havel_2002
Havel_2000
Havel_2003
Havel_1997
Svoboda_1974
Svoboda_1973
Svoboda_1972
Husák_1981
Husák_1976
Husák_1977
Husák_1978
Husák_1975
Husák_1983
Husák_1982
Husák_1986
Husák_1985
Husák_1987
Husák_1984
Husák_1980
Husák_1979
Husák_1989
Husák_1988
Novotný_1958
Gottwald_1949
Svoboda_1971
Svoboda_1970
Svoboda_1969
Zápotocký_1956
Zápotocký_1954
Zápotocký_1957
Zápotocký_1955
Gottwald_1953
Gottwald_1952
Novotný_1961
Gottwald_1951
Gottwald_1950
Novotný_1967
Novotný_1965
Novotný_1966
Novotný_1964
Novotný_1968
Novotný_1962
Novotný_1960
Novotný_1959
Novotný_1963


Hácha_1943
Hácha_1941
Hácha_1944
Hácha_1942
Hácha_1940
Masaryk_1935
Beneš_1936
Beneš_1948
Beneš_1947
Hácha_1945
Hácha_1939
Beneš_1938
Beneš_1945
Beneš_1937
Beneš_1941
Beneš_1943
Beneš_1942
Beneš_1946
Beneš_1944
Klaus_2010
Klaus_2006
Klaus_2011
Klaus_2009
Klaus_2013
Klaus_2012
Klaus_2005
Klaus_2007
Klaus_2004
Klaus_2008
Zeman_2018
Zeman_2017
Zeman_2016
Zeman_2015
Zeman_2014
Havel_2001
Havel_1998
Havel_1995
Havel_1994
Havel_1996
Havel_1992
Havel_1991
Havel_1999
Havel_1990
Havel_2002
Havel_2000
Havel_2003
Havel_1997
Svoboda_1974
Svoboda_1973
Svoboda_1972
Husák_1981
Husák_1976
Husák_1977
Husák_1978
Husák_1975
Husák_1983
Husák_1982
Husák_1986
Husák_1985
Husák_1987
Husák_1984
Husák_1980
Husák_1979
Husák_1989
Husák_1988
Novotný_1958
Gottwald_1949
Svoboda_1971
Svoboda_1970
Svoboda_1969
Zápotocký_1956
Zápotocký_1954
Zápotocký_1957
Zápotocký_1955
Gottwald_1953
Gottwald_1952
Novotný_1961
Gottwald_1951
Gottwald_1950
Novotný_1967
Novotný_1965
Novotný_1966
Novotný_1964
Novotný_1968
Novotný_1962
Novotný_1960
Novotný_1959
Novotný_1963


Hácha_1943
Hácha_1941
Hácha_1944
Hácha_1942
Hácha_1940
Masaryk_1935
Beneš_1936
Beneš_1948
Beneš_1947
Hácha_1945
Hácha_1939
Beneš_1938
Beneš_1945
Beneš_1937
Beneš_1941
Beneš_1943
Beneš_1942
Beneš_1946
Beneš_1944
Klaus_2010
Klaus_2006
Klaus_2011
Klaus_2009
Klaus_2013
Klaus_2012
Klaus_2005
Klaus_2007
Klaus_2004
Klaus_2008
Zeman_2018
Zeman_2017
Zeman_2016
Zeman_2015
Zeman_2014
Havel_2001
Havel_1998
Havel_1995
Havel_1994
Havel_1996
Havel_1992
Havel_1991
Havel_1999
Havel_1990
Havel_2002
Havel_2000
Havel_2003
Havel_1997
Svoboda_1974
Svoboda_1973
Svoboda_1972
Husák_1981
Husák_1976
Husák_1977
Husák_1978
Husák_1975
Husák_1983
Husák_1982
Husák_1986
Husák_1985
Husák_1987
Husák_1984
Husák_1980
Husák_1979
Husák_1989
Husák_1988
Novotný_1958
Gottwald_1949
Svoboda_1971
Svoboda_1970
Svoboda_1969
Zápotocký_1956
Zápotocký_1954
Zápotocký_1957
Zápotocký_1955
Gottwald_1953
Gottwald_1952
Novotný_1961
Gottwald_1951
Gottwald_1950
Novotný_1967
Novotný_1965
Novotný_1966
Novotný_1964
Novotný_1968
Novotný_1962
Novotný_1960
Novotný_1959
Novotný_1963


Klaus_2010
Klaus_2006
Klaus_2011
Klaus_2009
Klaus_2013
Klaus_2012
Klaus_2005
Klaus_2007
Klaus_2004
Klaus_2008
Zeman_2018
Zeman_2017
Zeman_2016
Zeman_2015
Zeman_2014
Havel_2001
Havel_1998
Havel_1995
Havel_1994
Havel_1996
Havel_1992
Havel_1991
Havel_1999
Havel_1990
Havel_2002
Havel_2000
Havel_2003
Havel_1997

4. QuitaUp

- aneb jak jednoduše měřit vybrané charakteristiky textů…

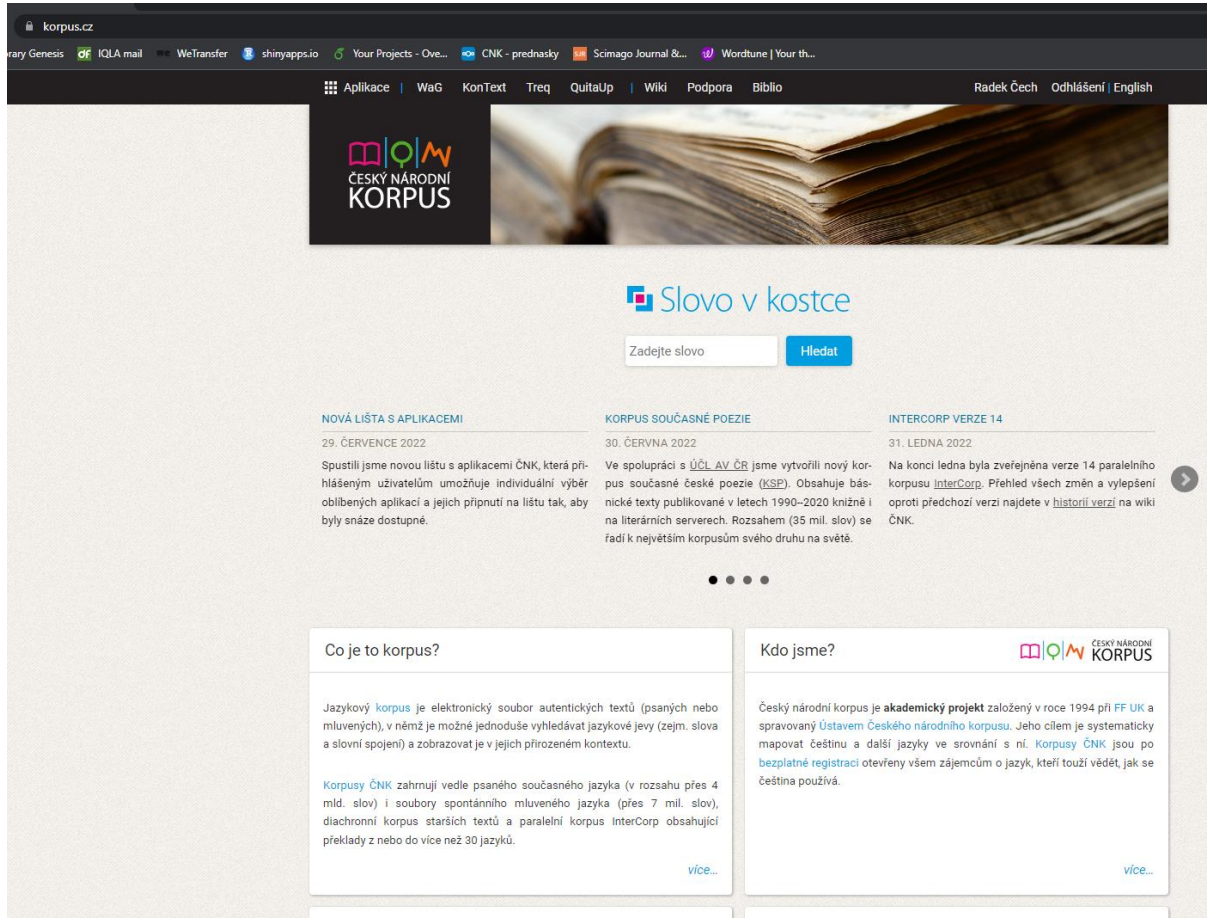
4. QuitaUp

Katedra českého jazyka

Filozofická fakulta Ostravské univerzity

4. QuitaUp

- https://www.korpus.cz/



4. QuitaUp

- https://korpus.cz/quitaup/

4. QuitaUp

- texty
  - http://www.zemanmilos.cz/cz/

  - projev prezidenta republiky při udílení státních vyznamenání
    - 2013
    - 2015
    - 2018
    - 2022

Děkuji za pozornost!

https://cechradek.cz/