

Stylistika V-VI

ZS 2024

Obsah

- tematická slova
- tematická koncentrace textu
 - možné vnímat jako projev aktualizace
- literatura
 - Čech, R. (2016). Tematická koncentrace textu v češtině. Praha: UFAL.
 - https://www.cechradek.cz/publ/2016_Cech_TK_v_cestine.pdf

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)

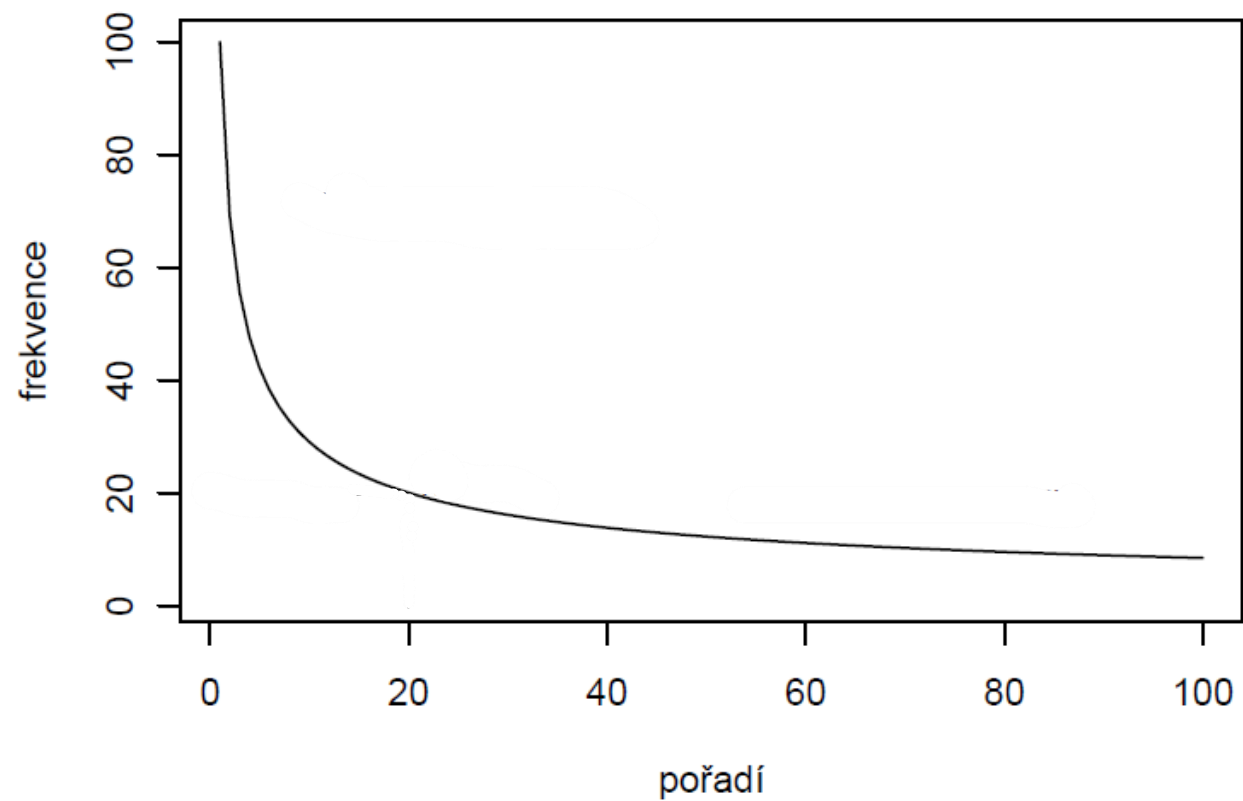
- v různých textech se autor na dané téma či témata může zaměřovat s různou intenzitou
- lze identifikovat jazykové jednotky, které je možné chápat jako nositele určitého tématu či témat
- míru zaměření se na dané téma či témata lze detekovat analýzou frekvenčních charakteristik textu

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. přepokládá se její systematické chování vzhledem jak k jiným vlastnostem textu, tak k faktorům pragmatickým

Frekvenční struktura textu

- uspořádání slov podle frekvence

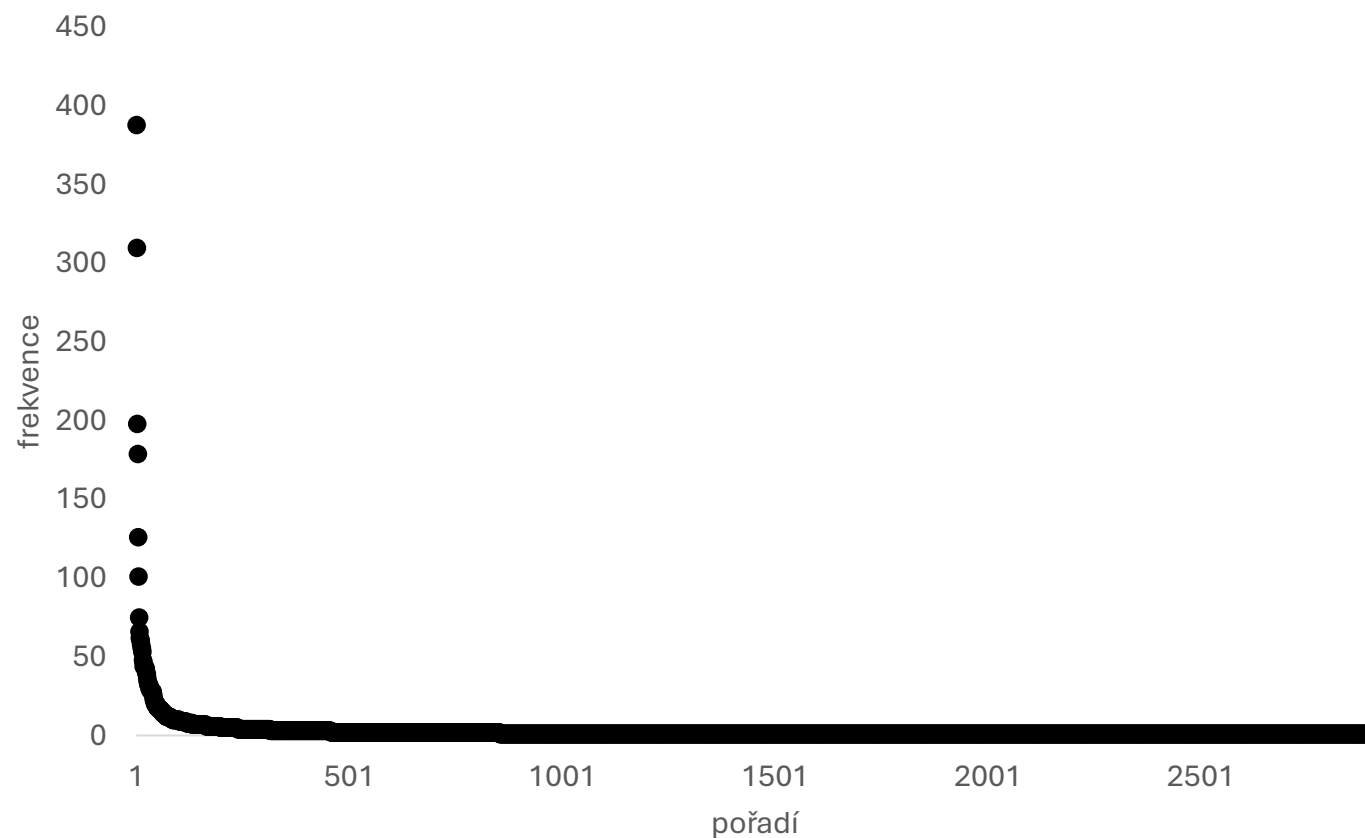


Frekvenční struktura textu

- J. Škvorecký: Neurčité kontury
- Word Frequency Counter
 - <https://ezcalc.me/word-frequency-counter/>

Frekvenční struktura textu

- J. Škvorecký: Neurčité kontury



Frekvenční struktura textu

- J. Skácel: Odvaha k tomu

Odvaha k tomu

Lhal jsem a říkal, že tam mrtvý není.

Tak pozdě v noci – nebude nikdo na ty housle hrát.

A byl jsem zděšený a prázdný

jak v zimě sad.

A byl tam. Docela tam byl.

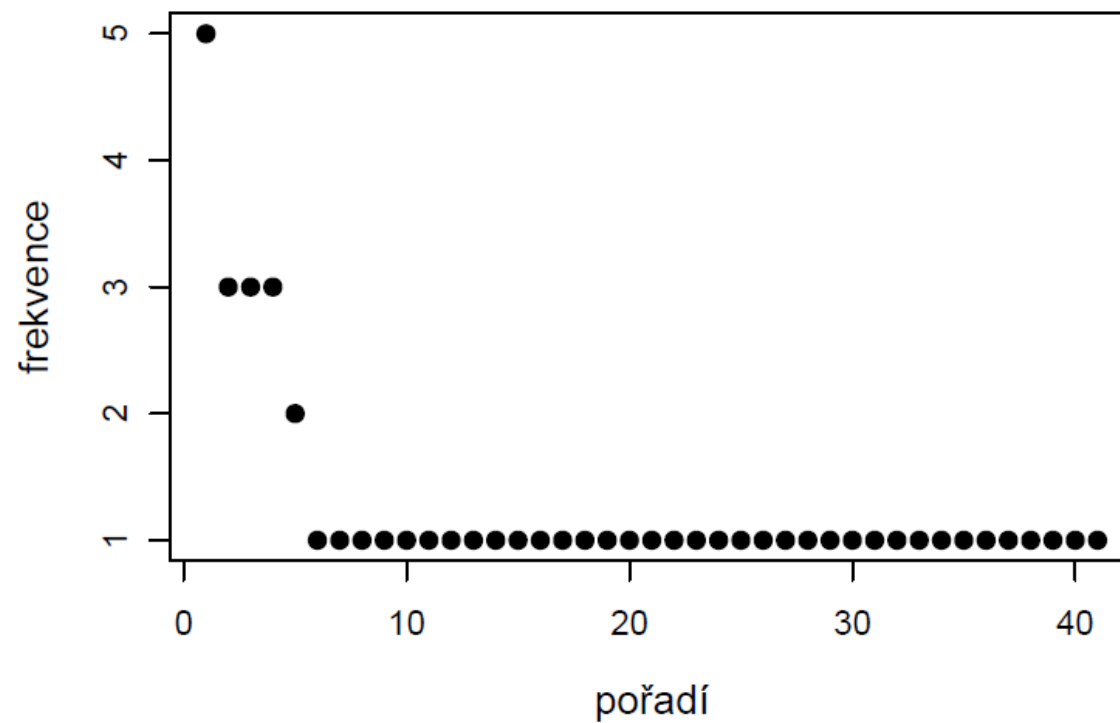
Tím rovným dílem ticha jsme to znali.

Dovedli v stráni ukřižovat les –

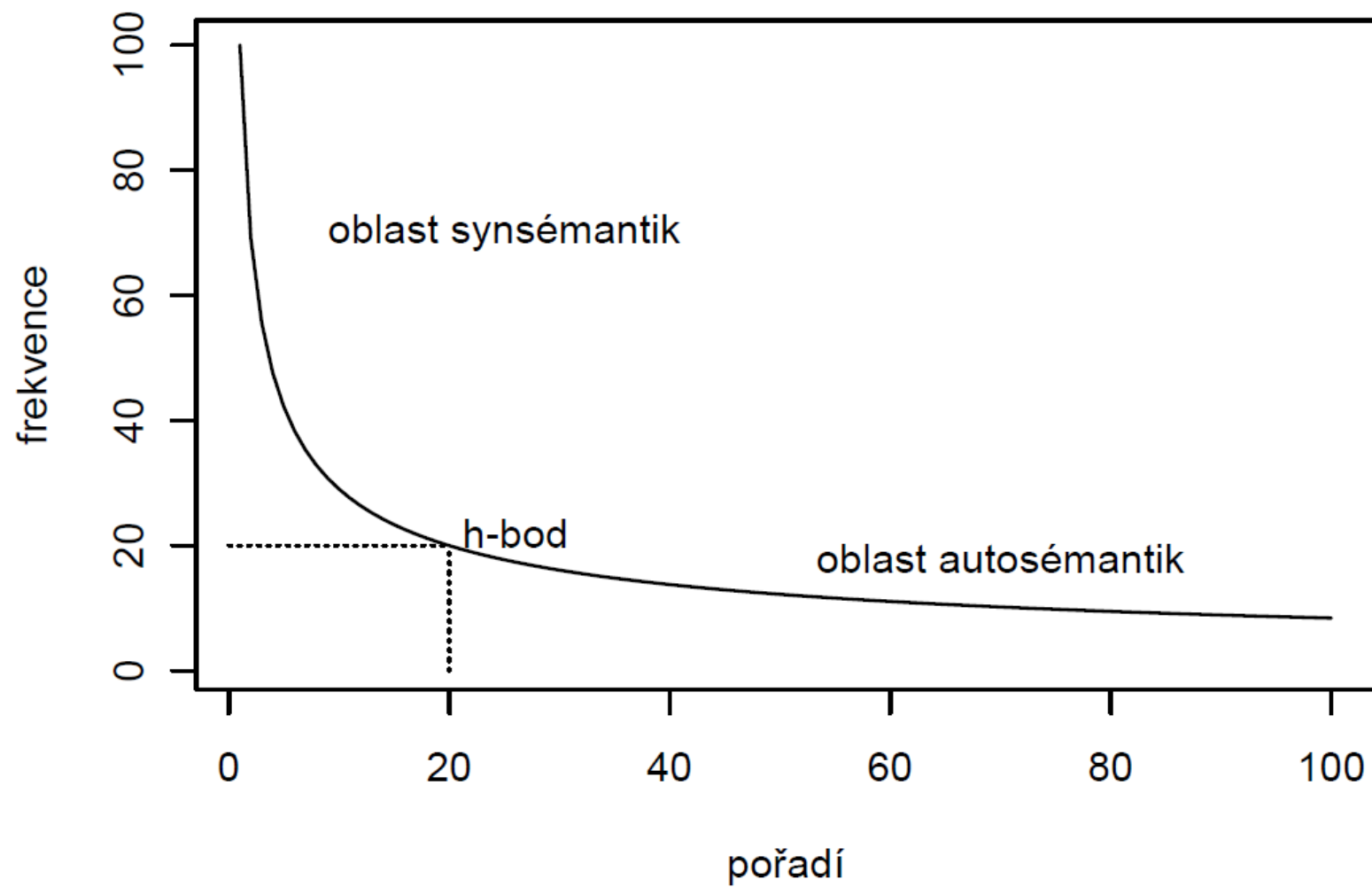
a vodu ukamenovali.

Frekvenční struktura 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, odpust')
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, odpust')
zpívá si na poli smuténka
a chodí
po konopných cestách podzimu.*

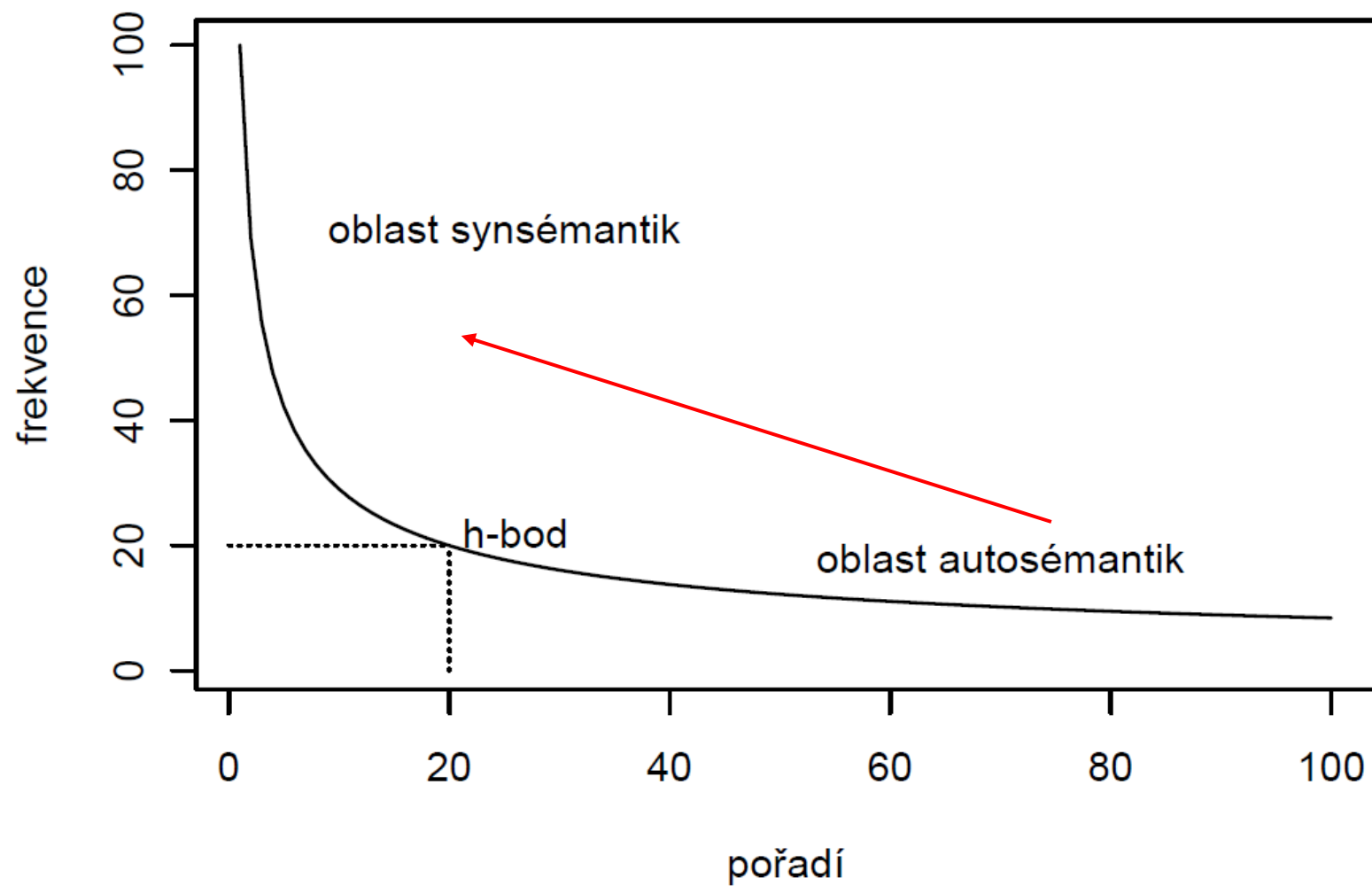
pořadí	slovní tvar	frekvence
1	<i>smuténka</i>	4
2	<i>a</i>	3
3	<i>chodí</i>	3
4	<i>zpívá</i>	3
5	<i>poli</i>	2
6	<i>po</i>	2
7	<i>než</i>	2
8	<i>starší</i>	1
9	<i>příběh</i>	1
(...)	(...)	(...)
39	<i>září</i>	1

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



H-bod

- pořadí slova r se rovná jeho frekvenci $f(r)$

$$r = f(r)$$

H-bod

- pořadí slova r se rovná jeho frekvenci $f(r)$

$$r = f(r)$$

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

H-bod

- pokud $r \neq f(r)$, pak

$$h = \frac{f(i)j - f(j)i}{j - i + f(i) - f(j)}$$

kde i a j jsou pořadí slov a $f(i)$ a $f(j)$ jsou jejich frekvence, přičemž $i < j$, kde i je největší takové číslo, pro které $i < f(i)$, a j je nejmenší takové číslo, pro které $j > f(j)$.

H-bod

- pokud $r \neq f(r)$, pak

$$h = \frac{f(i)j - f(j)i}{j - i + f(i) - f(j)}$$

kde i a j jsou pořadí slov a $f(i)$ a $f(j)$ jsou jejich frekvence, přičemž $i < j$, kde i je největší takové číslo, pro které $i < f(i)$, a j je nejmenší takové číslo, pro které $j > f(j)$.

pořadí	slovní tvar	frekvence
1	<i>a</i>	7
2	<i>s</i>	4
3	<i>na</i>	2
4	<i>za</i>	2
5	<i>lod'</i>	2
6	<i>pojedu</i>	2
7	<i>kávu</i>	2

Pořadí

- slova se stejnou frekvencí → průměrné pořadí

pořadí	slovní tvar	frekvence
1	<i>že</i>	14
2	<i>a</i>	12
3	<i>se</i>	11
4	<i>Berdych</i>	8
5	<i>za</i>	8
6	<i>z</i>	7
7	<i>na</i>	7
8	<i>trestu</i>	7
9	<i>je</i>	6
10	<i>v</i>	6
(...)	(...)	(...)
373	<i>neuvěřil</i>	1

Pořadí

- slova se stejnou frekvencí → průměrné pořadí

pořadí	slovní tvar	frekvence
1	<i>že</i>	14
2	<i>a</i>	12
3	<i>se</i>	11
4	<i>Berdych</i>	8
5	<i>za</i>	8
6	<i>z</i>	7
7	<i>na</i>	7
8	<i>trestu</i>	7
9	<i>je</i>	6
10	<i>v</i>	6
(...)	(...)	(...)
373	<i>neuvěřil</i>	1

Pořadí

- slova se stejnou frekvencí → průměrné pořadí

$$\bar{r}_{\text{slovo}} = \frac{\sum r_{f_i}}{n_{f_i}}$$

$$\bar{r}_{\text{Berdych}} = \frac{r_{\text{Berdych}} + r_{\text{za}}}{2} = \frac{4 + 5}{2} = 4,5$$

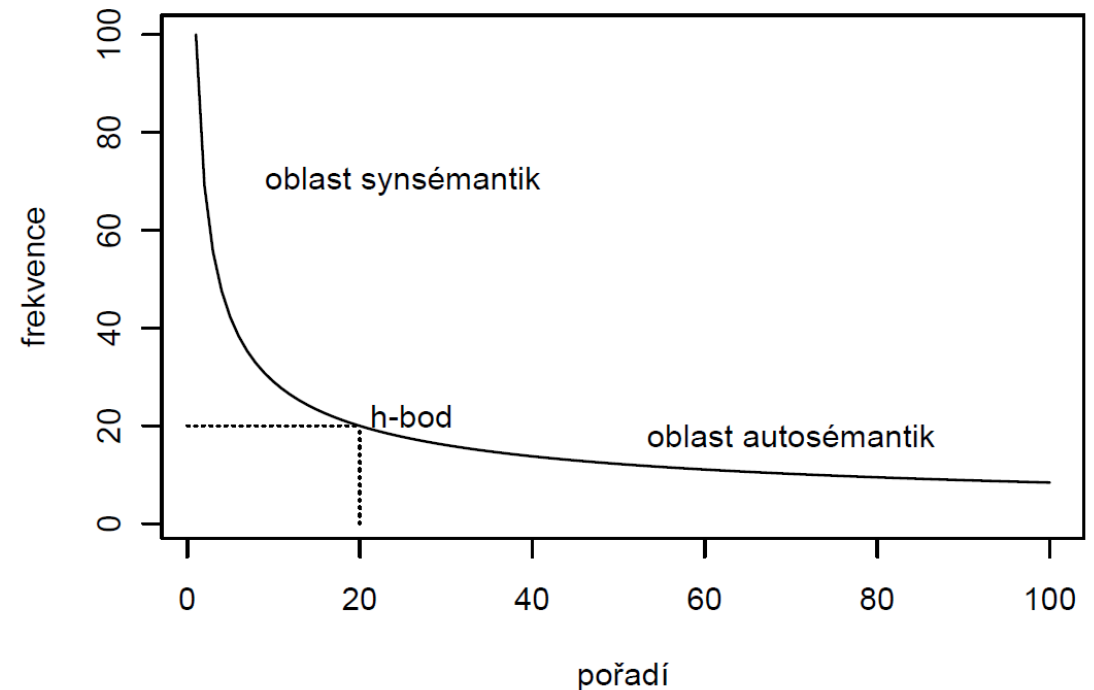
$$\bar{r}_{\text{trestu}} = \frac{r_z + r_{na} + r_{trestu}}{3} = \frac{6 + 7 + 8}{3} = 7$$

Tematická váha slova

- vzdálenost od h-bodu: $(h - r')$
- frekvence: $f(r')$
- normalizace: $h(h - 1) \cdot f(1)$

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

- srov. strany 16nn

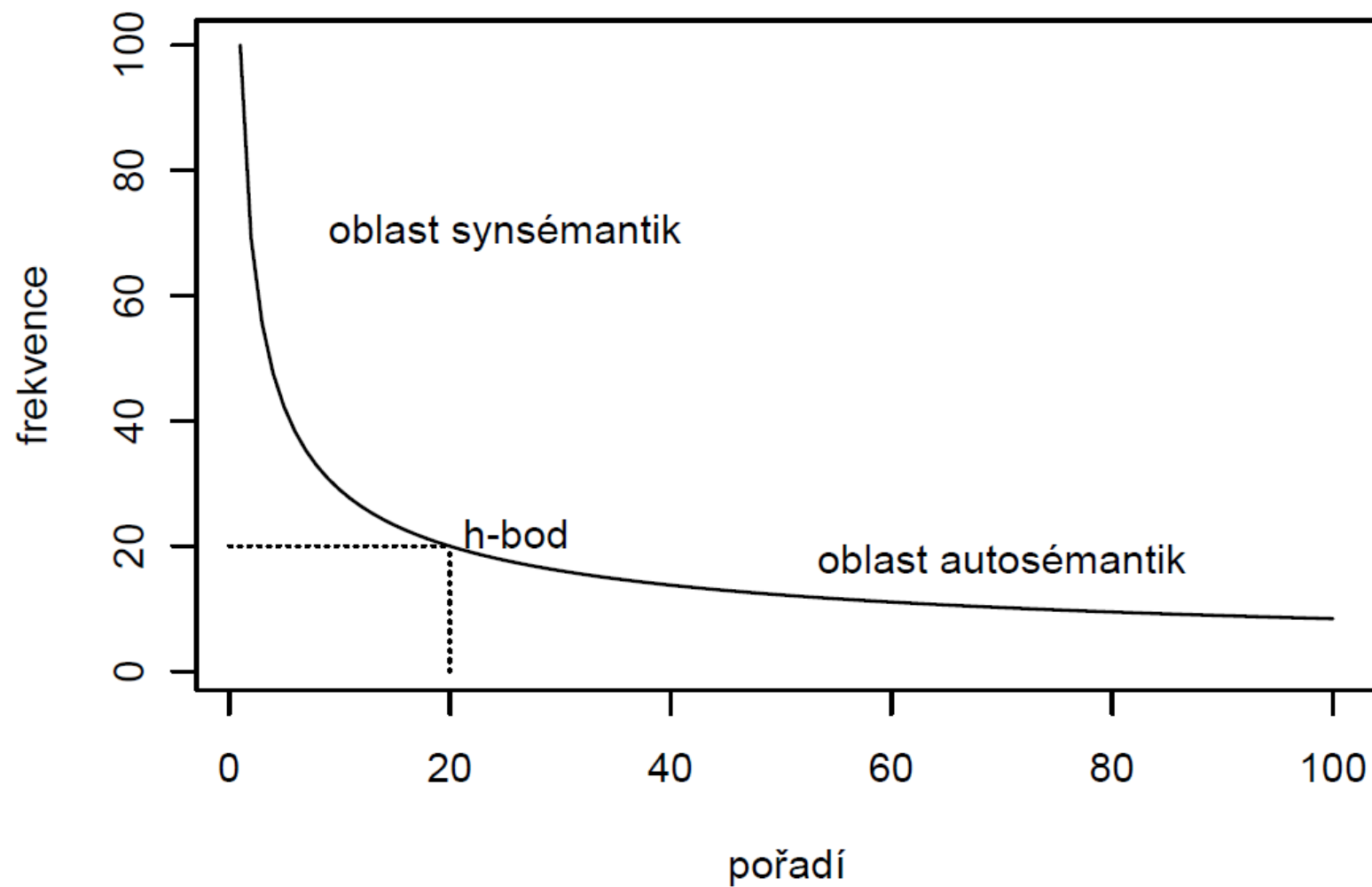


Tematická koncentrace textu

- součet tematických vah jednotlivých slov

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

Frekvenční struktura textu – „neutrální“ text

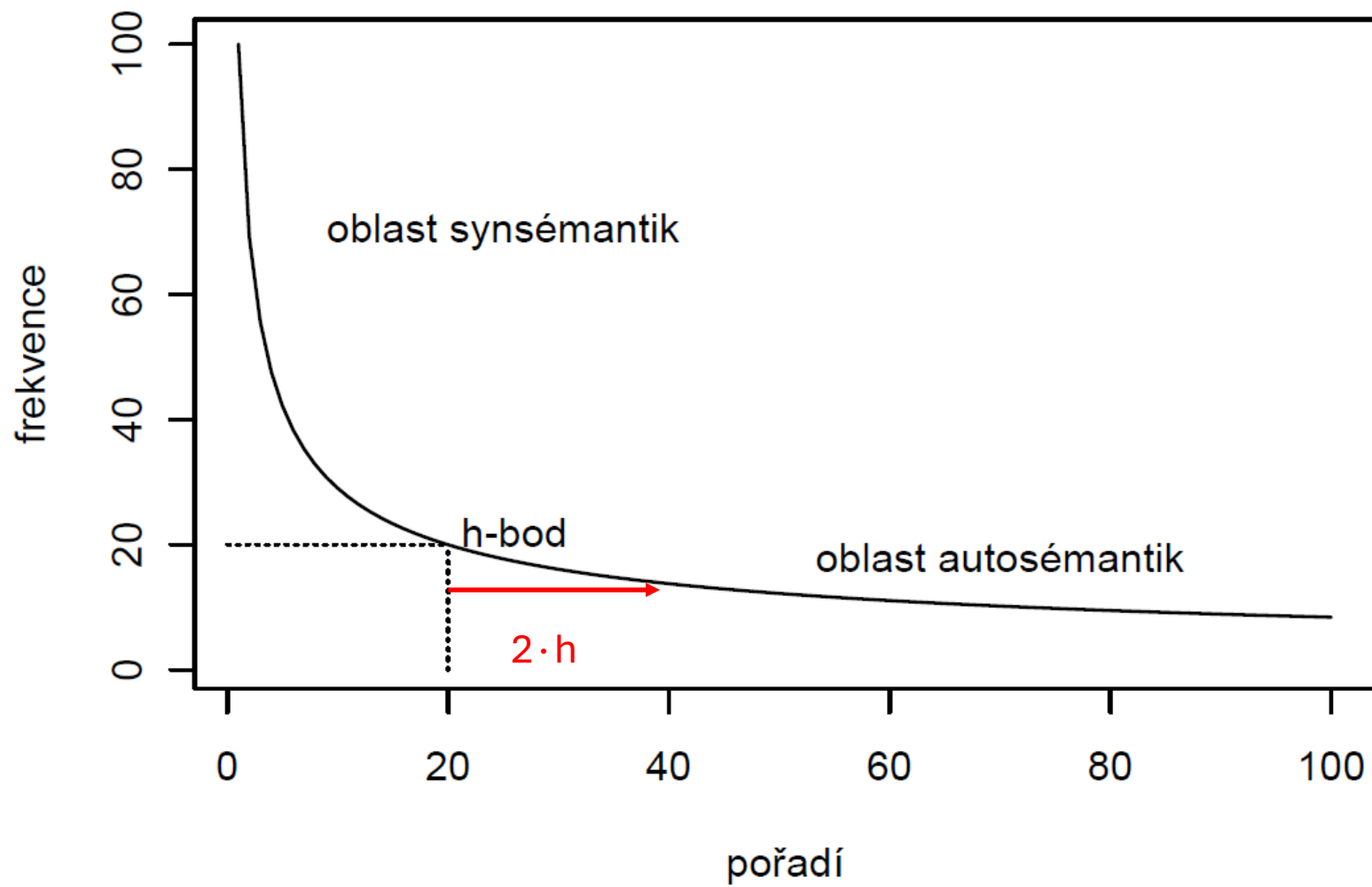


Sekundární tematická koncentrace textu

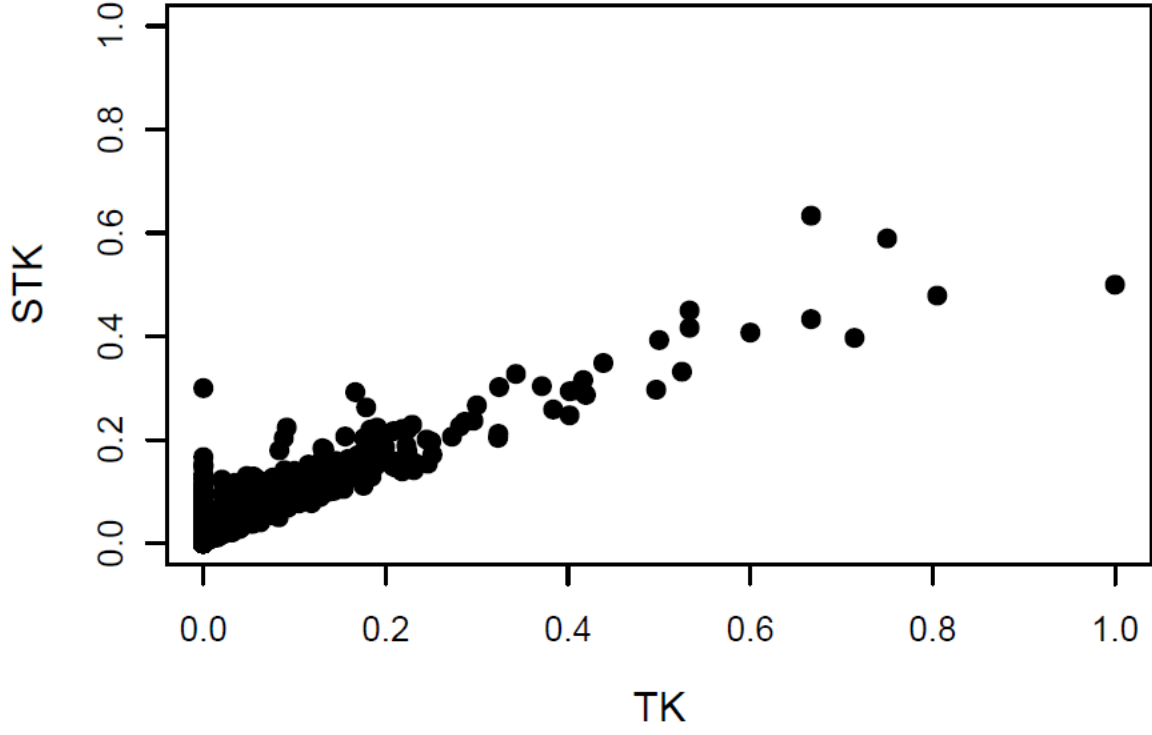
- h-bod se násobí 2

$$STK = \sum_{j=1}^T \frac{(2h - r'_{(j)})f(r'_{(j)})}{h(2h - 1)f(1)}$$

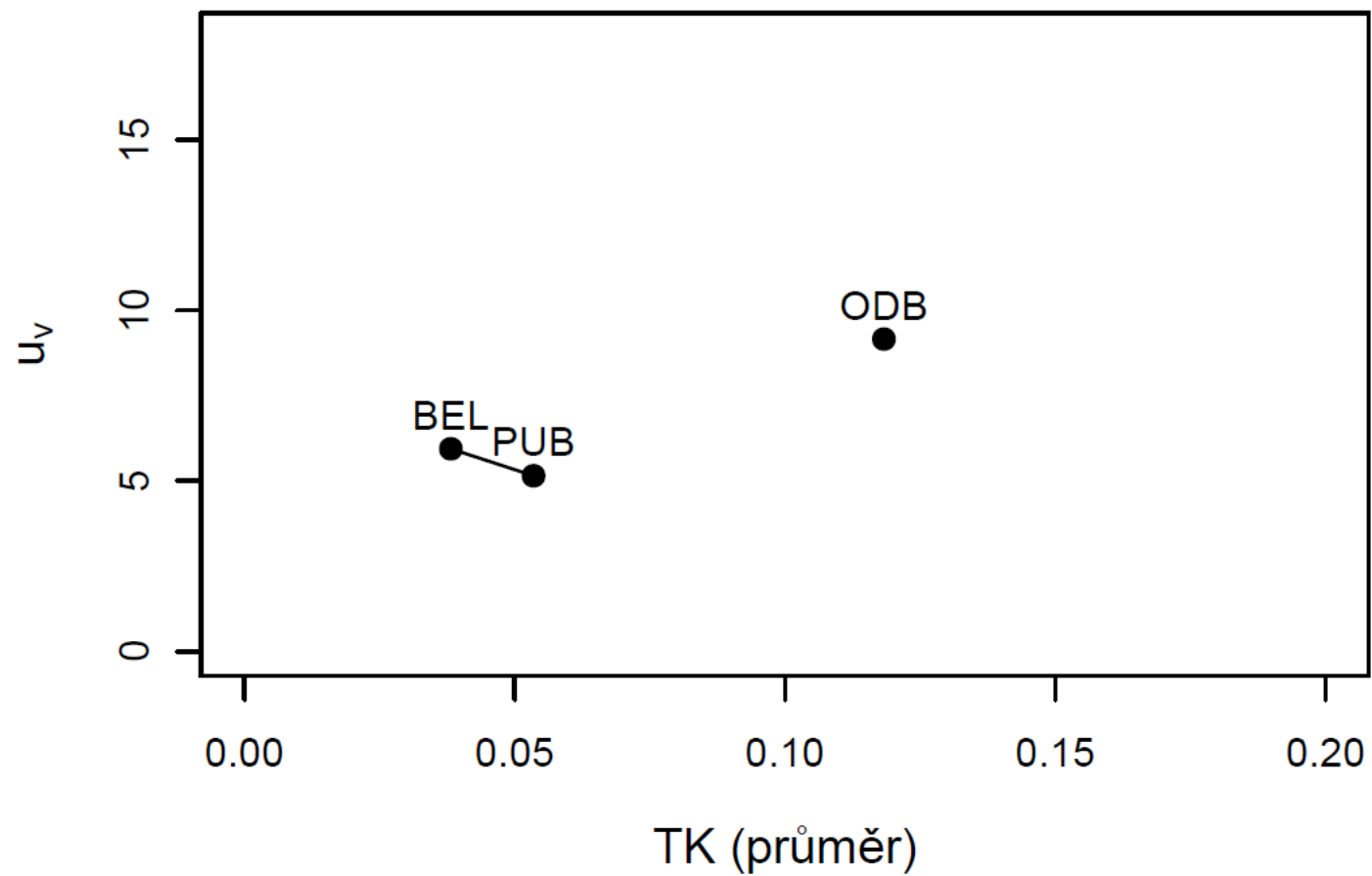
Frekvenční struktura textu



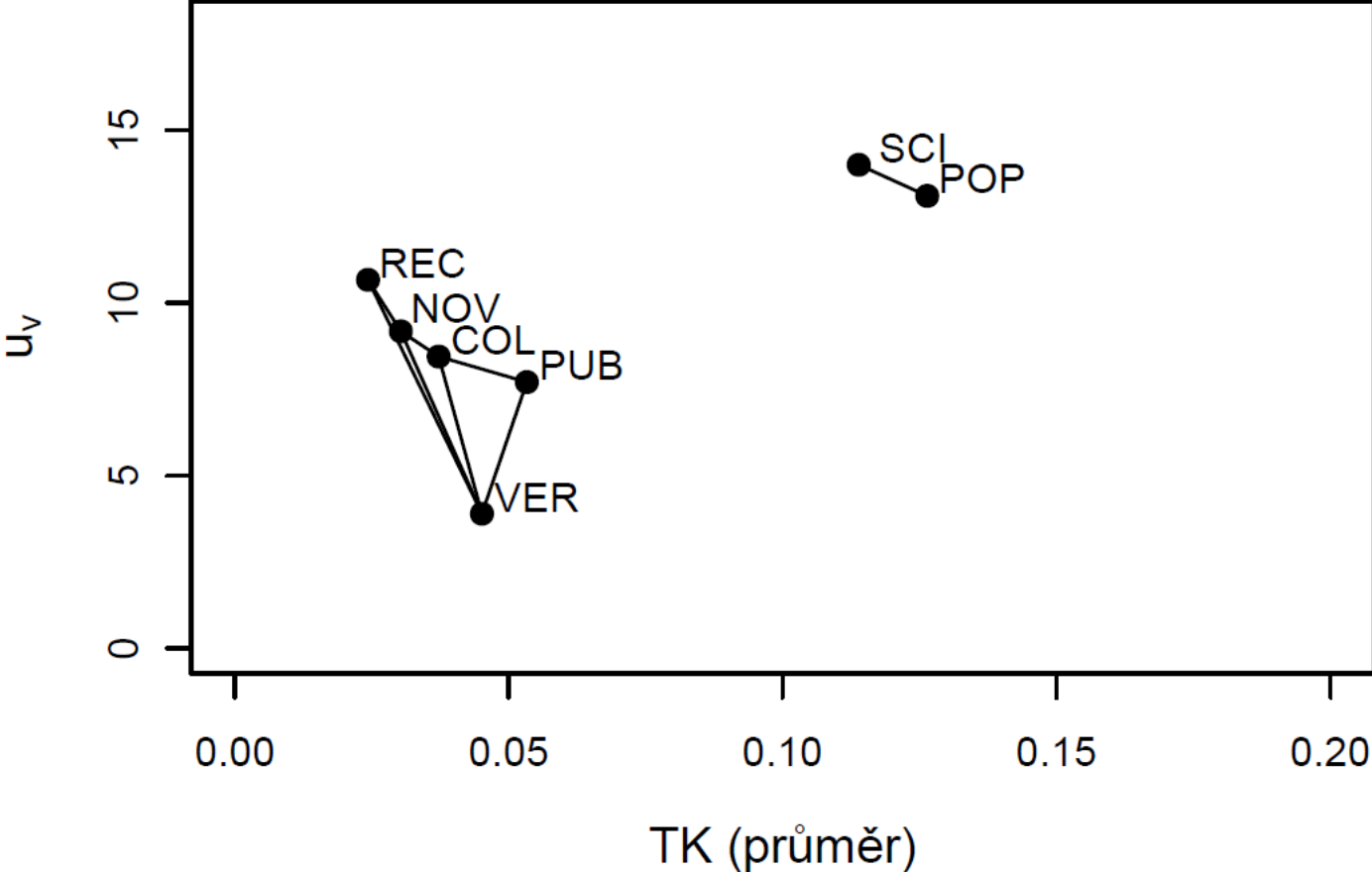
STK & TK



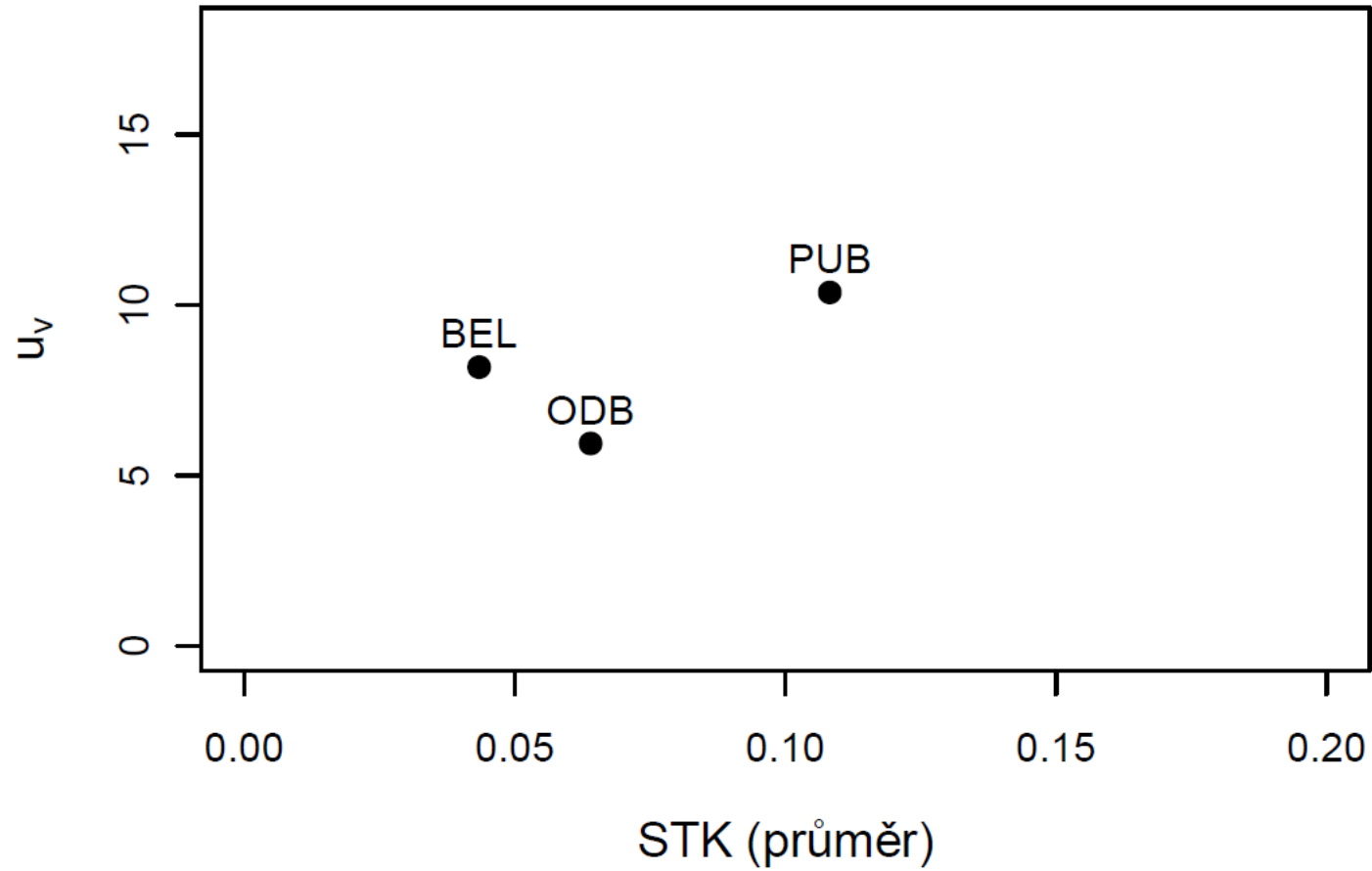
TK a klasifikace textů



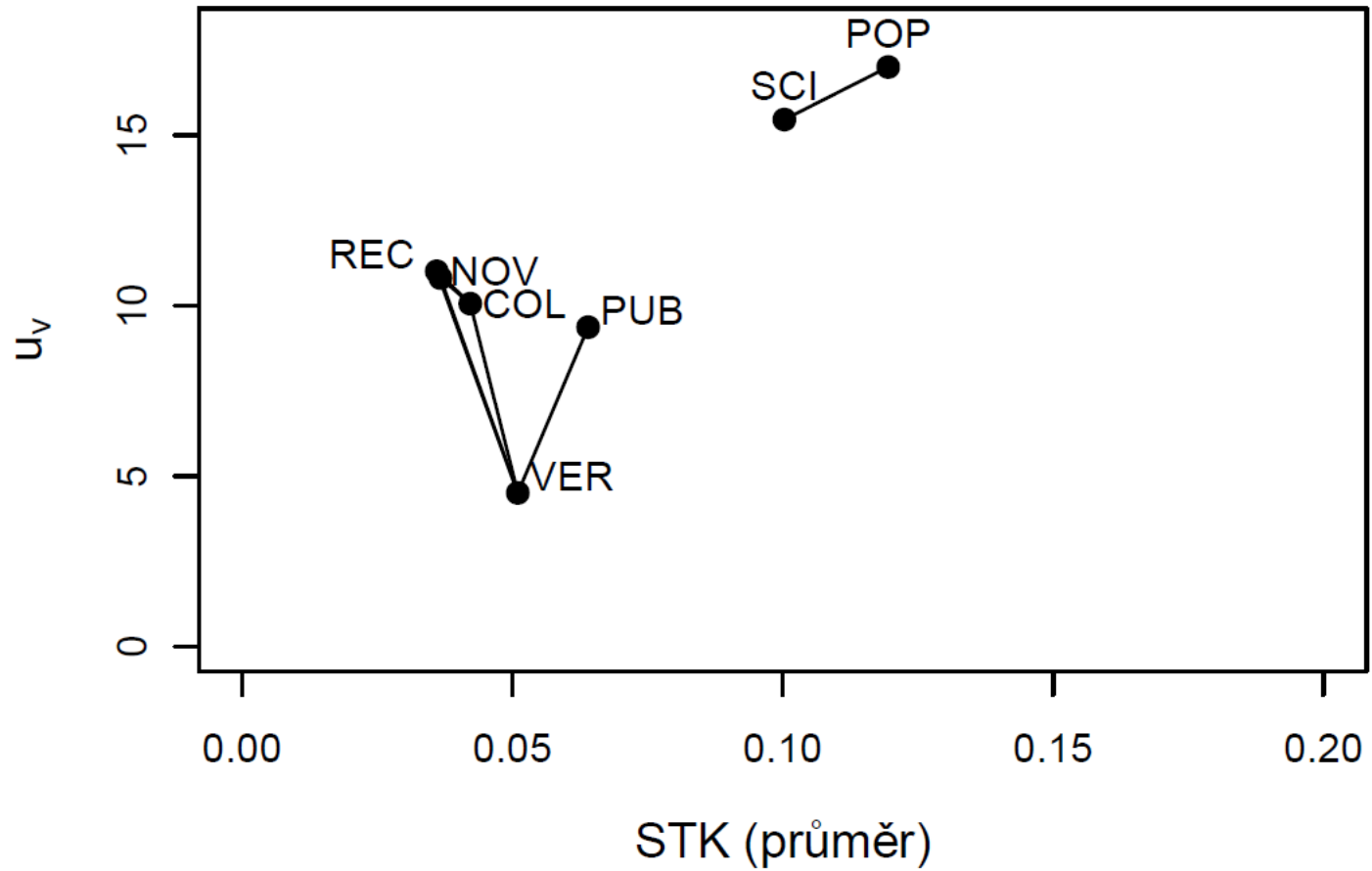
TK a klasifikace textů



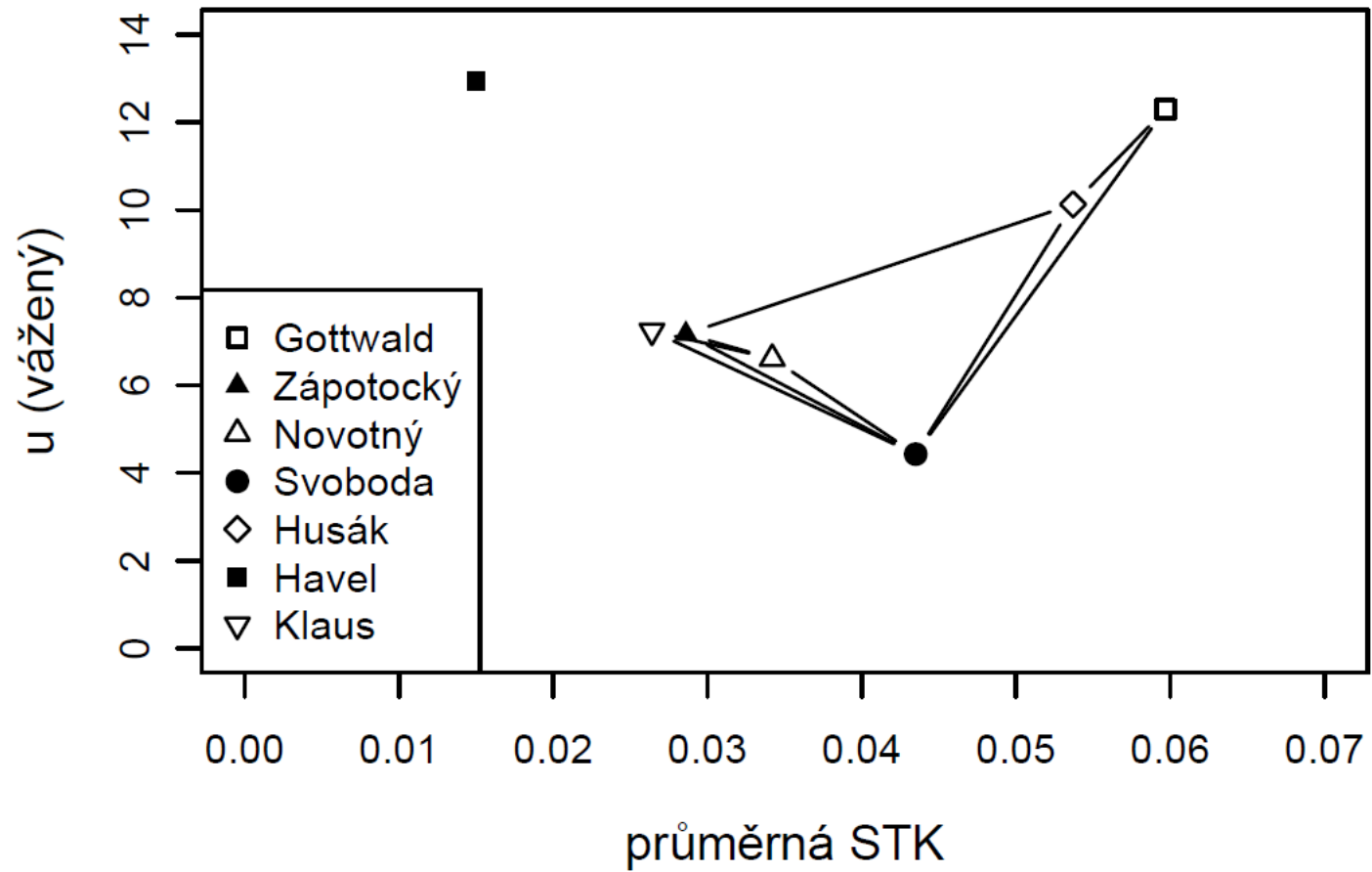
STK a klasifikace textů



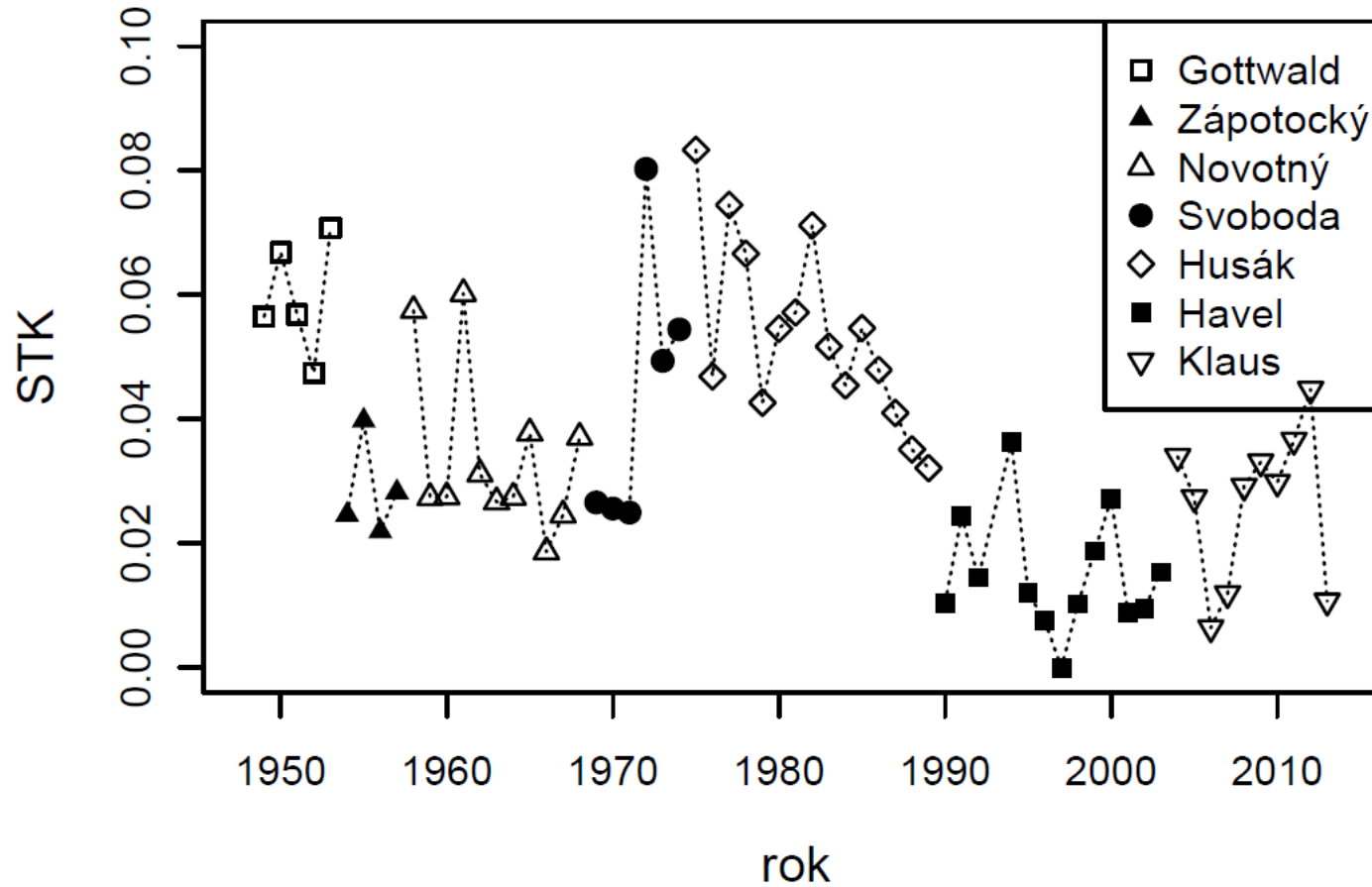
STK a klasifikace textů



STK a klasifikace textů



STK a klasifikace textů



Vývoj tematické koncentrace v textu

- s. 73nn