

Stylistika V

ZS 2024

Obsah

- „globální“ vlastnosti textu
- literatura
 - Kubát, M. Kvantitativní analýza žánrů. 2016.
 - grafy a příklady převzaty z této publikace

Celkové vlastnosti textu

- 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.

Slovní bohatství / lexikální diverzita

- type-token poměr
- proporce hapax legomenon
- index opakování slov (repeat rate)
- entropie
- Giniho index
- ...

Poměr typů a tokenů

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

Poměr typů a tokenů

- Marie miluje Petra. Petra miluje taky Jana. Petr miluje Emu.
- tokeny = 10
- typy = ?

Poměr typů a tokenů

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

- tokeny = 10

- typy = 7

Poměr typů a tokenů

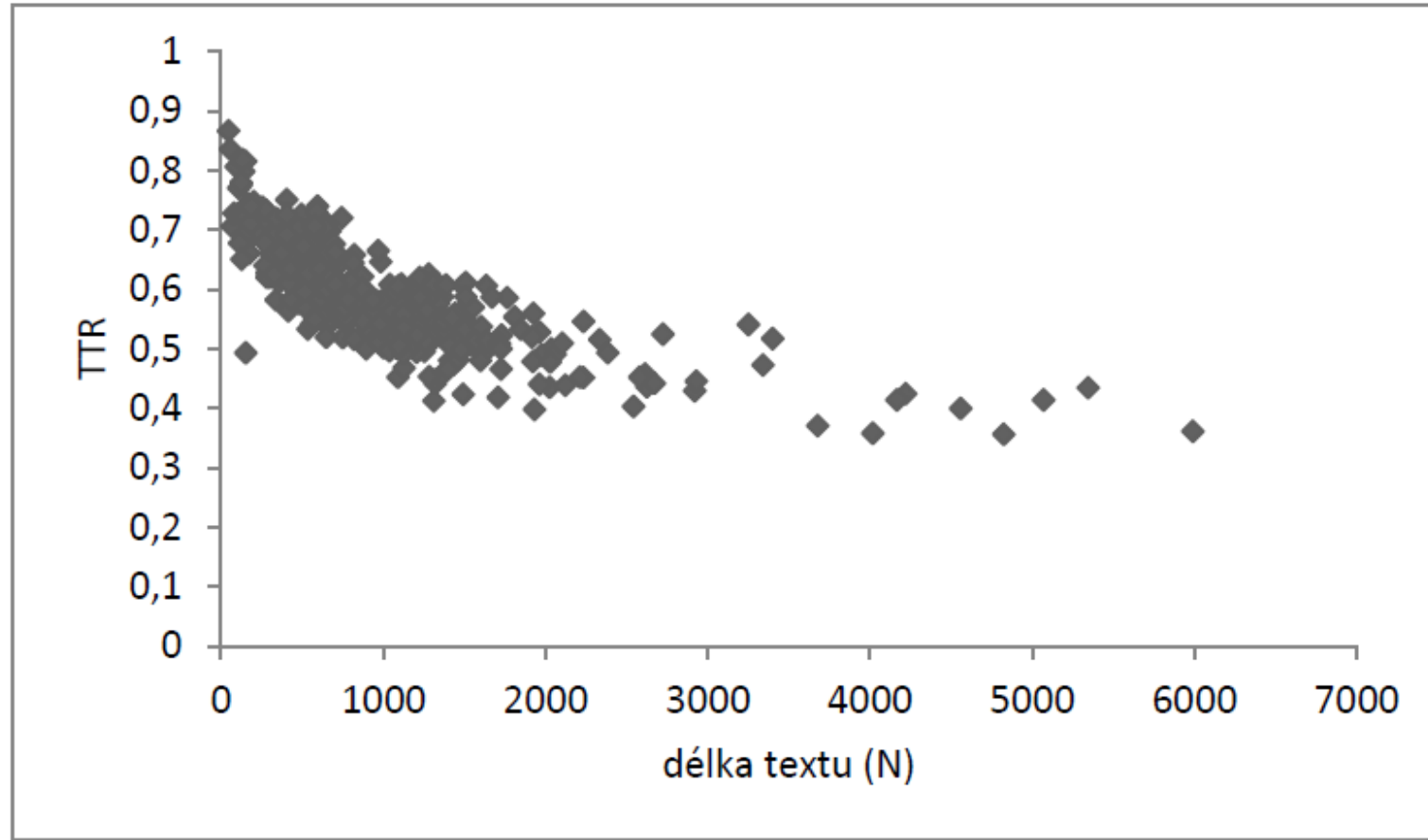
• **Marie miluje Petra. Petra miluje taky Jana. Petr miluje Emu.**

• tokeny = 10

• typy = 7

$$TTR = \frac{\textit{typy}}{\textit{tokeny}} = \frac{7}{10} = 0.7$$

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.

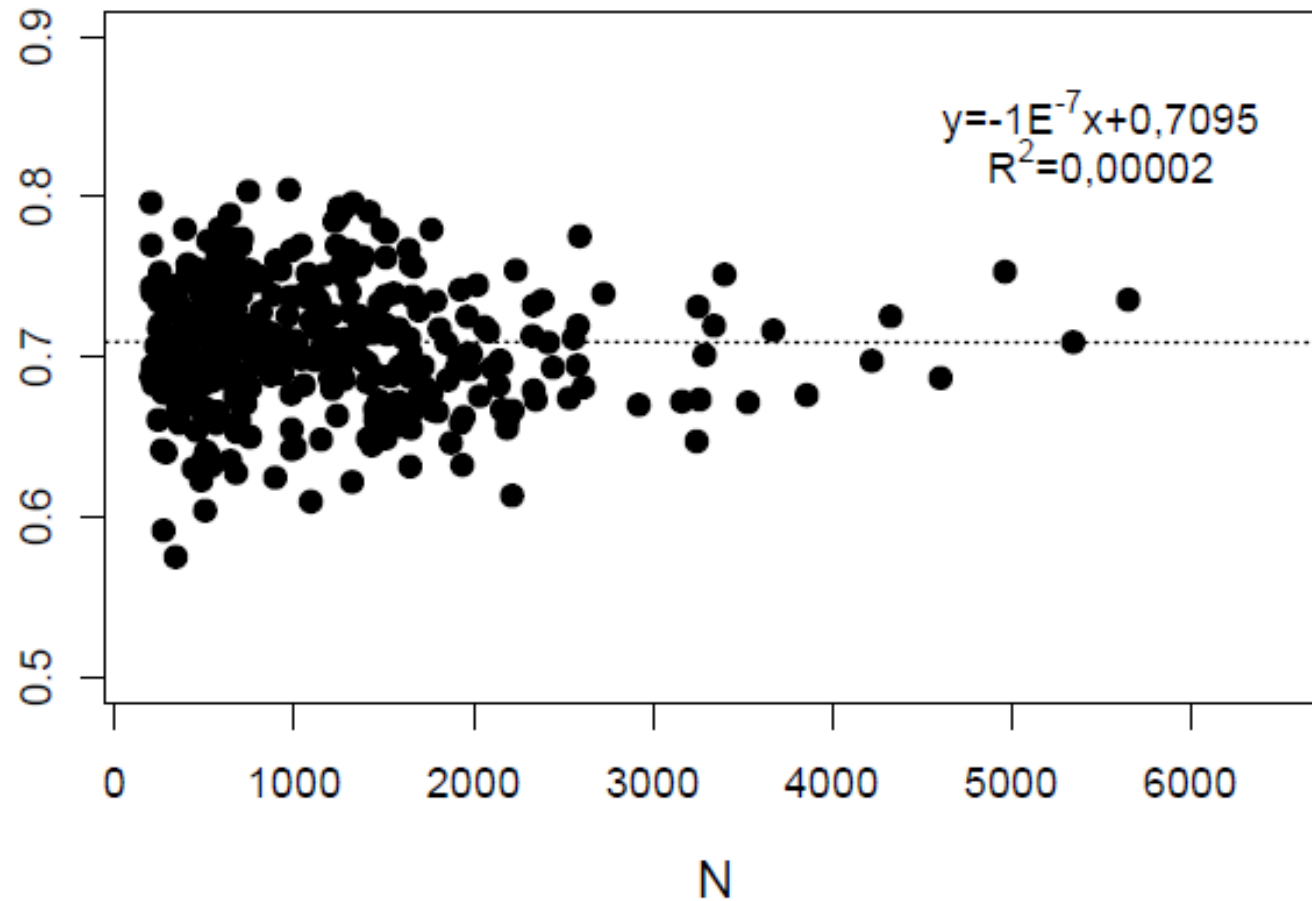
$$MATTR = \frac{\sum_{i=1}^{N-L} V_i}{L(N-L+1)} = \frac{3 + 3 + 4 + 5 + 4 + 5}{5(10 - 5 + 1)} = \frac{24}{30} = 0.8$$

L ... velikost okna

V_i ... počet typů v daném okně

N ... délka textu

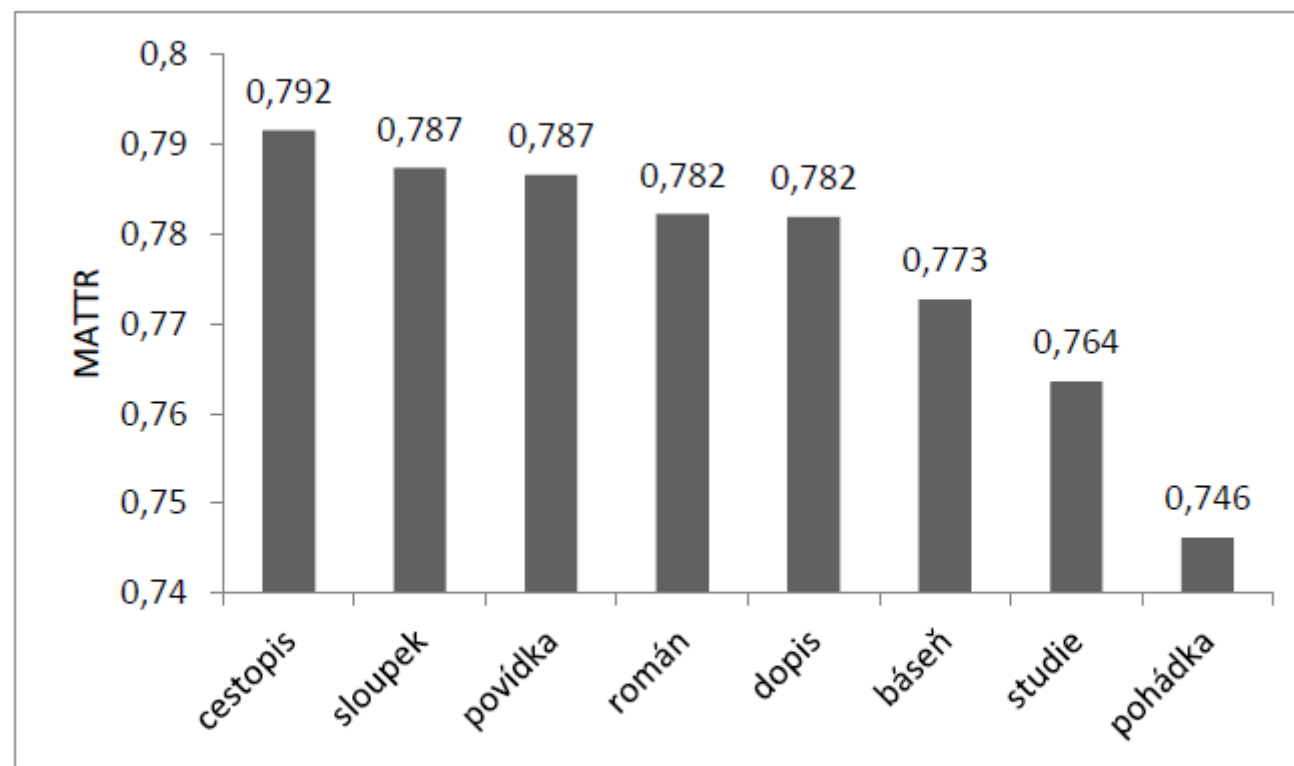
Klouzavý průměr TTR – MATTR



Klouzavý průměr TTR – MATTR

- velikost okna obvykle
 - $L = 100$
 - $L = 500$
- nástroje
 - QuitaUp
 - <https://korpus.cz/quitaup/>

Klouzavý průměr TTR – MATTR



Obr. 19. Výsledné hodnoty *MATTR* v různých žánrech v Čapkových textech

Podíl hapaxů k tokenům (PHL)

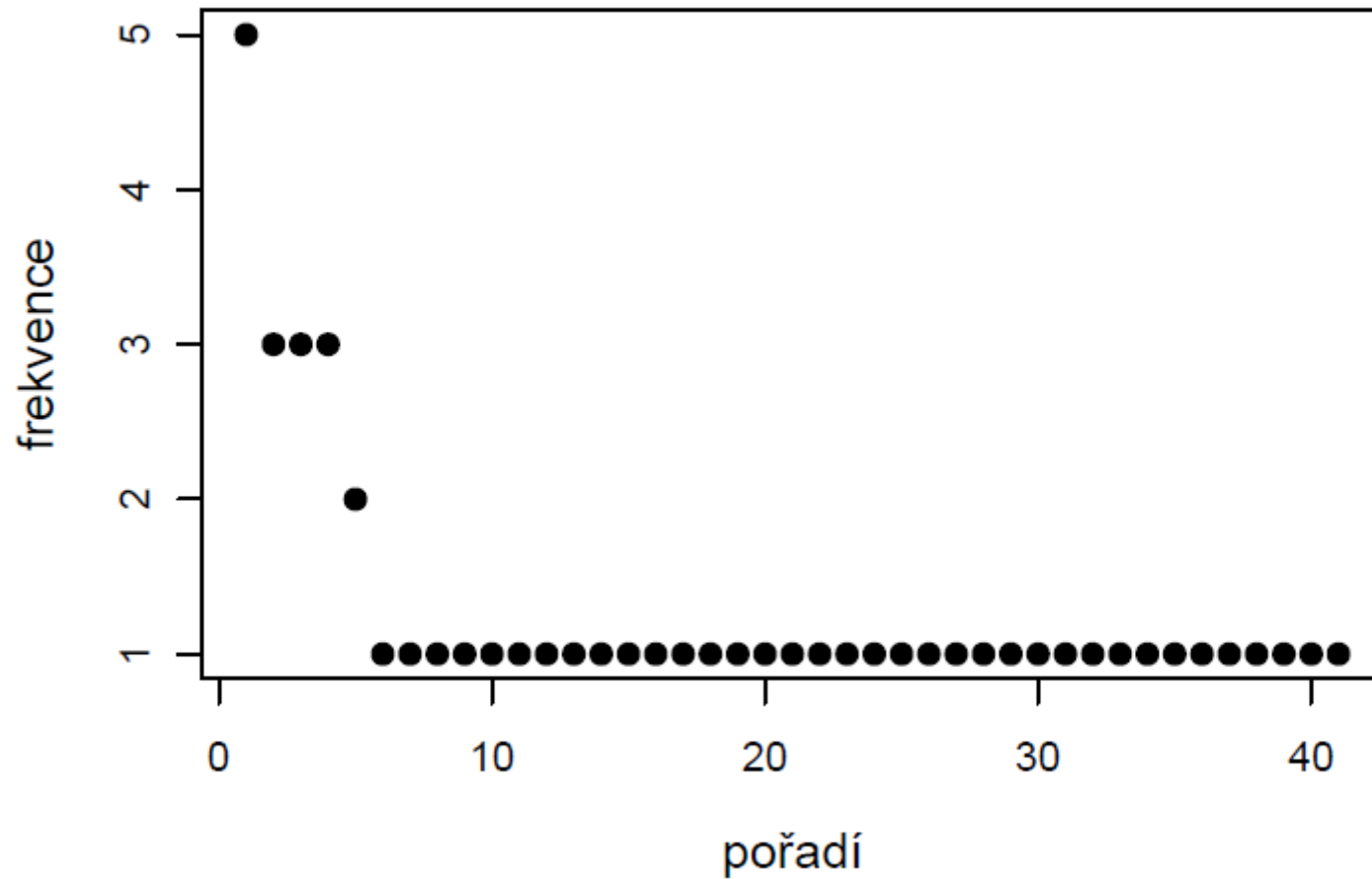
- hapax legomenon
 - výraz, který se vyskytuje v textu jednou

$$PHL = \frac{\sum HL_i}{N}$$

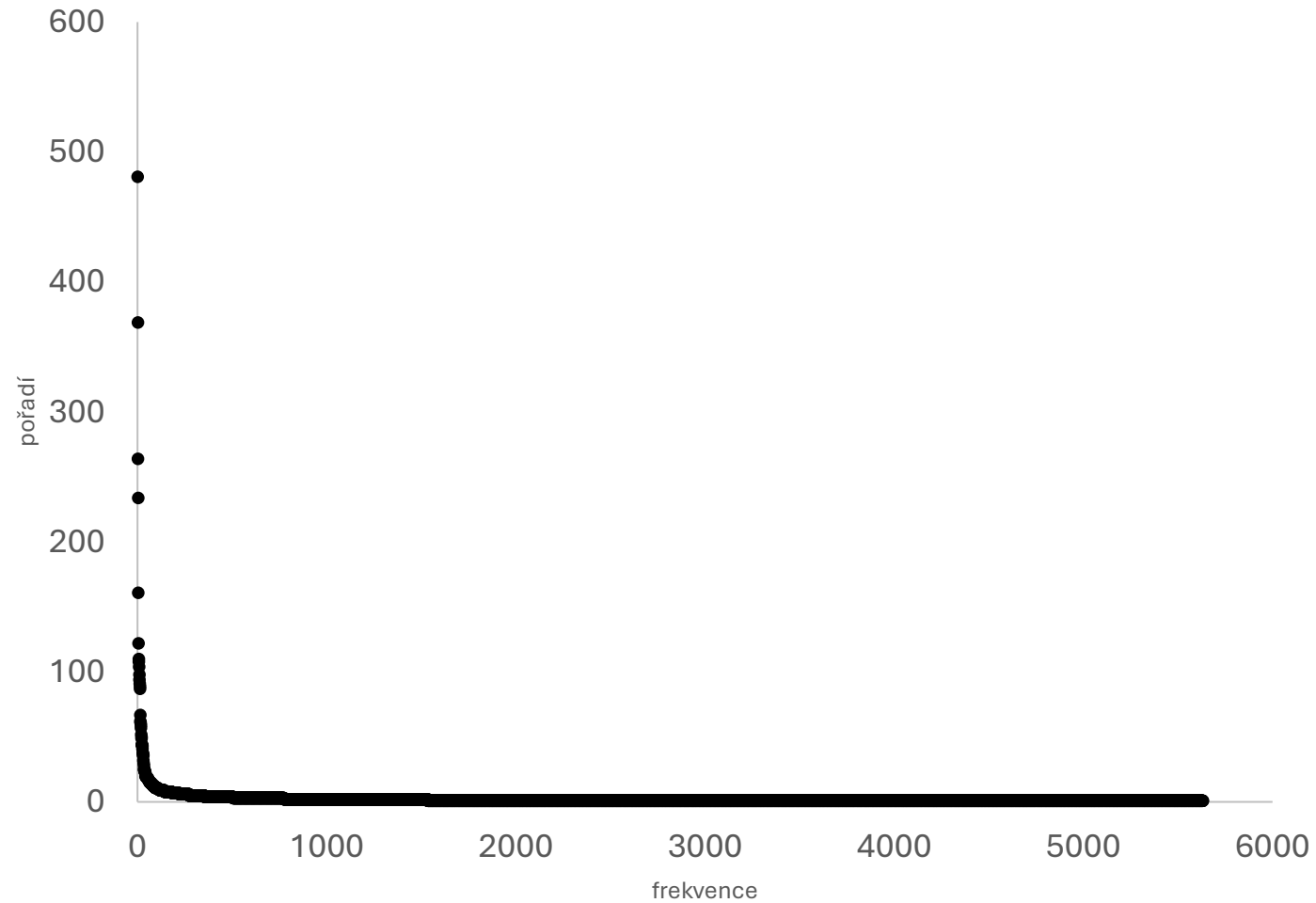
HL...hapax legomenon

N... počet tokenů

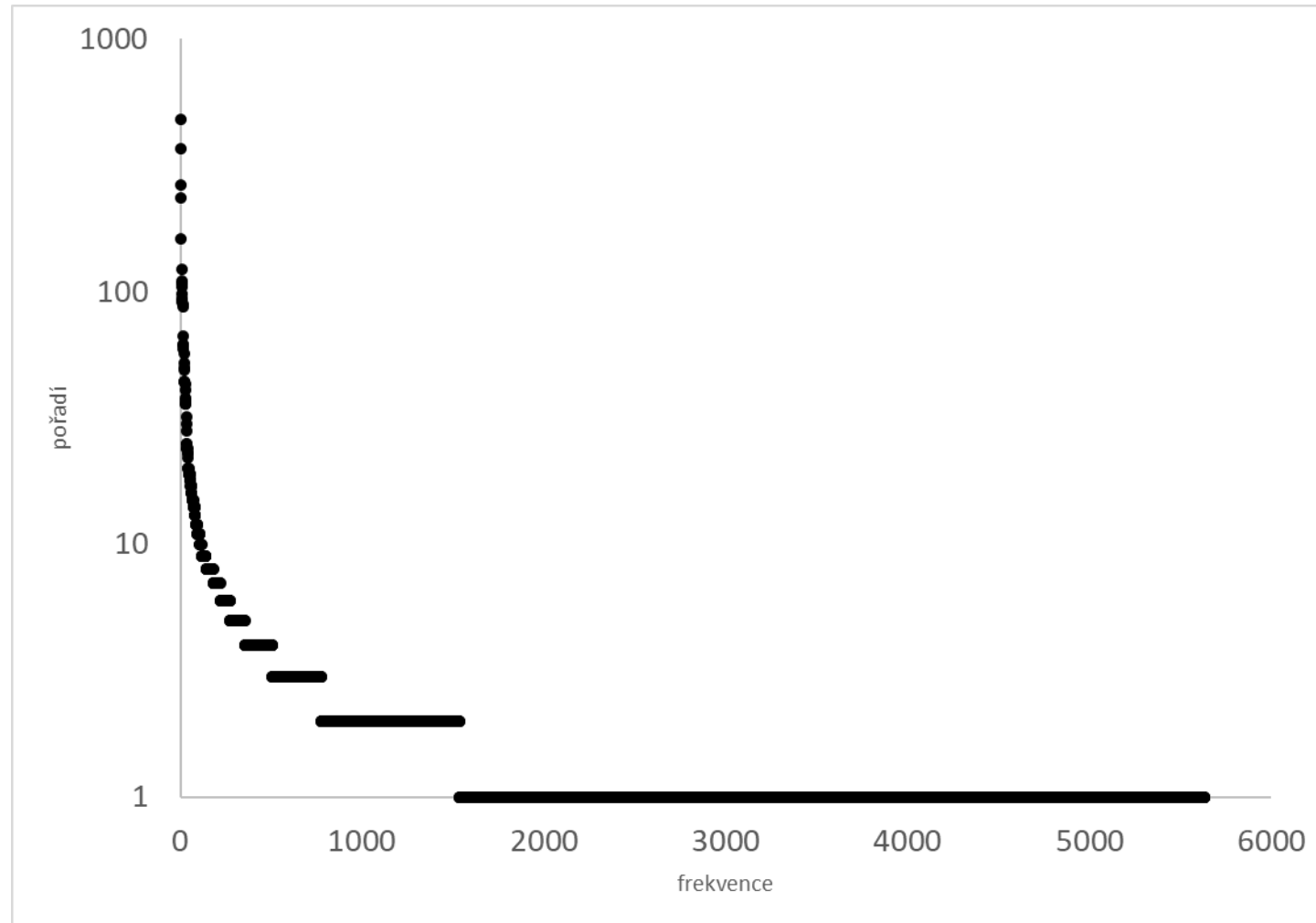
Podíl hapaxů k tokenům (PHL)



Podíl hapaxů k tokenům (PHL)



Podíl hapaxů k tokenům (PHL)

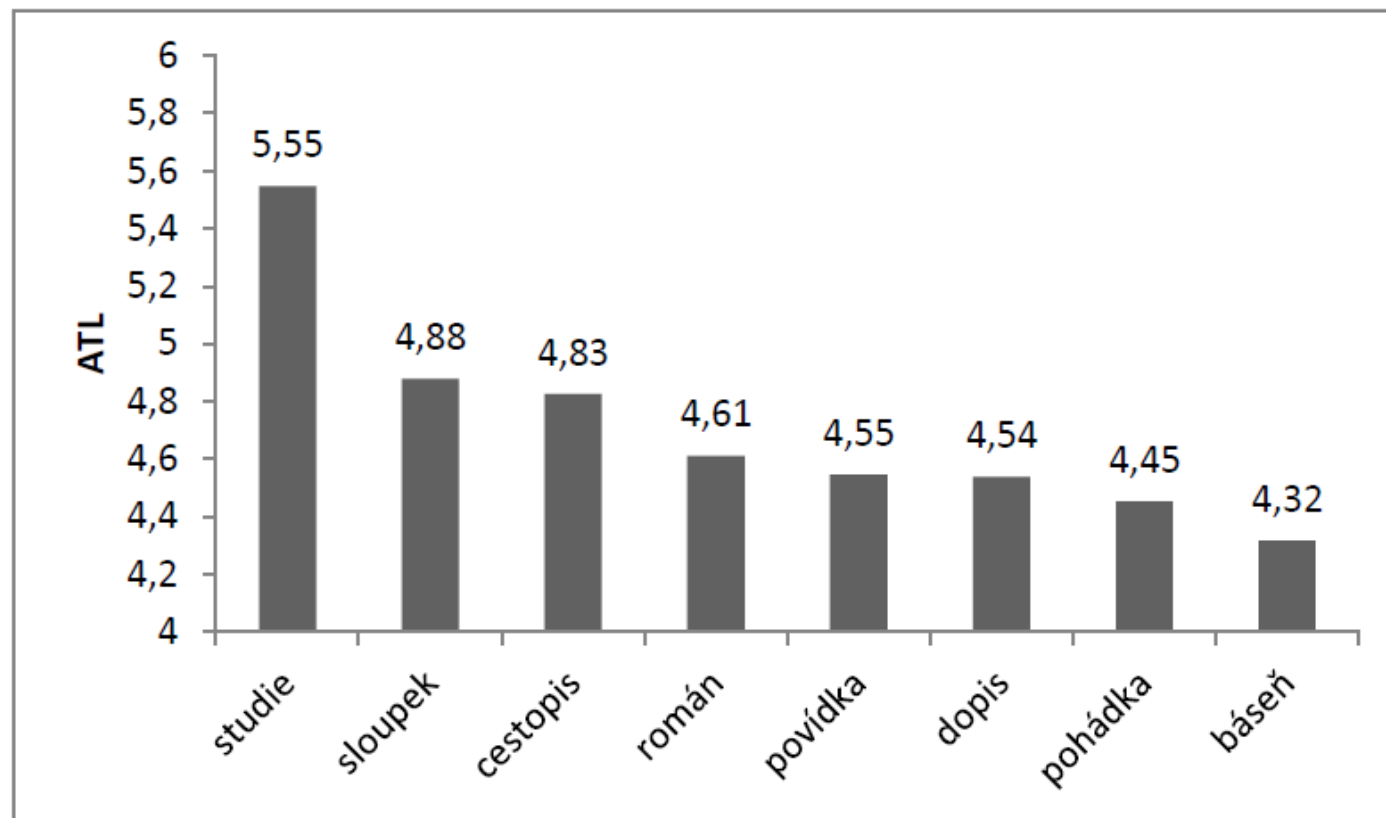


Průměrná délka tokenu (ATL)

- vyjadřuje hodnotu aritmetického průměru délek tokenů v textu
- délka tokenu je počítána v počtu znaků
 - např. slovo chyba má tedy délku 5 znaků

$$ATL = \frac{1}{N} \sum_{i=1}^N x_i$$

Průměrná délka tokenu (ATL)



Obr. 40. Průměrná délka slova v grafémech v různých žánrech v Čapkových textech

Aktivita (Q)

- míra dějovosti textu
- v kontrastu s deskriptivitou (popisností)
- poměr sloves k součtu sloves a adjektiv, které se v textu vyskytují

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

- obrácenou hodnotou je deskriptivita

$$D = 1 - Q$$

Aktivita (Q)

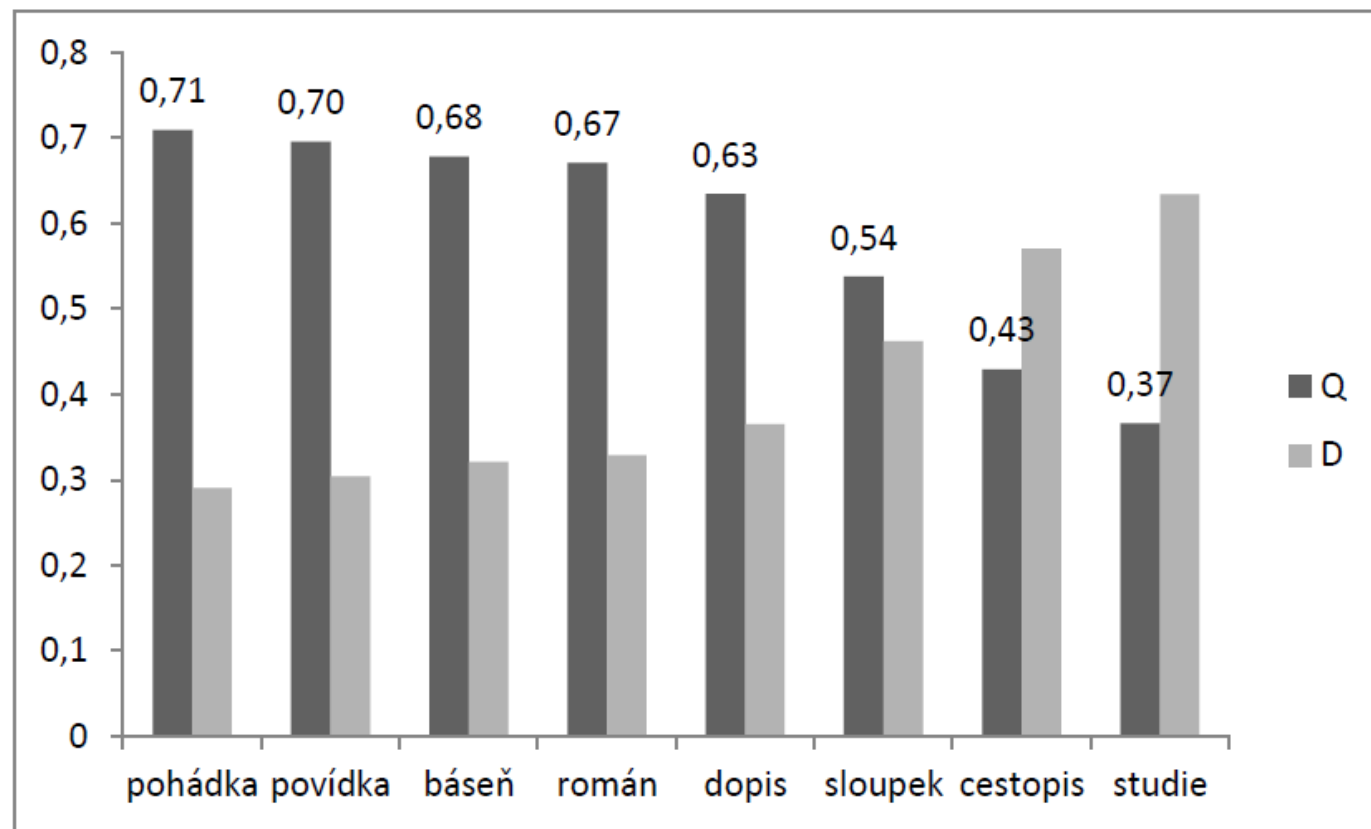
Na malém stole ležela rozbitá váza a pod stolem se válely prázdné lahve.

Aktivita (Q)

Na malém stole ležela rozbitá váza a pod stolem se válely prázdné lahve.

$$Q = \frac{2}{2 + 3} = 0.4$$

Aktivita (Q)



Obr. 42. Aktivita a deskriptivita v různých žánrech v Čapkových textech

Verb distance (VD)

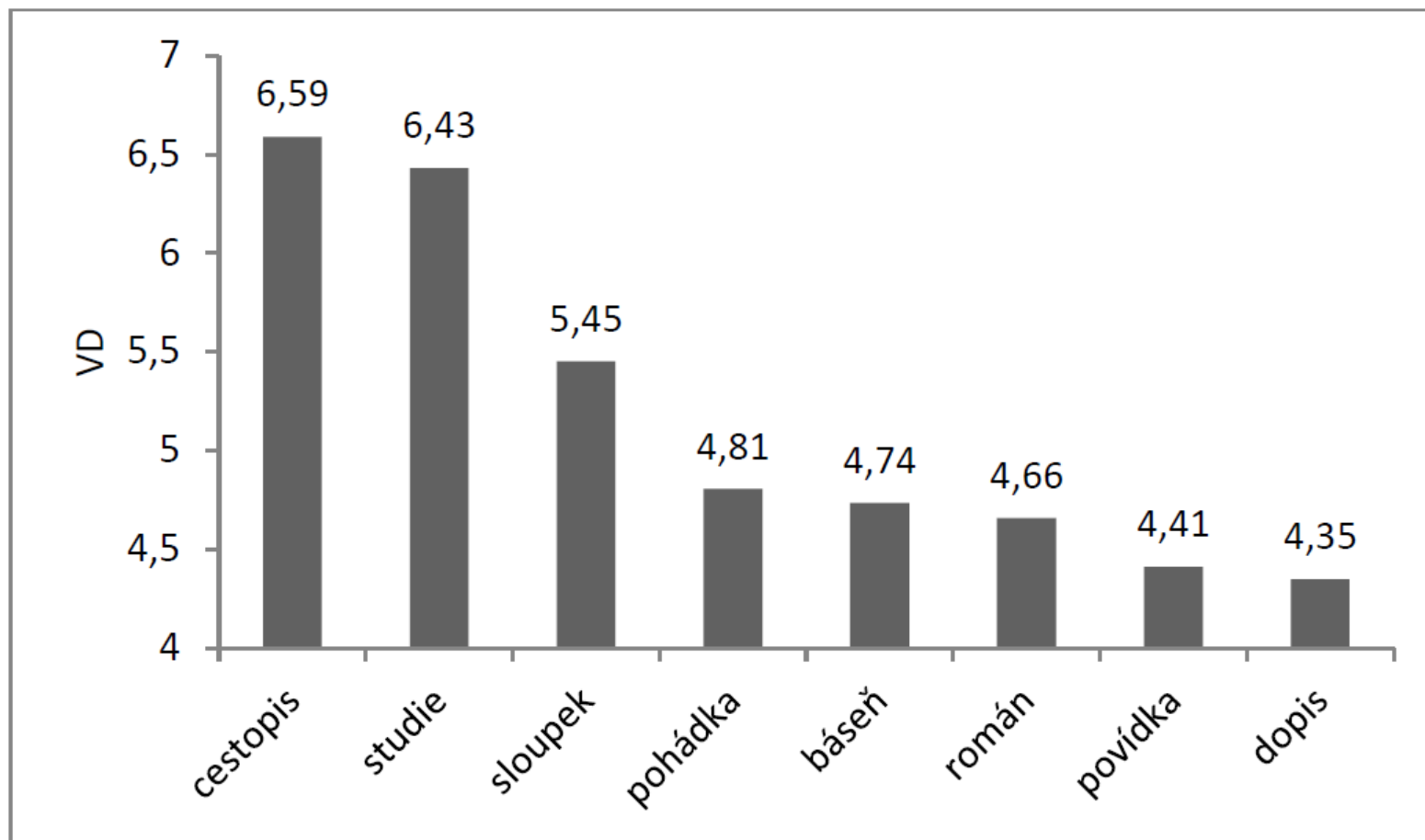
- aritmetický průměr počtu tokenů mezi dvěma po sobě následujícími slovesy v textu
 - nezapočítávají se pomocná slova

Verb distance (VD)

*Potichu **ulehl** naznak. Slunce ho **udeřilo** v oči a **proniklo** zavřenými víčky; červené a černé kruhy se **roztočily** a palčivě **tančí** před očima.*

$$VD = \frac{3 + 3 + 7 + 2}{4} = 3,75$$

Verb distance (VD)



Obr. 38. Vzdálenosti mezi slovesy v různých žánrech v Čapkových textech