

# Úvod do kvantitativní lingvistiky

ZS 2022

# Opakování 1

- předpokládejme, že máme dva texty: A a B
- v textu A je průměrná délka věty (měřena v počtu klauzích) rovna 3, směrodatná odchylka je rovna 6
- v textu B je průměrná délka věty (měřena v počtu klauzích) rovna 3, směrodatná odchylka je rovna 2
- interpretujte tento výsledek

# Tematická analýza textu

- jak na ni?

# Tematická analýza textu

- jak na ni?
- přečíst text...

# Tematická analýza textu

- jak na ni?
- přečíst text...
- viz texty
  - *V Beskydech blesk zapálil chatu, vítr lámal stromy*
  - *Odvaha k tomu*
  - *Smuténka*
    - vyberte z textu slova, která b bylo možno označit za tematická/klíčová

# Tematická analýza textu

- jak na ni?
- přečíst text...
- limity?

# Tematická analýza textu

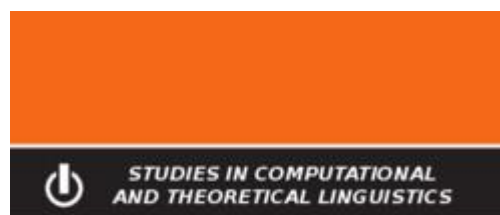
- frekvence?

# Tematická analýza textu

- analýza klíčových slov
  - <https://wiki.korpus.cz/doku.php/manualy:keywords>
  
- analýza tematických slov



# Tematická koncentrace textu



TEMATICKÁ KONCENTRACE TEXTU  
V ČEŠTINĚ

Radek Čech



# H-bod

- vypočítejte h-bod z následujících r-f distribucí

r	f
1	30
2	15
3	10
4	8
5	8
6	6
7	3
8	2
9	1
10	1
...	
52	1

r	f
1	60
2	30
3	20
4	15
5	13
6	12
7	8
8	5
9	3
10	2
...	
52	1