

Orthography system of broadside ballads from 17th century. Quantitative approach



Radek Čech
Ján Mačutek



Motivations

- theoretical
 - philological insights, hints, speculations...
 - in principle, empirically testable
 - Voit (2016), Koupil (2017)

Motivations

- theoretical
 - philological insights, hints, speculations...
 - in principle, empirically testable
 - Voit (2016), Koupil (2017)
- pragmatic
 - availability of electronic transliterated texts

Motivations

- theoretical
 - philological insights, hints, speculations...
 - in principle, empirically testable
 - Voit (2016), Koupil (2017)
- pragmatic
 - availability of electronic transliterated texts
- methodological
 - application of statistical methods to linguistics

Language material

- 1650 Litomyšl, M. V. Březina, MZK 0000.542_22
- 1664 Litomyšl, J. Arnolt, MZK 0000.542_09
- 1652 Olomouc, M. Hradecký, MZK 0000.385_11
- 1685 Olomouc, J. J. Kylián, MZK 0000.542_07
- transliteration
 - project: *Broadside Ballads in the Historical Collections in Brno*

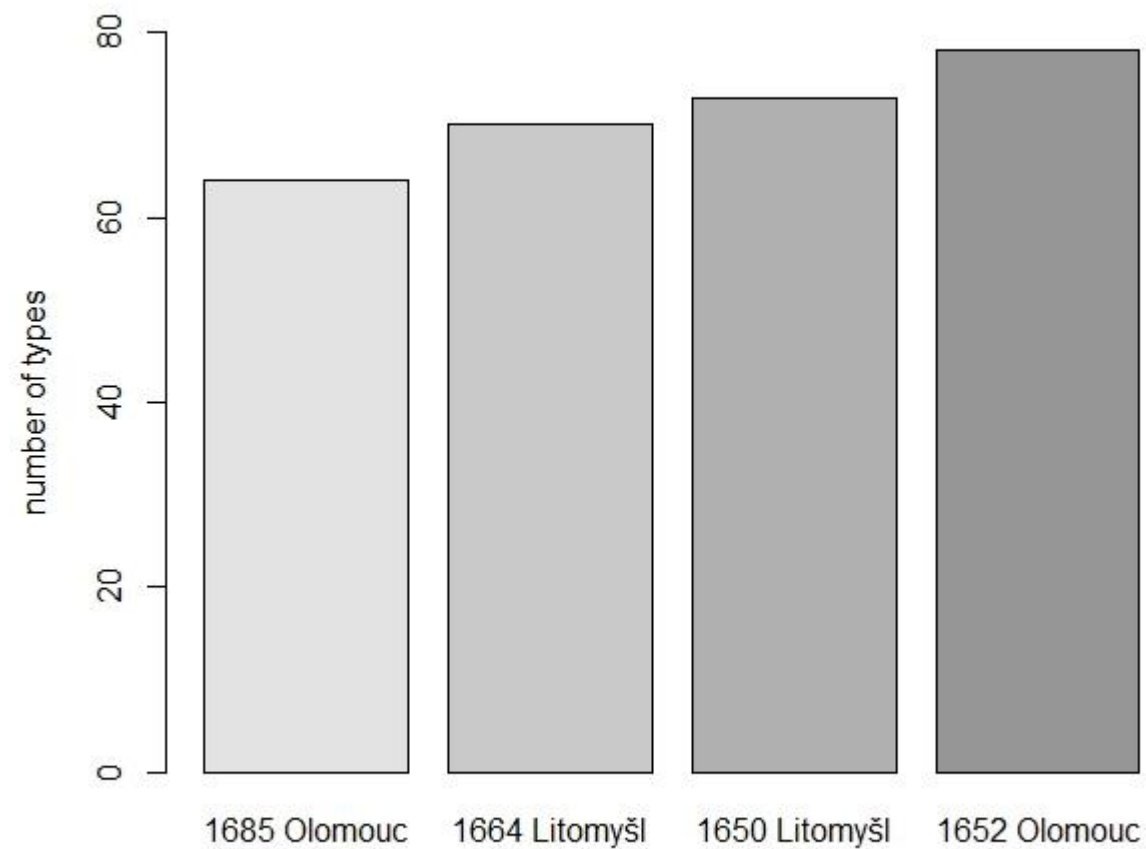
Types

- the same phonological system
- absence of codification of orthography = variability
- a comparison of types = differences among printing houses
- differences
 - period
 - tradition of particular printing houses
 - typesetter's and corrector competence

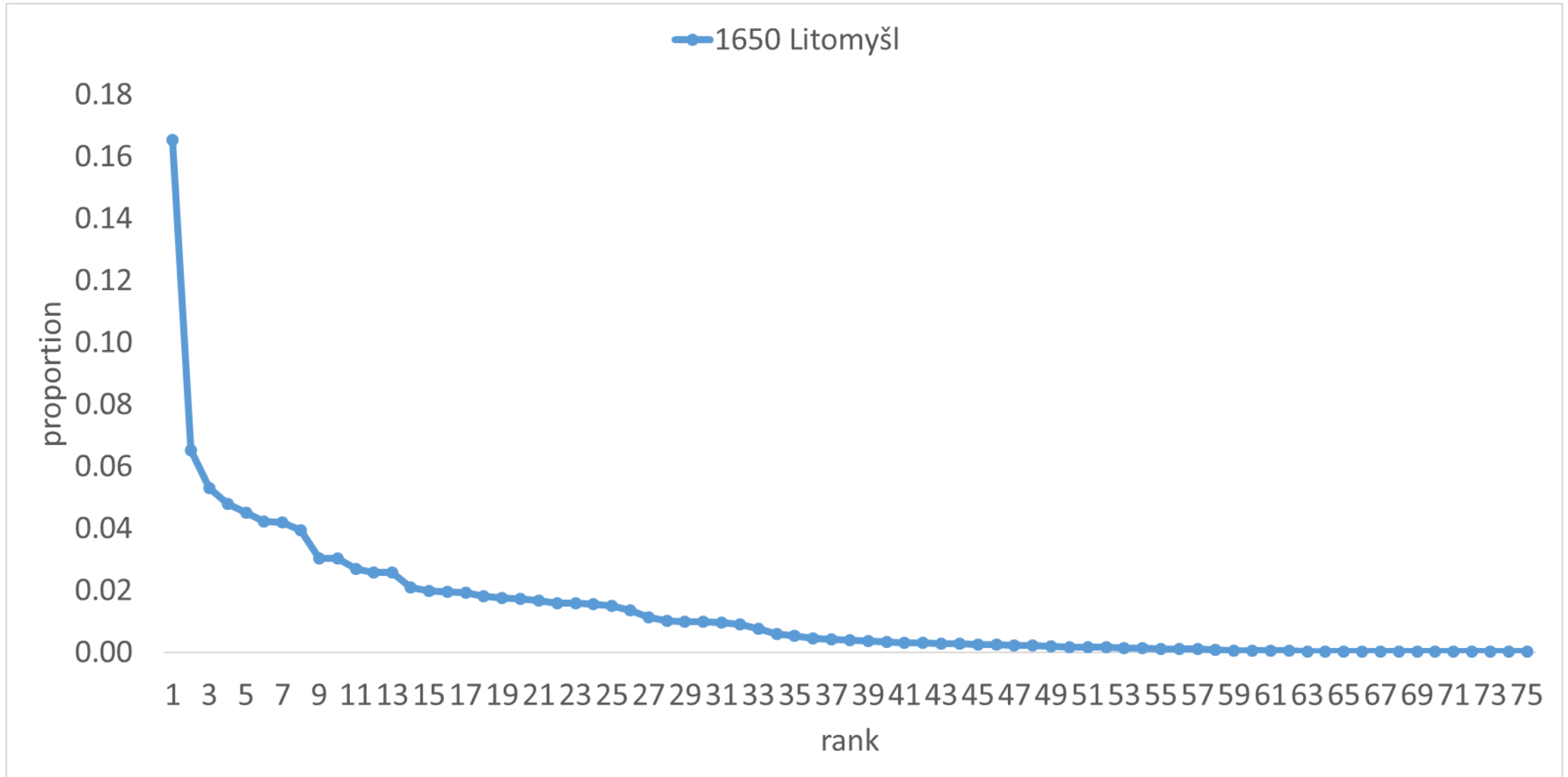
Types

- comparison of texts
 - inventory of types
 - rank-frequency distribution of types

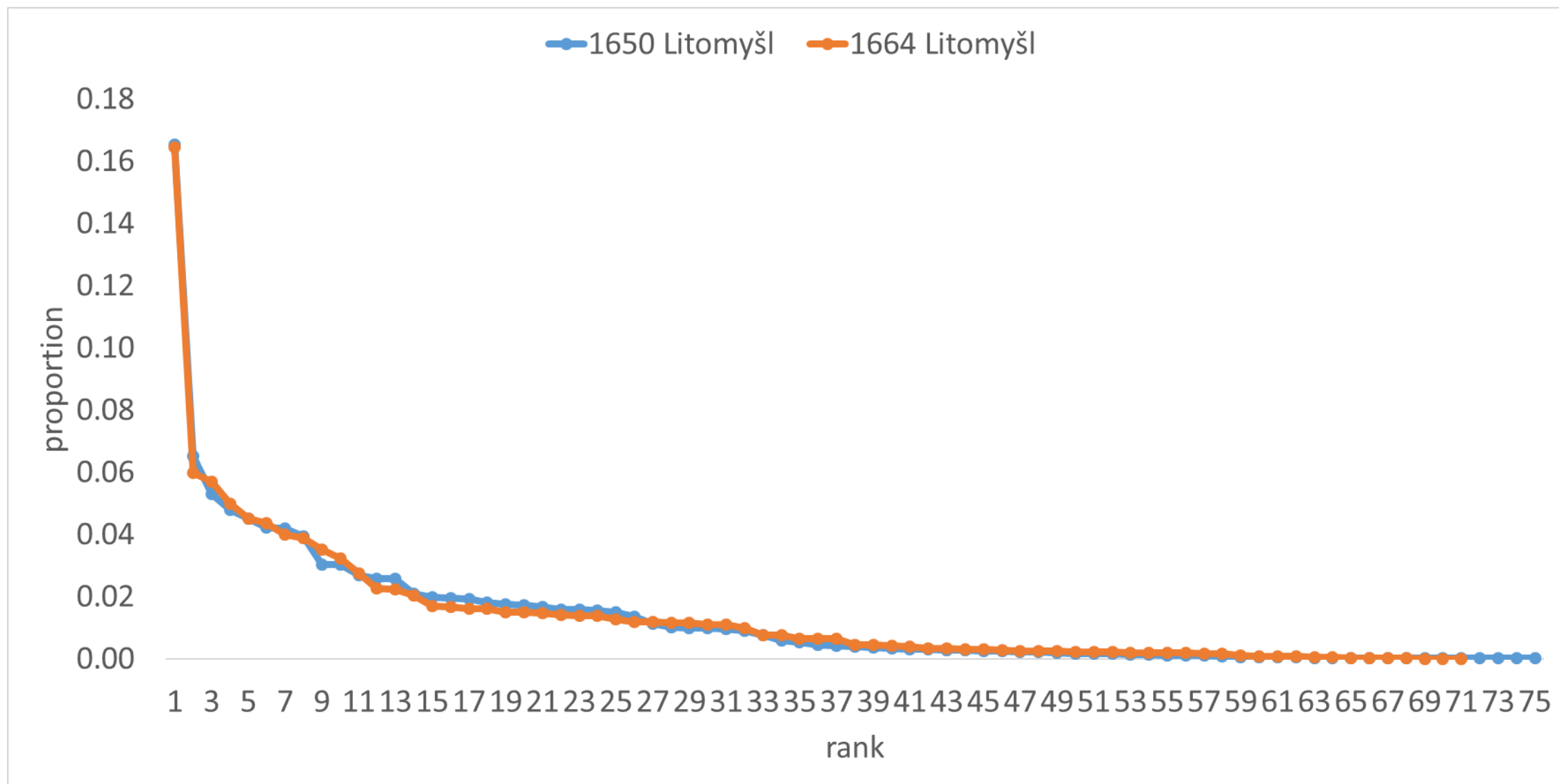
Inventory of types



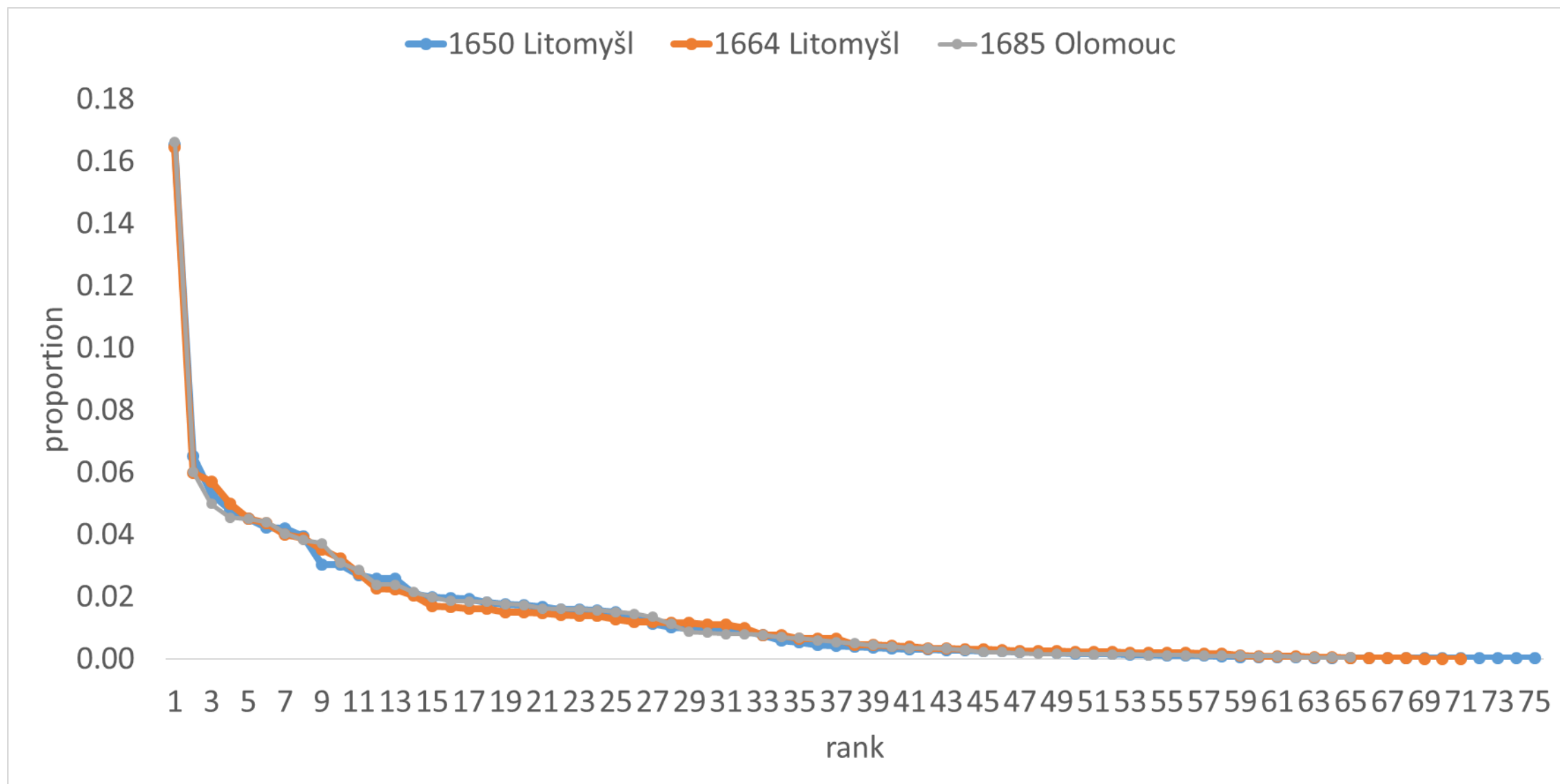
Rank-frequency distribution of types



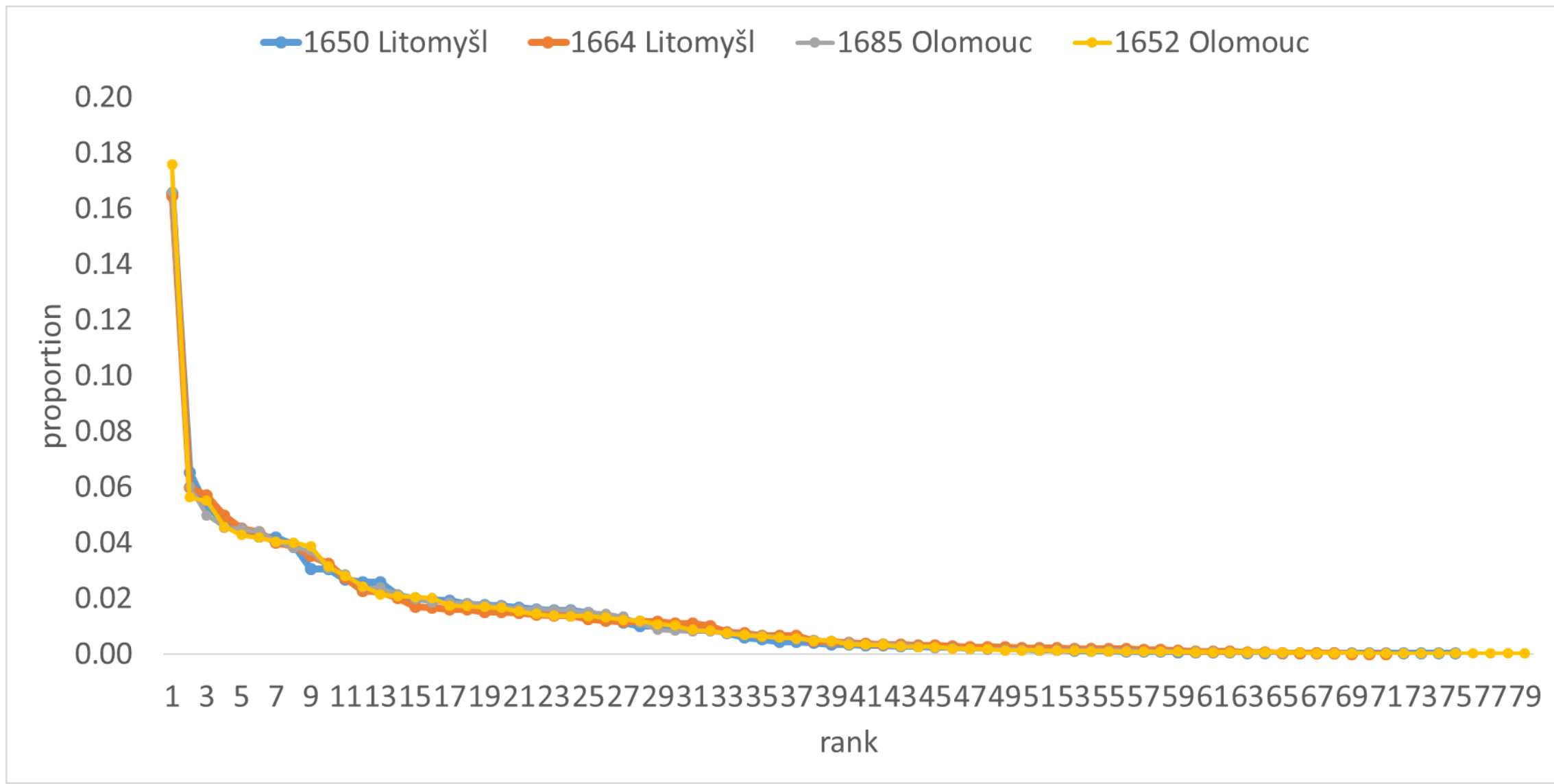
Rank-frequency distribution of types



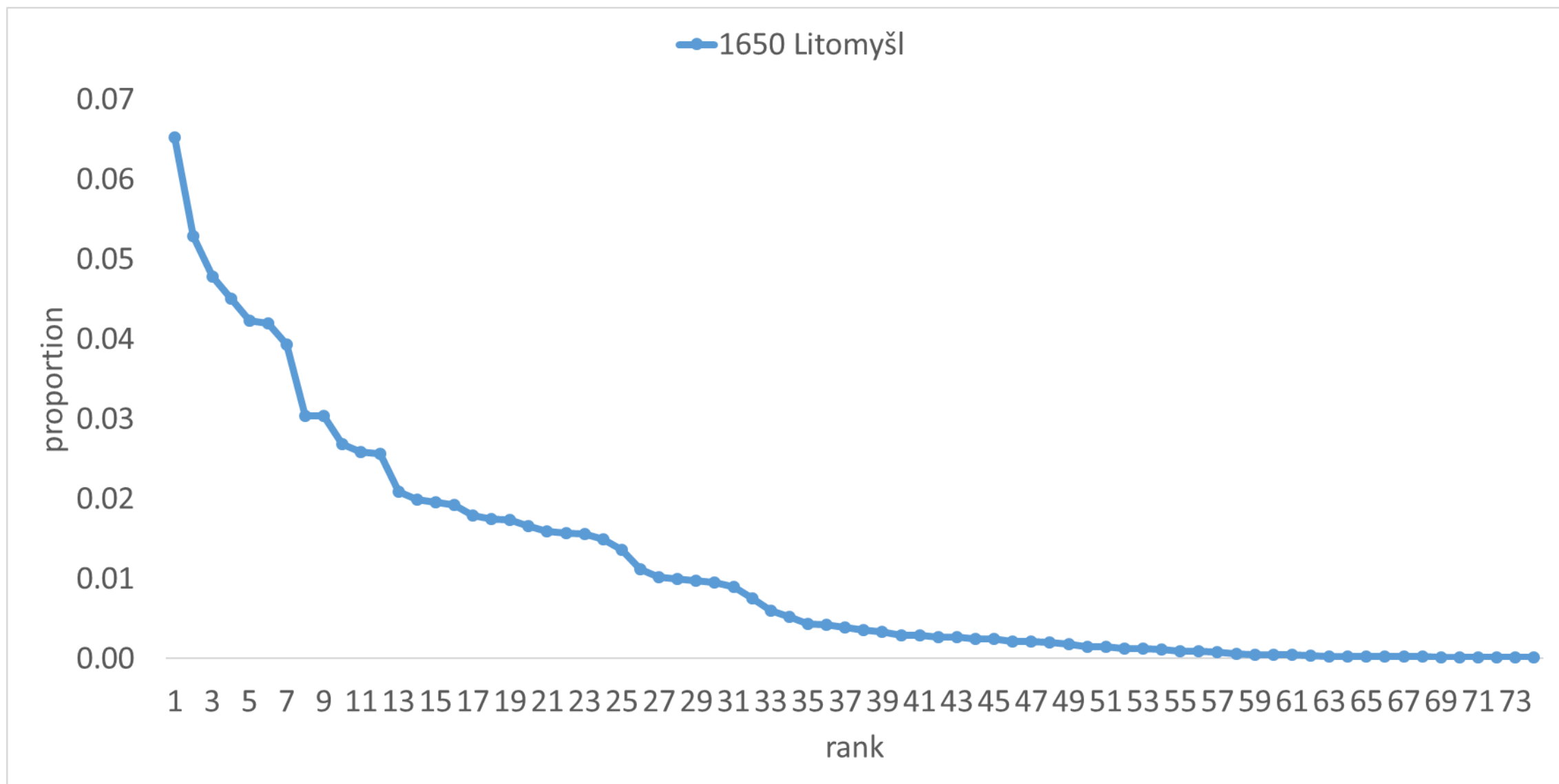
Rank-frequency distribution of types



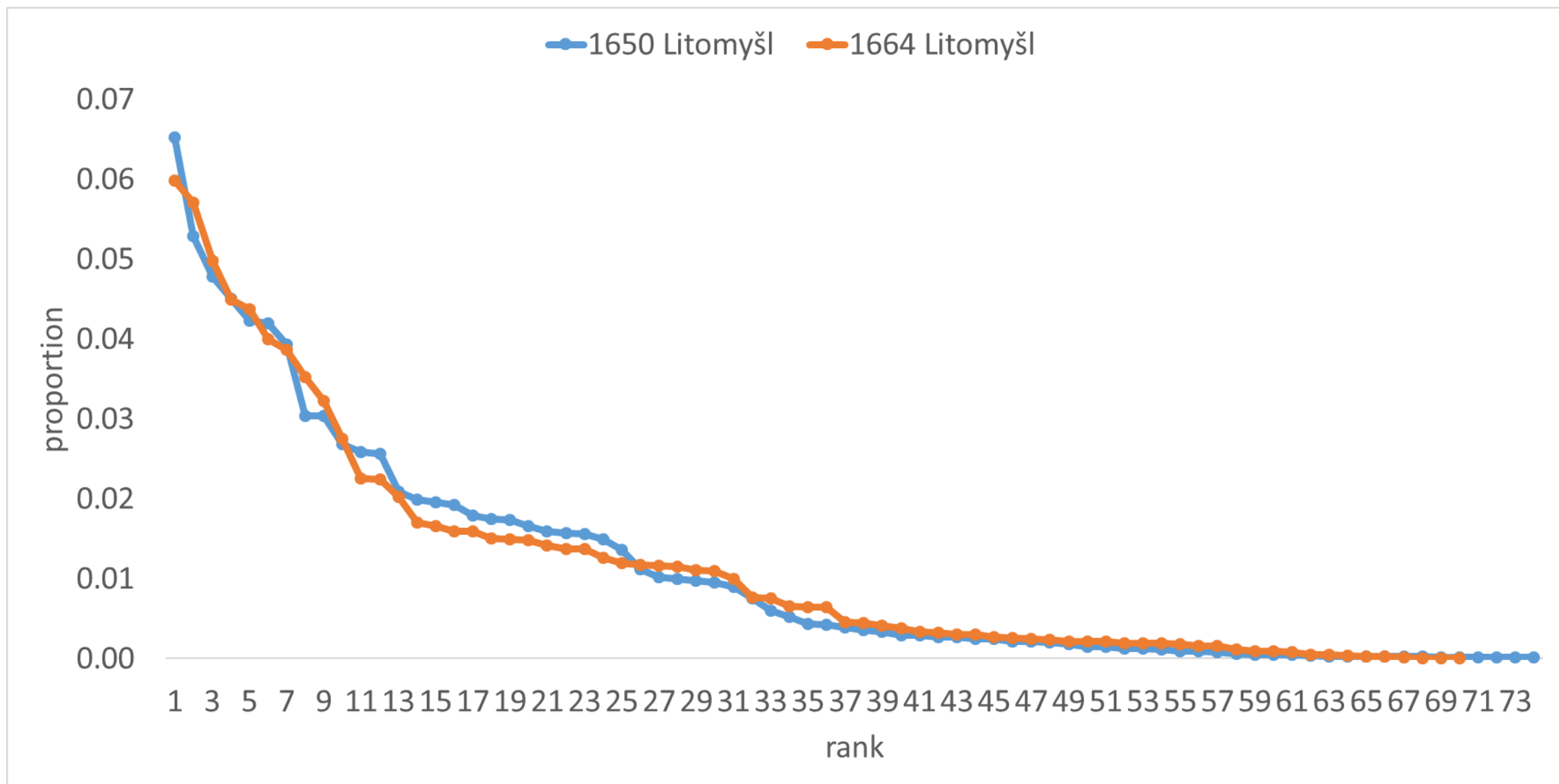
Rank-frequency distribution of types



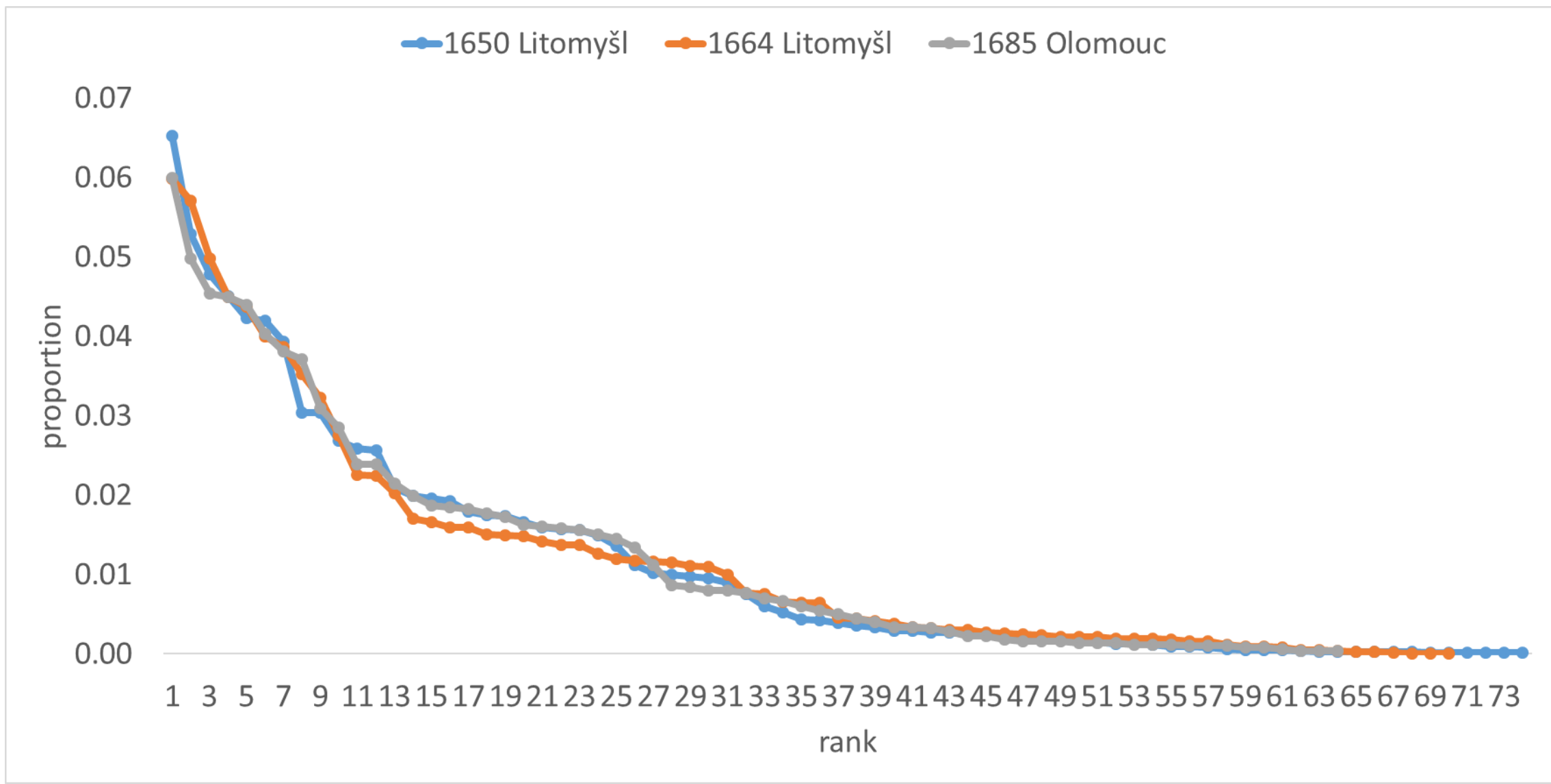
Rank-frequency distribution of types



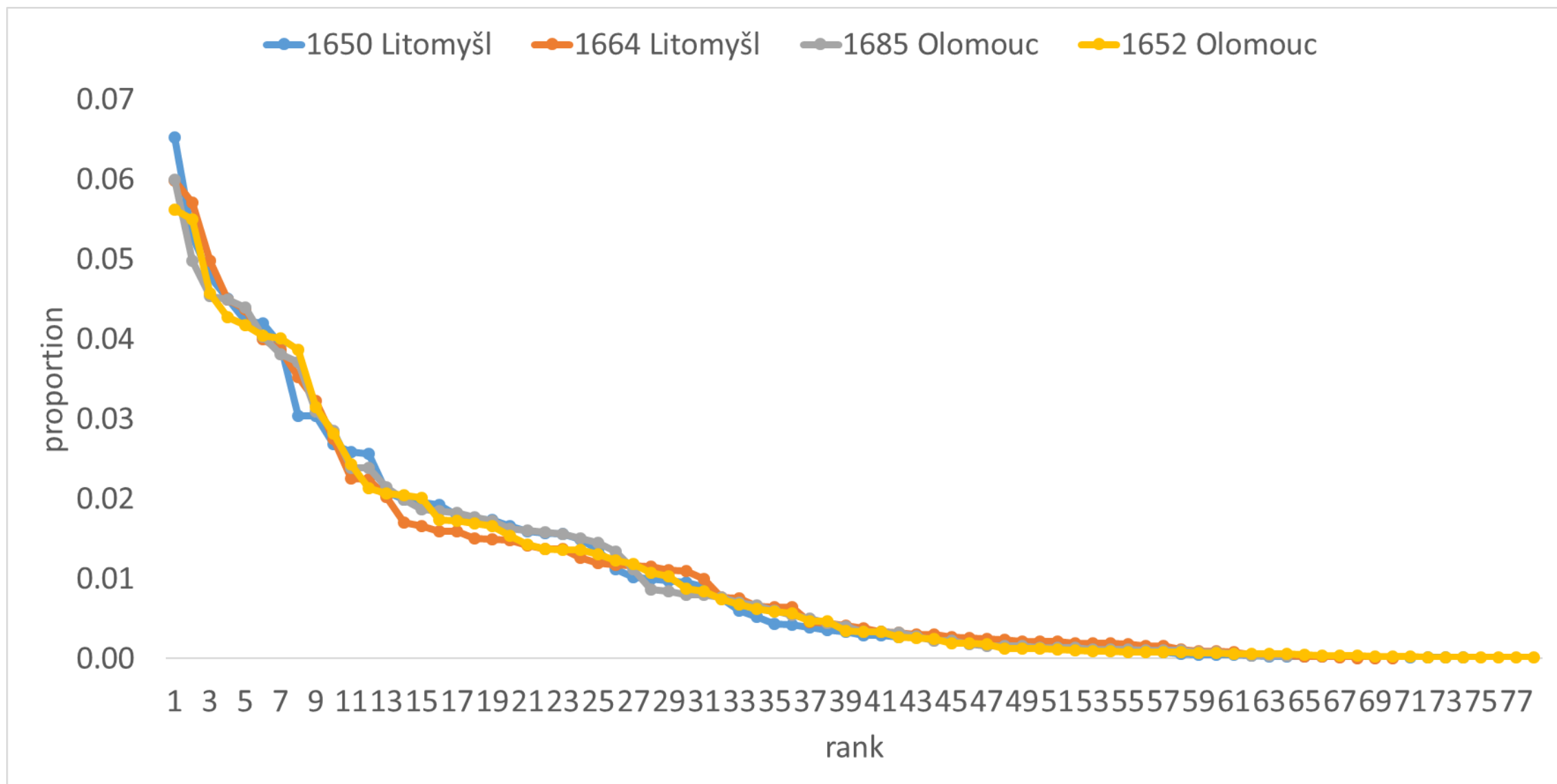
Rank-frequency distribution of types



Rank-frequency distribution of types



Rank-frequency distribution of types



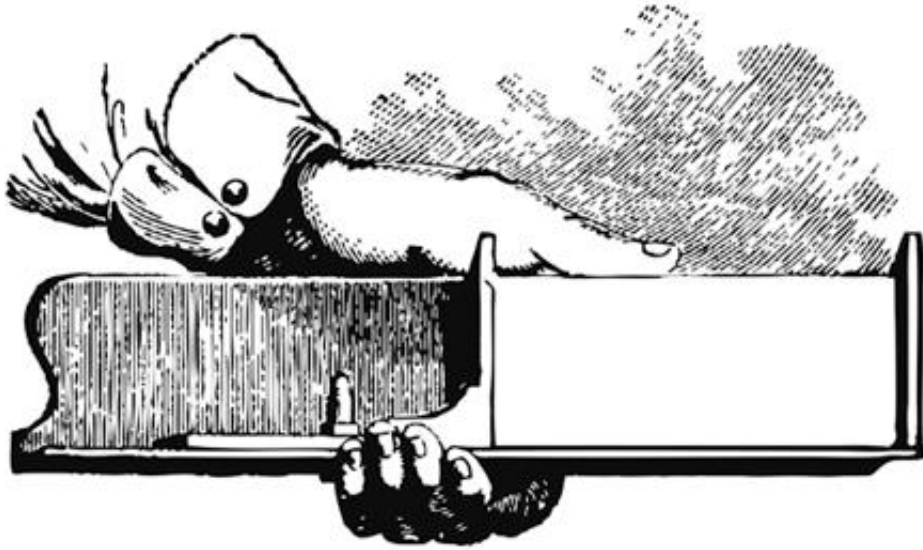
Mathematical model for frequencies

- frequencies in all texts can be modelled by the negative hypergeometric distribution

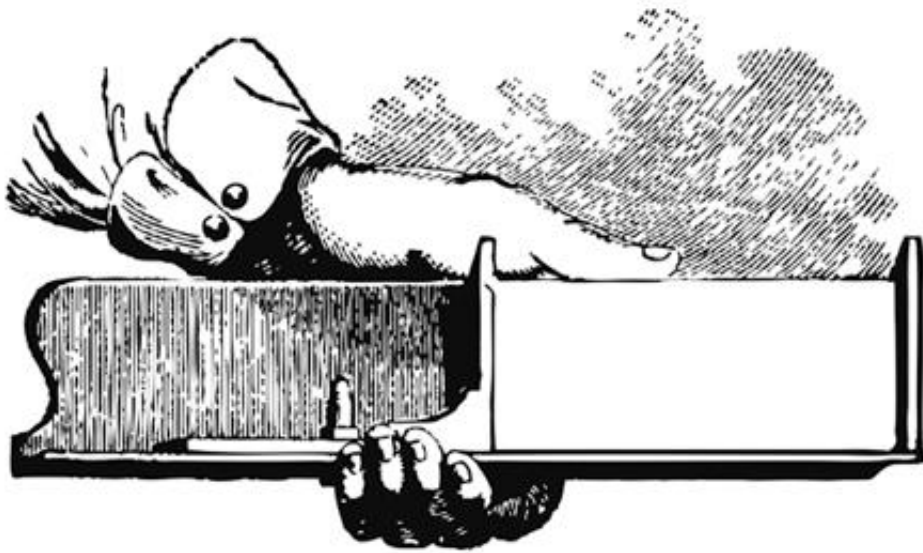
$$P_x = \frac{\binom{M+x-2}{x-1} \binom{K-M+n-x}{n-x+1}}{\binom{K+n-1}{n}}$$

- the same model as for grapheme frequencies in many other languages (Slavic, Celtic, West African,...)
- i.e. the same system behaviour, but with different parameter values (K, M, n)

Length of the line



Length of the line



- How to fit the text to the width of composing stick?

Length of the line

- multiply spaces
- hyphenation
- punctuation
- variation of ortographics and typography
 - *buoh – bŭh*
 - *býti – být*
 - *naffe – naše*
- ...

Length of the line

- multiply spaces
- hyphenation
- punctuation
- **variation of ortographics and typography**
 - *buoh – bŭh*
 - *býti – být*
 - *naffe – naše*
- these variations can be a consequence of a **need to fit the borders of composing stick**

Length of the line

- multiply spaces
- hyphenation
- punctuation
- **variation of ortographics and typography**
 - *buoh – bŭh*
 - *býti – být*
 - *naffe – naše*
- these variations can be a consequence of a **need to fit the borders of composing stick**
- they are **not** necessarily caused by **stylistic** or „inner“ **linguistic** reasons

Length of the line & variation of ortographics

- hypothesis:
 - the more types in the line, the higher probability of occurring of digraphic grapheme

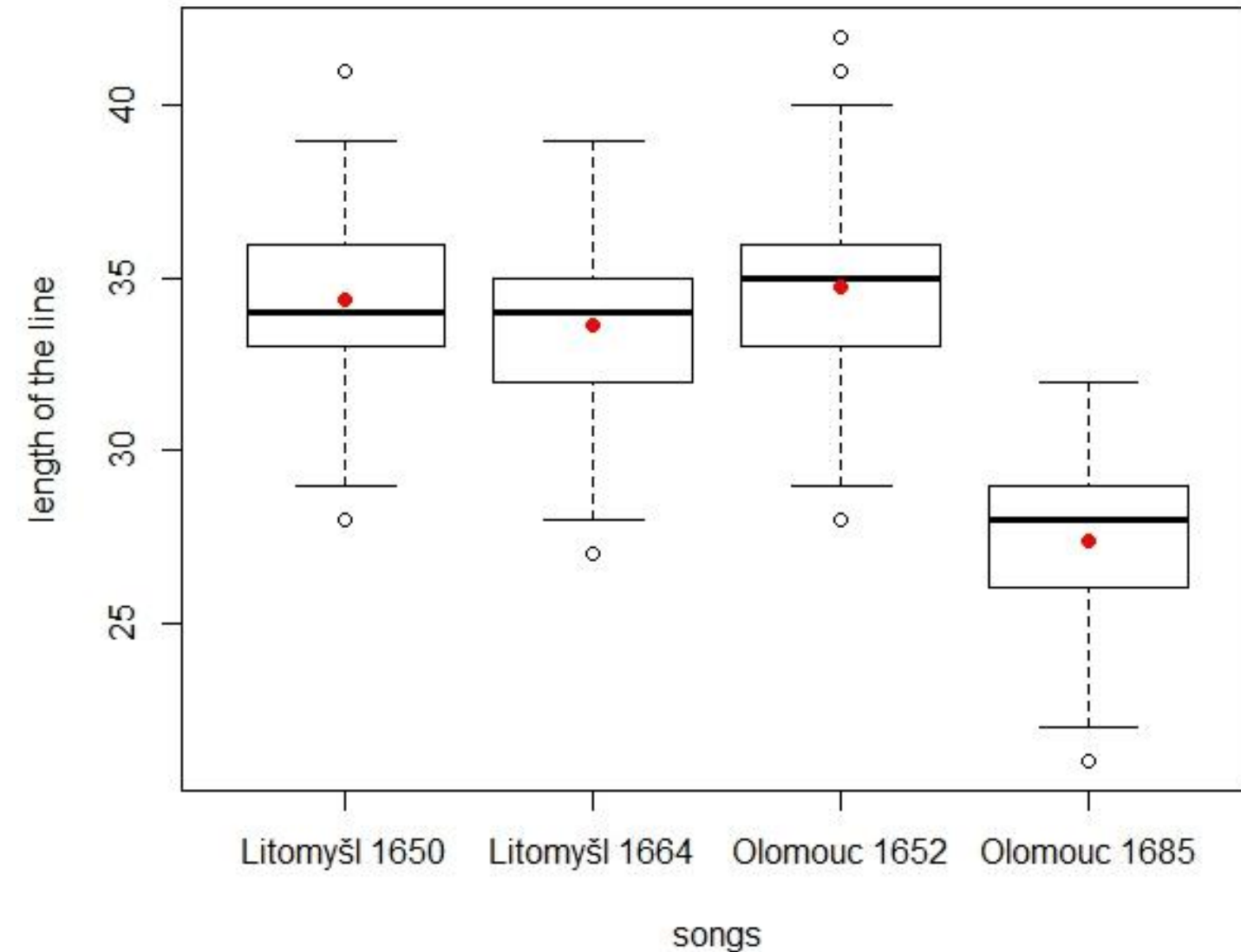
Length of the line & variation of ortographics

- hypothesis:
 - the more types in the line, the higher probability of occurring of bigramic grapheme
- **however**
 - multiply spaces
 - different width of some types
 - special types
 - pilcrow ¶
 - interpunction
 - initial
 -

Length of the line – first insights

- the width of the line
 - measured in number of types
- excluded
 - lines with initials
 - „loosing“ typesetting
 - incomplete lines

Length of the line – first insights



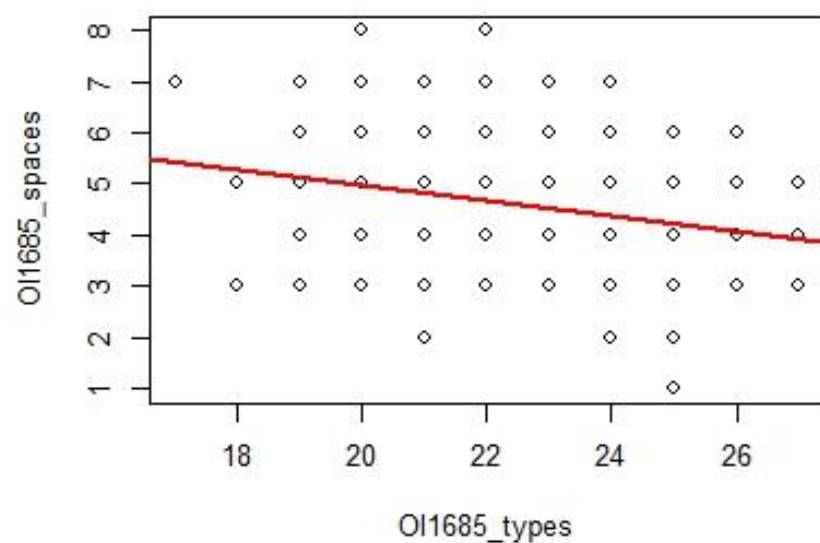
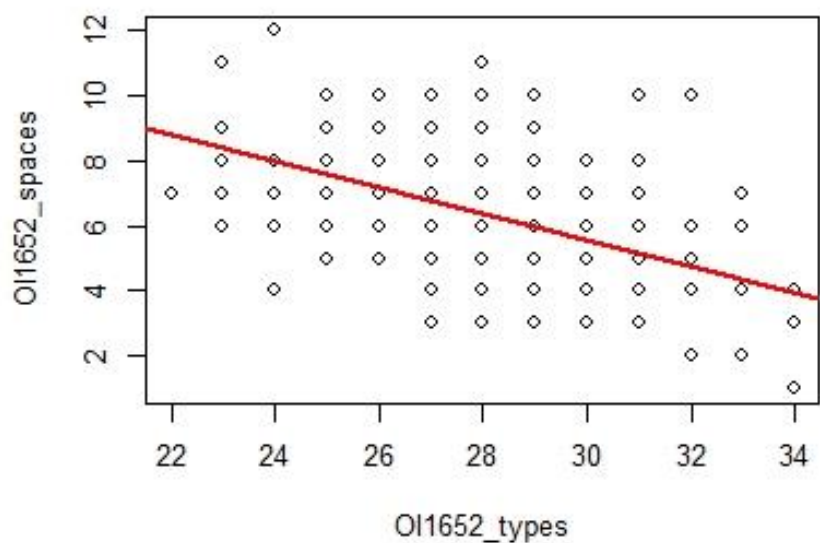
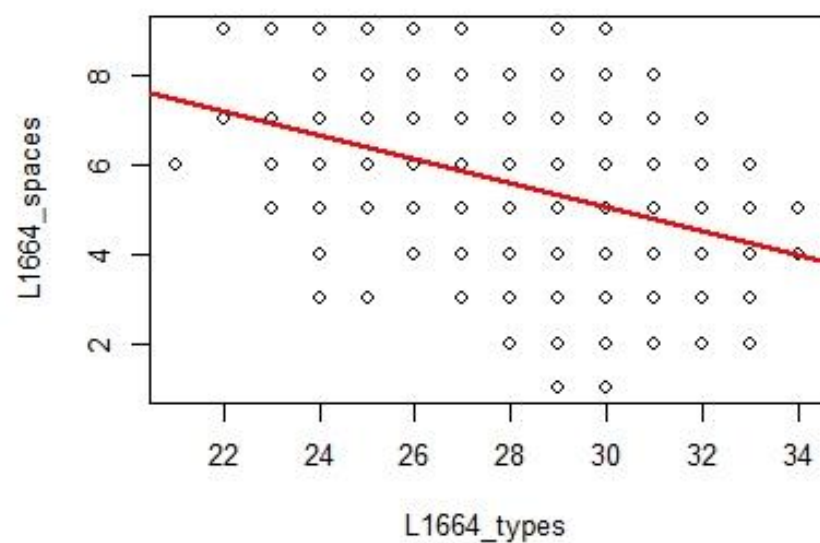
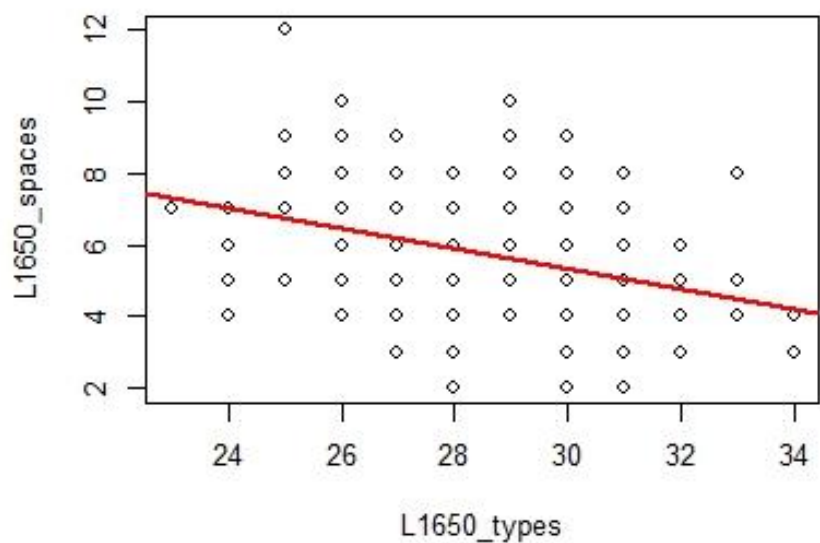
	Lit 1650	Lit 1664	OI 1652	OI 1685
Max.	41	39	42	32
3rd Quartile	36	35	36	29
Median	34	34	35	28
Mean	34.38	33.65	34.74	27.36
1st Quartile	33	32	33	26
Min.	28	27	28	21

Types & spaces

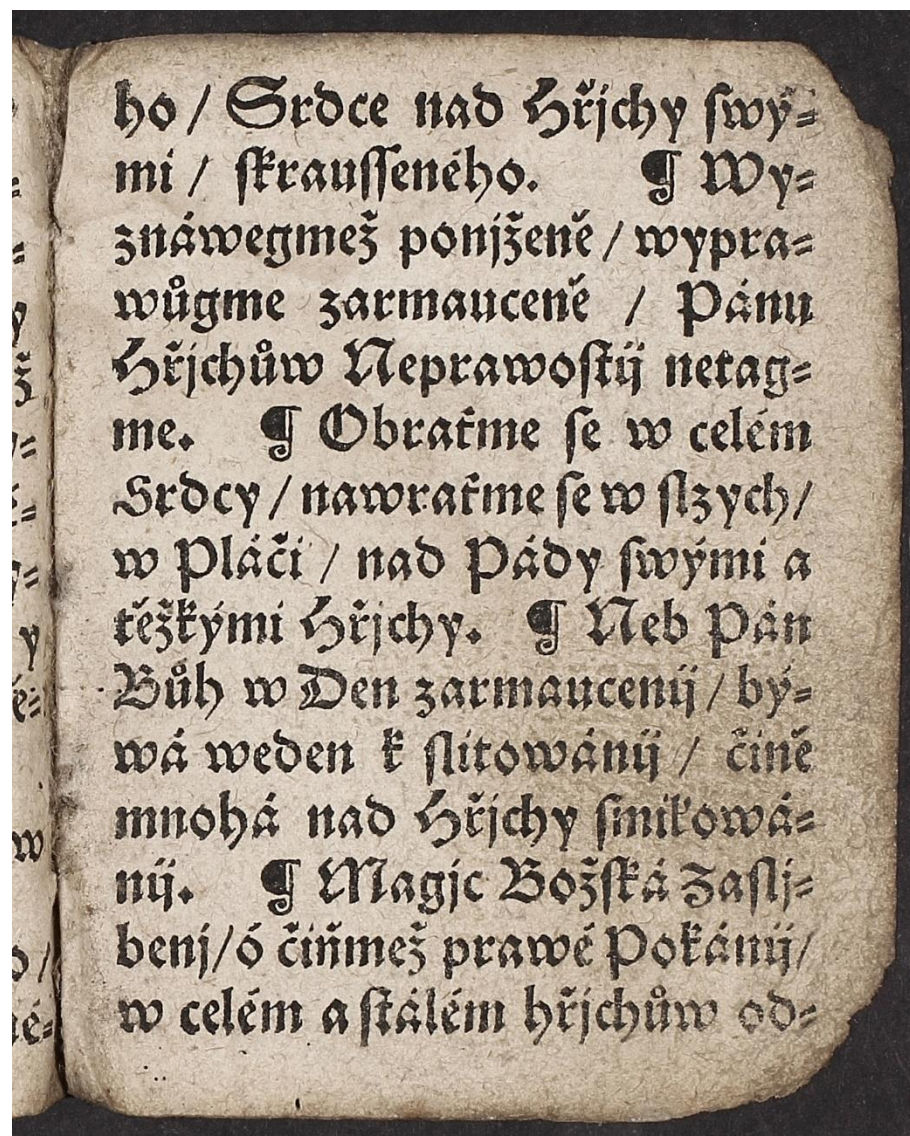
- hypothesis

The more types in the line, the less spaces in the line

Types & spaces



Length of the line – the impact of the pilcrow ¶

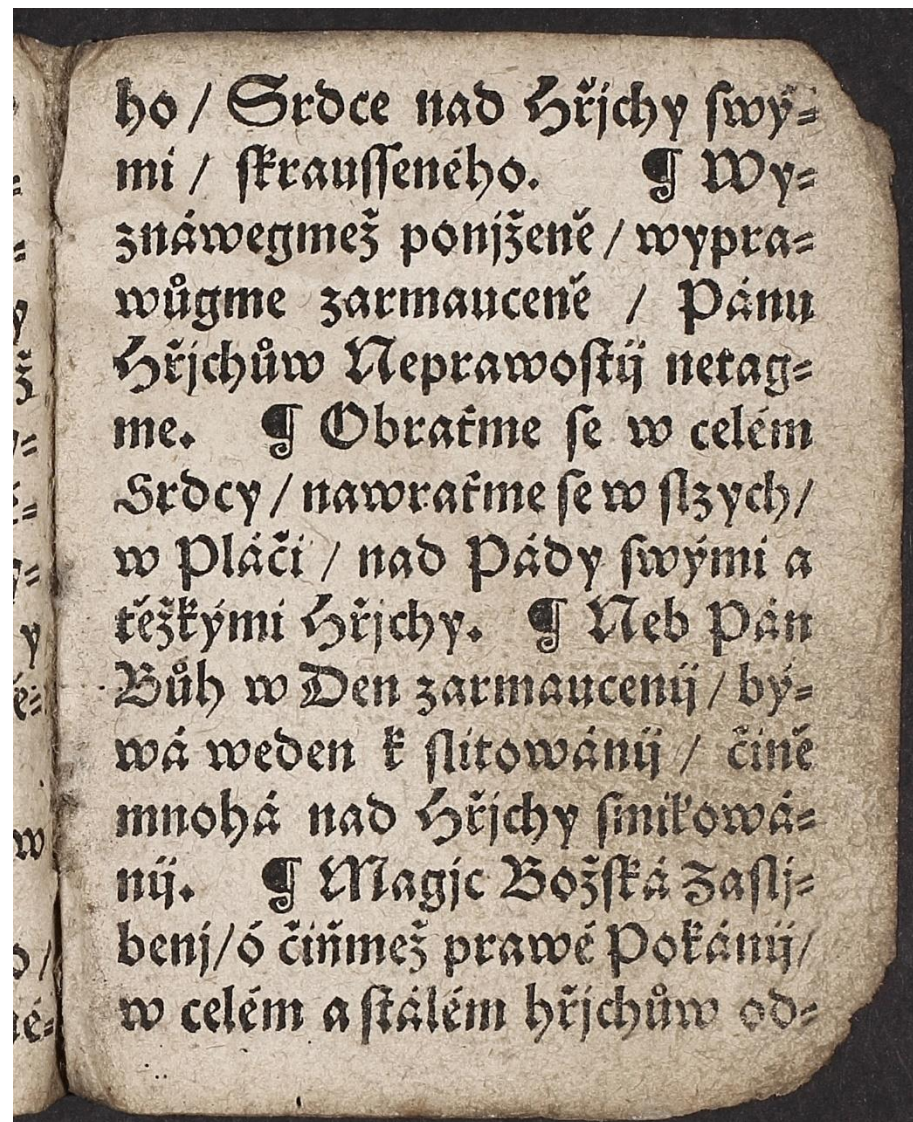


Length of the line – the impact of the pilcrow ¶

		Lit 1650	Lit 1664	OI 1652	OI 1685
with the pilcrow	mean	34.61	32.91	34.82	27.85
	sd	2.05	2.11	2.29	1.98
without the pilcrow	mean	34.26	33.87	34.72	27.22
	sd	2.15	2.25	2.38	2.27

Length of the line – the impact of the pilcrow ¶

- hyphenation



Length of the line – the impact of the pilcrow ¶

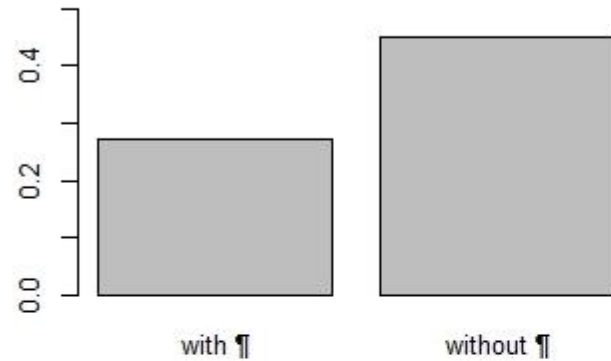
- hypothesis:

The use of the pilcrow reduce to the use of hyphenation

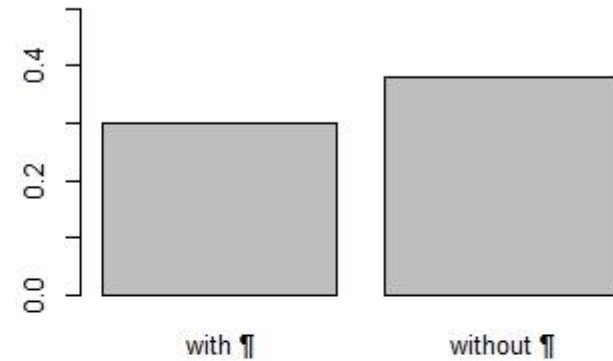
- the proportion of hyphenation in lines with the pilcrow is lower than in the lines without the pilcrow

Proportion of hyphenation

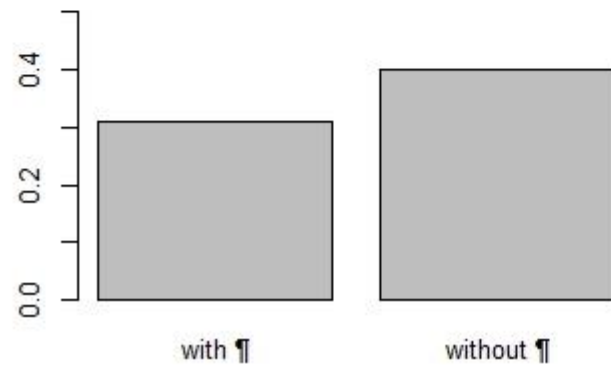
Litomyšl 1650



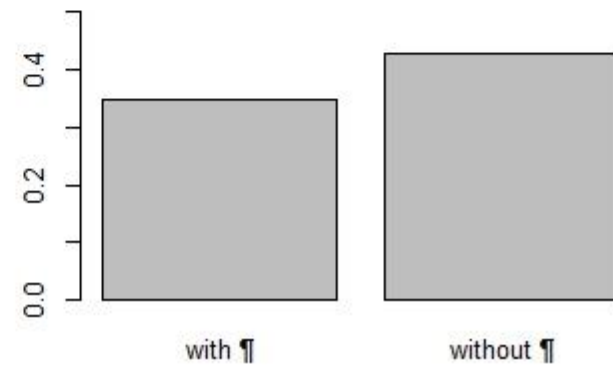
Litomyšl 1664



Olomouc 1652



Olomouc 1685



Capitals

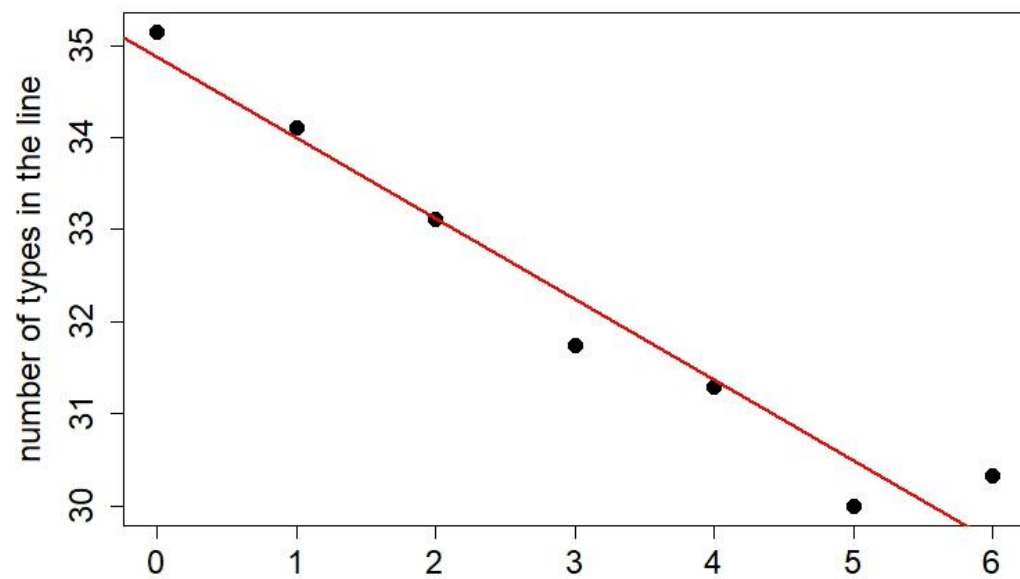
ho / Srdce nad Hřichy swými /
mi / krausseneho. ¶ Wy-
znáwegmež poniženě / wypra-
wüğme zarmauceně / Pánu
Hřichůw Neprawostij netag-
me. ¶ Obratme se w celém
Srdcy / nawratme se w slzych /
w Pláči / nad Pády swými a
těžkými Hřichy. ¶ Ueb Pán
Bůh w Den zarmaucenij / by-
wá weden k slitowanij / čině
mnohá nad Hřichy smilowá-
nij. ¶ Magje Božská Zasli-
benj / ó činmež prawě Pořánij /
w celém a stálém hřichůw od-

Capitals

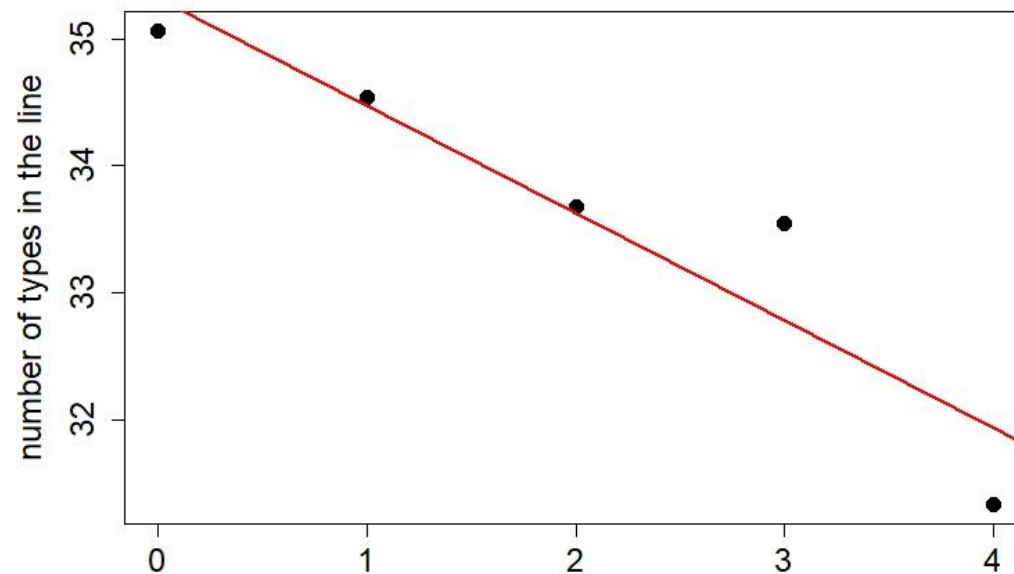
- hypothesis:

The more capitals in the line, the less types in the line

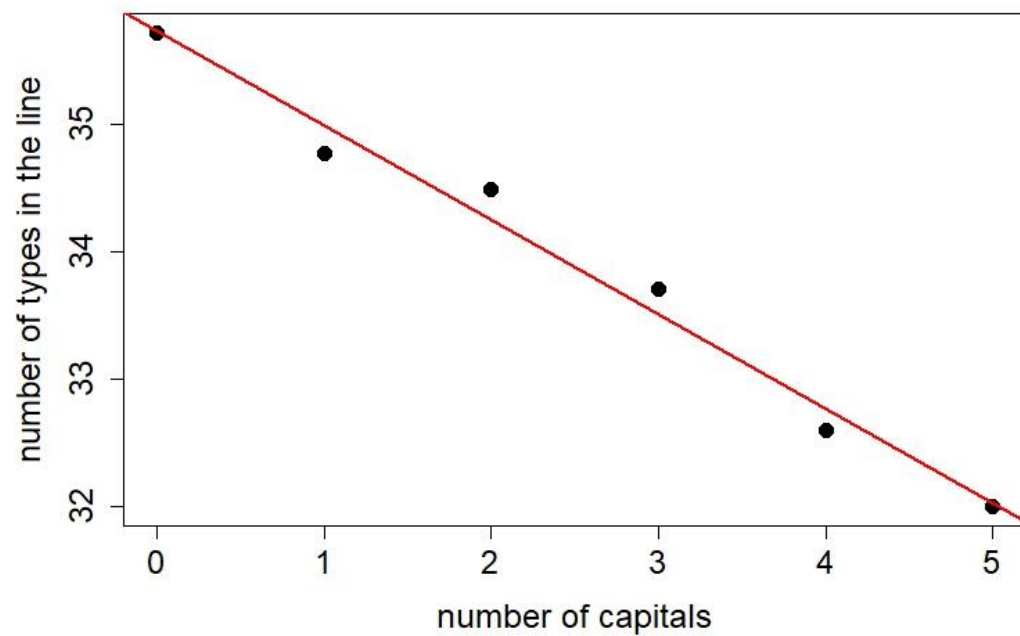
Litomyšl 1664



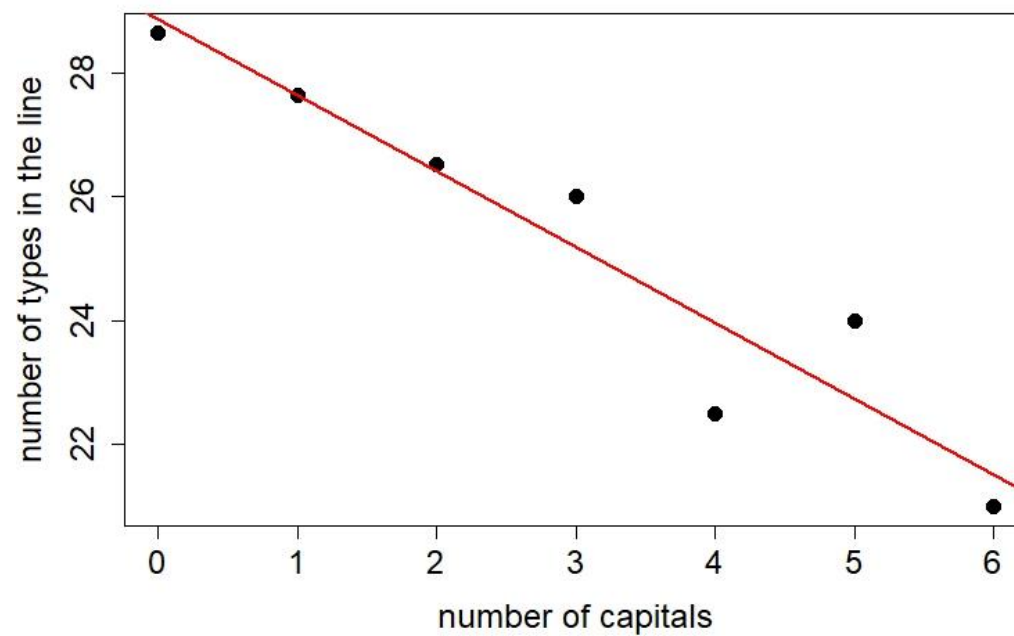
Litomyšl 1650



Olomouc 1652



Olomouc 1685



Summary & further research

- differences among inventories of types
- regular rank-frequency distribution of types
- the relationship between the number of types and number of spaces
- the pilcrow (and multiply spaces) as a tool for fitting the width of the line
- the relationship between the occurrence of the pilcrow and hyphenation
- the role of capitals
- more songs
- analysis of graphs (graphemes)
- determination of width of particular types
- distribution of positions of spaces in the line

