

Michal Místecký

VYUŽITÍ RANKOVÉ FREKVENČNÍ DISTRIBUCE PŘI VÝZKUMU ANTROPONYM V LITERÁRNÍCH DÍLECH

USING THE RANK-FREQUENCY DISTRIBUTION IN ANALYSING
PERSONAL NAMES IN FICTION

ABSTRACT

The paper deals with the proper names of the characters from Lesser Town Short-Stories (Povídky malostranské) by Jan Neruda; it analyses these via the newly designed methods of primary and secondary anthroponym-based text concentrations. They stem from similar formulae which are used in analysing thematic structures of texts, and are based upon rank-frequency distributions of language units (most frequently, words or lemmata). The goal of the article is both to present the methods, and to apply them in the analysis of the short-story collection; the attention will be paid to the relation between the two indexes, the foci of individual stories, and to the anthroponymy structure of the entire book. The counts will be accompanied by sketches of literary-criticism interpretations. The methods could be used in investigating authorial styles and narrative techniques.

KEYWORDS

anthroponymy-based thematic concentration; anthroponym; quantitative linguistics; Jan Neruda

1 Úvod

Kvantitativní zkoumání jazyka nabývá v posledních letech na významu (srov. ČECH – POPESCU – ALTMANN 2014; ANDREEV – MÍSTECKÝ – ALTMANN 2018 aj.); ukazuje se, že celou řadu vlastností textu lze vyjádřit pomocí vzorců s intersubjektivní platností – takový výzkum má pak tu výhodu, že přináší výsledky reprodukovatelné, exaktní, a jsou-li dané metody přijaty, univerzálně použitelné. Srovnání děl se tak

může odehrávat na přesnější rovině; jsou-li vlastnosti (například tematická zatíženost textu, aktivita, slovní bohatství, koheze nebo dekorativnost) kvantifikovány, pak je možné seřadit texty podle míry jejich zastoupení a srovnávat je mezi sebou.

Tato studie se pokusí o aplikaci jedné z těchto metod – tematické koncentrace textu – na antroponymii v literárních dílech.¹ Jejím cílem je jednak uvedenou modifikaci základního vzorce představit, jednak ji použít při analýze konkrétního díla – totiž *Povídek malostranských* Jana Nerudy. Ty byly vybrány především proto, že reprezentují tvorbu zaměřenou na psychologii postav, jíž je v povídkách věnována značná pozornost (např. POHOŘSKÝ 1961, 158); článek si tak klade za cíl prozkoumat jednotlivé Nerudovy texty a zjistit, zdali mezi nimi neexistují rozdíly ve způsobu, jímž autor nakládá se svými postavami. K tomuto účelu bude výzkum využívat jak metod specializovaných tematických koncentrací, tak statistických testů, a konečně i shlukové analýzy. Ve výhledu může představená metoda posloužit v literární vědě, kde ji lze využít ke zkoumání autorského stylu a propojit s výzkumem dalších kompozičních prostředků (např. v naratologické oblasti).

2 Metodologie

Do jaké míry v daném textu dominují antroponyma, lze zjistit různými způsoby; zde bude využito úpravy vzorce, jímž se v kvantitativně zaměřené lingvistice zkoumá téma textu. Základem této metody je ranková frekvenční distribuce, tedy soupis lemmat² seřazených od nejčtenějšího po ta s jedinečným výskytem (tzv. hapax legomena). Postup bude vyložen na příkladu povídky *O měkkém srdci paní Rusky*, protože na ní lze modifikace dobře ilustrovat.

1 Antroponyma, ale v textech publicistických, byla tématem knihy DAVIDA et al. (2013), která pracuje i s metodami tematické koncentrace. Literárnímu užití antroponym se pak komplexně věnuje DVOŘÁKOVÁ (2017).

2 Lemmatem rozumíme konvencí daný základní tvar slova. U podstatných jmen a zájmen se jedná o nominativ singuláru, u adjektiv o nominativ singuláru maskulina a u sloves o infinitiv.



Tabulka č. 1 – Ranková frekvenční distribuce prvních třiceti lemmat v povídce *O měkkém srdci paní Rusky*; tučně je vyznačen h-bod.

Lemma	Pořadí	Frekvence
být	1	64
se	2	63
a	3	60
on	4	45
v	5	40
na	6	31
pan	7	26
paní	8	25
ten	9	25
Ruska	10	18
z	11	16
ale	12	14
by	13	14
mít	14	14
že	15	13
s	16	12
svůj	17	12
jako	18	11
jeho	19	10
Velš	20	10
tu	21	9
co	22	9
stát	23	8
všechno	24	8
také	25	8
k	26	8
komisař	27	8
moci	28	8
funus	29	8
teď	30	7

Tabulka č. 1 zachycuje rankovou frekvenční distribuci uvedeného textu (z prostorových důvodů se omezujeme na prvních třicet lemmat). Na prvních místech se standardně objevují slova neplnovýznamová (např. „být“, „se“, „a“ či „on“), která jsou nezbytná k tomu, aby byl text gramaticky funkční; na spodních příčkách se naproti tomu sdružují autosémantika, která textu dodávají význam. Empiricky bylo zjištěno, že hranicí mezi oběma skupinami je tzv. h-bod (srov. POPESCU 2007; POPESCU et al. 2009); ten odpovídá místu, kde se pořadí kryje s frekvencí, tedy –

$$r = f(r).$$

V našem případě nabývá h-bod hodnoty 14 (lemma „mít“). V případě, že takovou rovnost nelze nalézt, vypočítá se h-bod podle vzorce –

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

v němž r_i je největší rank, pro který platí $r_i < f(i)$, a r_j je nejmenší rank, pro který platí $r_j > f(j)$. H-bod tak představuje jakousi „průměrnou“ hodnotu mezi mezními pozicemi i a j .

Ač lze h-bod interpretovat jako pomyslnou hranici mezi synsémantiky a autosémantiky, v některých textech nad něj pronikají i slova plnovýznamová; takové texty se pak označují jako tematicky koncentrované. Ve frekvenčním seznamu v tabulce č. 1 jsou takovými slovy „pan“, „paní“ a „Ruska“. Popescu a Altmann (2011) kvantifikují míru tematické koncentrace textu tak, že vedle h-bodu (h) zohledňují též rank daného plnovýznamového slova (r'), jeho frekvenci [$f(r')$] a frekvenci slova nejčastějšího ($f(1)$ – tedy slova o ranku 1). Vzorec tematické váhy daného slova pak vypadá takto –

$$TV = 2 * \frac{(h - r') * f(r')}{h * (h - 1) * f(1)}$$

přičemž jmenovatel slouží především k neutralizaci hodnoty TV vůči délce textu. Tematická koncentrace daného textu se pak spočítá jako součet tematických vah všech autosémantik nad h-bodem:

$$TK = \sum TV.$$

V modifikaci, kterou přináší tato studie, se z tematických slov nad h-bodem vyberou pouze antroponyma; způsob výpočtu zůstane stejný. Pracovně se tato úprava nazývá antroponymická koncentrace textu. Obecně lze psát:

$$AV = 2 * \frac{(h - r_a) * f(r_a)}{h * (h - 1) * f(1)},$$

kde r_a je rank antroponyma, $f(r_a)$ jeho frekvence a AV jeho (tedy antroponymická) váha; a dále:

$$AK = \sum AV.$$

V povídce *O měkkém srdci paní Rusky* se nad h-bodem objevuje jediné osobní jméno – „Ruska“. Lze tedy dosadit tyto hodnoty:

$$AV = 2 * \frac{(14 - 10) * 18}{14 * (14 - 1) * 64} = 0,012.$$

Protože je „Ruska“ jediným tematickým antroponymem textu, odpovídá hodnota jeho antroponymické váhy i antroponymické koncentraci celé povídky.

Nově zavedená veličina umožňuje vyčíslit, jak moc jsou povídky zaměřeny na vlastní jména; podobně jako u tematické koncentrace klasické však obecně platí, že počet textů, které by vykazovaly tematická antroponyma, není dostatečný. Jejich počet lze rozšířit zavedením tzv. sekundární tematické koncentrace, která pracuje s dvojnásobnou hodnotou h-bodu ($2h$).³ Sekundární antroponymická koncentrace textu je pak její obdoba, tedy součtem sekundárních antroponymických vah jednotlivých vlastních jmen; matematicky –

$$SAK = \sum SAV = \sum 2 * \frac{(2h - r_a) * f(r_a)}{h * (2h - 1) * f(1)}.$$

Dlužno podotknout, že zdvojnásobení h-bodu nemá žádnou oporu v empiricky vyzkoumaných vlastnostech rankové frekvenční distribuce; jeho účelem je pouze pragmaticky zvýšit počet tematických slov, respektive antroponym.

Dvojnásobkem h-bodu ve zkoumané povídce je hodnota 28; před ní se objevují dvě tematická antroponyma – „Ruska“ a „Velš“. Nejprve je tedy třeba spočítat sekundární antroponymické váhy obou jmen –

$$SAV_{Ruska} = \frac{(28 - 10) * 18}{14 * (28 - 1) * 64} = 0,0134;$$

$$SAV_{Velš} = \frac{(28 - 20) * 10}{14 * (28 - 1) * 64} = 0,0035.$$

Celková sekundární antroponymická koncentrace povídky je pak jejich součtem:

$$SAK_{O\ měkkém} = SAV_{Ruska} + SAV_{Velš} = 0,0162.$$

V analýze Nerudovy povídkové sbírky bude využito obou metod měření antroponymické koncentrace.

3 Metodu poprvé představuje Čech (2016).

3 Výsledky a interpretace

V této části práce budou prezentovány několikery výsledky: nejprve bude pozornost věnována tematickým antroponymům a hodnotám obou koncentrací, které u nich byly naměřeny; dále bude představen korelační diagram srovnávající oba výpočty a konečně bude interpretován graf sledující vývoj hodnot v průběhu Nerudovy sbírky.

Tabulka č. 2 zachycuje výsledky antroponymické koncentrace v jednotlivých povídkách. Už zde se ukazuje, že Nerudova sbírka není co do této vlastnosti homogenní, protože se zde objevují tři texty, v nichž jména postav prominentní nejsou (*Svatováclavská mše*, *Týden v tichém domě*, *U tří lilií*), tři, ve kterých se exponují dvě vlastní jména (*Jak to přišlo...*, *Pan Ryšánek a pan Schlegl*, *Večerní šplechty*), a zbytek, který je více či méně zacílen na psychologický portrét jedince (*Doktor Kazisvět*, *Hastrman*, *Jak si nakouřil pan Vorel pěnovku* atd.). Dá se tak říci, že Neruda užívá různých antroponymických strategií, jimiž ozvláštňuje svou charakterizační techniku: objevuje se zde povídka zcela bez jmen (*U tří lilií*), dualistické střety (*Pan Ryšánek a pan Schlegl*), vzpomínka z mládí (*Svatováclavská mše*), novela s románově rozvětveným dějem (*Týden v tichém domě*) a plejáda postav s velmi slabou tematizací (*Figurky*). Ač je tedy většina povídek u Nerudy výrazně zaměřena na postavy (traktované vlastními jmény), přece jen nelze v práci s nimi nevidět rozdíly.

Tabulka č. 2 - Antroponymické váhy (AV) tematických vlastních jmen ve zkoumaných textech a jejich celková antroponymická koncentrace (AK).

Povídka	Tematická antroponyma	AV	AK
<i>Doktor Kazisvět</i>	Heribert	0,0023	0,0023
<i>Figurky</i>	Provazník	4,27991E-05	4,27991E-05
<i>Hastrman</i>	Rybář	0,0169	0,0169
<i>Jak si nakouřil pan Vorel pěnovku</i>	Vorel	0,0130	0,0130
<i>Jak to přišlo...</i>	Prokop	0,0054	0,0059
	Pohorák	0,0005	
<i>O měkkém srdci paní Rusky</i>	Ruska	0,0124	0,0124
<i>Pan Ryšánek a pan Schlegl</i>	Schlegl	0,0096	0,0155
	Ryšánek	0,0059	
<i>Přivedla žebráka na mizinu</i>	Vojtíšek	0,0231	0,0231
<i>Psáno o letošních Dušičkách</i>	Mára	0,0120	0,0120
<i>Svatováclavská mše</i>	-	0,0000	0,0000
<i>Týden v tichém domě</i>	-	0,0000	0,0000
<i>U tří lilií</i>	-	0,0000	0,0000
<i>Večerní šplechty</i>	Novomlýnský	0,0032	0,0040
	Kupka	0,0009	

Trendy zjištěné u měření antroponymické koncentrace se potvrzují u jejího pandánu, koncentrace sekundární (viz tabulku č. 3); i když rysy některých povídek zůstávají obdobné, zmnožuje se zde však, v souladu s očekáváním, počet tematických vlastních jmen. Výrazně individuální charakter si drží *Doktor Kazisvět* a *Hastrman*, do ostatních příběhů pak vstupují postavy další; bez vlivu jmen zůstává *Svatováclavská mše* a – z důvodů výše zmíněných logicky – *U tří lilíí*. K dualistickým povídkám (*Pan Ryšánek a pan Schlegl*) se přidávají další příběhy, u nichž má však pár jiný náboj (*Jak si nakouřil pan Vorel pěnovku*, *O měkkém srdci paní Rusky*, *Figurky*).⁴ Nově vzniklou skupinou jsou pak povídky zaměřené na větší počet postav (*Psáno o letošních Dušičkách*, *Večerní šlechtty*). Sekundární antroponymická koncentrace tak nebortí obrázek vykreslený prvním výpočtem, ale odhaluje další důležité postavy děl, ač tak někdy znepřehledňuje interpretaci jejich rolí.

Tabulka č. 3 – Sekundární antroponymické váhy (SAV) tematických vlastních jmen ve zkoumaných textech a jejich celková sekundární antroponymická koncentrace (SAK).

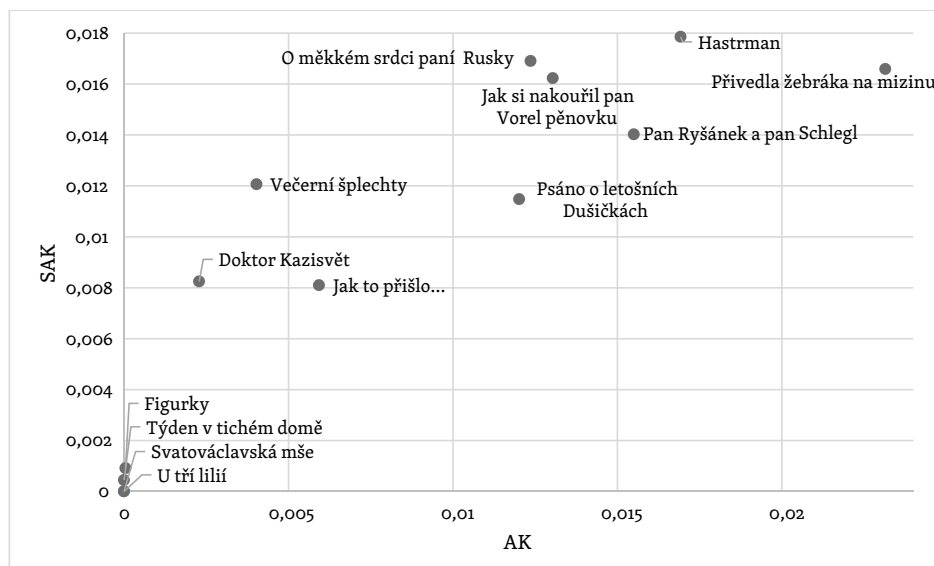
Povídka	Tematická antroponyma	SAV	SAK
<i>Doktor Kazisvět</i>	Heribert	0,0082	0,0082
<i>Figurky</i>	Provazník	0,0005	0,0009
	Pepík	0,0004	
<i>Hastrman</i>	Rybář	0,0179	0,0179
<i>Jak si nakouřil pan Vorel pěnovku</i>	Vorel	0,0157	0,0162
	Poldýnka	0,0005	
<i>Jak to přišlo...</i>	Pohorák	0,0042	0,0081
	Prokop	0,0021	
	Holý	0,0018	
<i>O měkkém srdci paní Rusky</i>	Ruska	0,0134	0,0169
	Velš	0,0035	
<i>Pan Ryšánek a pan Schlegl</i>	Schlegl	0,0077	0,0140
	Ryšánek	0,0064	
<i>Přivedla žebráka na mizinu</i>	Vojtíšek	0,0166	0,0166
<i>Psáno o letošních Dušičkách</i>	Mára	0,0090	0,0115
	Nocarová	0,0016	
	Cibulková	0,0009	
<i>Svatováclavská mše</i>	-	0,0000	0,0000

4 Tyto nově vzniklé dvojice by si zasloužily samostatný rozbor; ve *Figurkách* se například antroponymicky exponují podivínský Provazník a zlobivý hoch Pepík – naproti tomu jméno hlavní postavy, advokátního koncipienta Krumlovského, není v díle zásadněji frekventované. Podobně zajímavý je vztah mezi krupařem Vorlem a jeho jedinou zákaznicí, slečnou Poldýnkou, a panem Velšem, jehož mrtvola je oplakávána paní Ruskou.

Povídka	Tematická antroponyma	SAV	SAK
Týden v tichém domě	Václav	0,0003	0,0004
	Matylda	0,0002	
	Bavorová	1,89359E-06	
U tří lilíí	-	0,0000	0,0000
Večerní šplechty	Novomlýnský	0,0044	0,0121
	Kupka	0,0037	
	Hovora	0,0022	
	Jákl	0,0017	

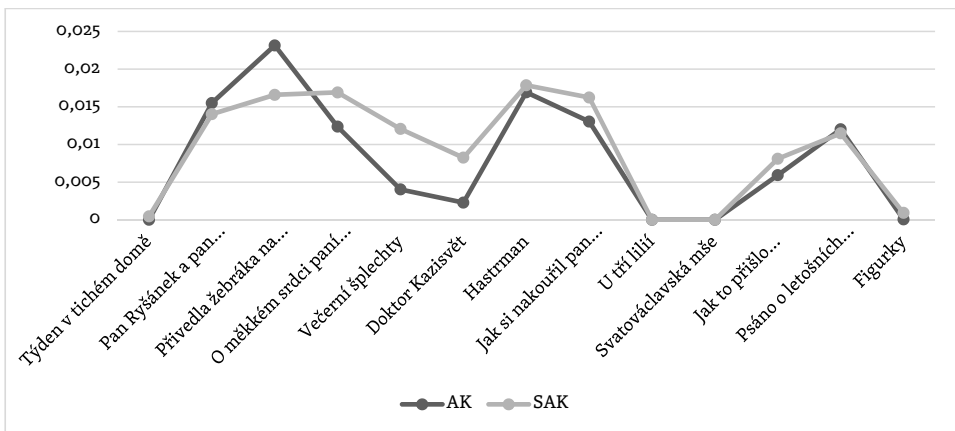
O vztahu primární a sekundární antroponymické koncentrace je možné si udělat představu z grafu č. 1. Jedná se o korelační diagram hodnot, v němž je na ose x umístěna koncentrace primární a na ose y sekundární. Aniž by vztah musel být modelován, je zřejmé, že veličiny jsou na sobě závislé; křivka, která mu odpovídá, bude logaritmická, neboť s přibývajícím hodnotou koncentrace primární (např. u povídek *Pan Ryšánek a pan Schlegl*, *Hastrman* nebo *Přivedla žebráka na mizinu*) se zpomaluje růst koncentrace sekundární. Teoreticky je tak možné předpokládat, že pro velmi vysoké hodnoty primární koncentrace se koncentrace sekundární bude měnit jen minimálně. Zdali je toto obecnou vlastností daných výpočtů, musí prokázat jiná, obecněji zaměřená studie.

Graf č. 1 – Korelační diagram povídek podle hodnot jejich antroponymických koncentrací.



Komplexním výstupem výzkumu je modelace průběhu hodnot antroponymických koncentrací v *Povídkách malostranských* jako celku. Z grafu č. 2 je patrné, že ačkoliv křivka neopisuje žádný jednoduchý matematický model, dá se zde intuitivně vysledovat tendence k modulovanému střídání povídek s nižší a vyšší prominencí vlastních jmen, a to v případě obou indexů. Křivky tak vytvářejí jakési „hrby“, které odpovídají antroponymickým vrcholům sbírky (*Přivedla žebráka na mizinu*, *Hastrman*, *Psáno o letošních dušičkách*), přičemž je vidět, že hodnoty na špičkách postupně klesají (tedy *Hastrman* je méně antroponymicky koncentrovaný než *Přivedla žebráka na mizinu*, *Psáno o letošních Dušičkách* zase méně než *Hastrman*). Z toho vyplývá, že povídky s největším vlastnějmenným zatížením se objevují v první části sbírky; v jejím průběhu pak tato antroponymická fokalizace klesá a ustupuje dějovějším⁵ příběhům vzpomínkového nebo novelistického charakteru (*U tří lilií*, *Svatováclavská mše*, *Figurky*). Lze jen těžko říci, nakolik byla tato struktura Nerudovým uměleckým záměrem, ale matematicky je jasně prokazatelná; může tak přispívat k dynamice sbírky a dovysvětlit její čtenářskou oblíbenost.

Graf č. 2 – Vývoj hodnot antroponymické a sekundární antroponymické koncentrace v textech *Povídek malostranských*.



Poslední interpretační vhléd naváže na předešlé úvahy. Korelační diagram (graf č. 1) byl podroben shlukové analýze; ta umožňuje setřídít povídky do skupin, které vykazují společné vlastnosti (srov. TREVINO 2016). Jako optimální počet těchto shluků se jeví dva nebo tři; vykázané výsledky (tabulky č. 3 a 4) odpovídají intuitivní interpretaci struktury sbírky: povídky z prvního shluku – s vyššími hodnotami obou koncentrací – se objevují především v první části knihy, kdežto v její druhé půlce

⁵ Dějovost lze v kvantitativní lingvistice měřit i pomocí indexu aktivity textu (srov. ČECH et al. 2014, 52–54); bylo by přínosné výsledky obou měření srovnat a zjistit, zdali spolu statisticky významně korelují.

přibývá povídek se střední (shluk č. 2) nebo nižší koncentrací, respektive žádnou (shluk č. 3). Sbírka je pak rámcována dvěma novelami s velmi nízkým antroponymickým zatížením, které odpovídá jejich zaměření na velký počet postav a dějové zápletky. Není tedy úplně pravdou, že by Nerudovy povídky byly jen charakterově zacílené; toto se nejvíce projevuje u šestice textů, které tvoří shluk č. 1. Zvláště texty ze shluku č. 3 však mají zcela jiný, většinou nesourodý charakter (románové náběhy, osobní vzpomínky).

Tabulka č. 3 – Rozdělení *Povídek malostranských* do shluků podle hodnot antroponymických koncentrací.

	Shluk č. 1	Shluk č. 2	Shluk č. 3
Hodnoty	vyšší	střední	nízké
	<i>Hastrman</i>	<i>Doktor Kazisvět</i>	<i>Figurky</i>
	<i>Jak si nakouřil pan Vorel pěnovku</i>	<i>Jak to přišlo...</i>	<i>Svatováclavská mše</i>
	<i>O měkkém srdci paní Rusky</i>	<i>Večerní šplechty</i>	<i>Týden v tichém domě</i>
	<i>Pan Ryšánek a pan Schlegl</i>		<i>U tří lilí</i>
	<i>Přivedla žebráka na mizinu</i>		
	<i>Psáno o letošních Dušičkách</i>		

Tabulka č. 4 – Struktura *Povídek malostranských* prizmatem shlukové analýzy.

Povídky (v pořadí)	Analýza (do tří shluků)
<i>Týden v tichém domě</i>	3
<i>Pan Ryšánek a pan Schlegl</i>	1
<i>Přivedla žebráka na mizinu</i>	1
<i>O měkkém srdci paní Rusky</i>	1
<i>Večerní šplechty</i>	2
<i>Doktor Kazisvět</i>	2
<i>Hastrman</i>	1
<i>Jak si nakouřil pan Vorel pěnovku</i>	1
<i>U tří lilí</i>	3
<i>Svatováclavská mše</i>	3
<i>Jak to přišlo...</i>	2
<i>Psáno o letošních Dušičkách</i>	1
<i>Figurky</i>	3

4 Závěry

Závěry studie budou shrnuty v následujících bodech.

1. Byly zavedeny a vyloženy dvě metody měření antroponymické koncentrace textu. Ty lze využít v literárněvědných analýzách, které zkoumají autorský styl, zaměření děl a strukturu cyklů (např. souboru povídek).

2. Metoda byla aplikována na Nerudovy *Povídky malostranské*. Ukázalo se, že povídky vykazují značnou mírou koncentrace na vlastní jména, ale že jednotlivé texty jsou jí zasaženy v různé míře. Přibližně polovina textu je silně antroponymocentrická, a to v různých konstelacích (psychologický portrét jedince, dualistické povídky, několik výrazných postav). Čtyři texty vykazují hodnoty velmi nízké; kromě dvou novel, které sbírku rámcují, jde o vzpomínkovou *Svatováclavskou mši* a obecně laděný text *U tří lilí*.

3. Sbírka vykazovala strukturaci do tří antroponymických vrcholů, jejichž hodnoty se v průběhu knihy snižují. Větší množství textů s nižšími koncentracemi je typické pro druhou část sbírky; psychologické črty se naproti tomu sdružují do první části. Tuto modulaci nelze jednoduše matematicky modelovat, ale lze ji interpretovat intuitivně.

4. Mezi antroponymickou koncentrací primární a sekundární je zřejmě logaritmický vztah: obě rostou srovnatelně do určitého bodu, od něhož sekundární roste méně. K prokázání trendu by však bylo třeba značně rozšířit zkoumaný korpus.

5. Nabízí se též řada drobnějších otázek teoretických i konkrétních. Především je třeba prozkoumat chování sekundární antroponymické koncentrace vůči primární; kupříkladu se ukazuje, že texty s vysokými hodnotami koncentrace primární (*Pan Ryšánek a pan Schlegl, Přiveďta žebráka na mizinu*) vykazují nižší míru koncentrace sekundární – změna metriky tak do určité míry „řadí“ jejich antroponymické zaměření. I toto je však téma spíše na stať samostatnou a obecněji orientovanou.

LITERATURA

- ANDREEV, Sergey – MÍSTECKÝ, Michal – ALTMANN, Gabriel. 2018. *Sonnets: Quantitative Inquiries*. Lüdenscheid: RAM-Verlag.
- ČECH, Radek – POPESCU, Ioan-Iovitz – ALTMANN, Gabriel. 2014. *Metody kvantitativní analýzy (nejen) básnických textů*. Olomouc: Univerzita Palackého.
- ČECH, Radek. 2016. *Tematická koncentrace textu v češtině*. Praha: ÚFAL.
- DAVID, Jaroslav – ČECH, Radek – DAVIDOVÁ GLOGAROVÁ, Jana – RADKOVÁ, Lucie – ŠÚSTKOVÁ, Hana. 2013. *Slovo a text v historickém kontextu. Perspektivy historickosémantické analýzy jazyka*. Brno: Host.
- DVOŘÁKOVÁ, Žaneta. 2017. *Literární onomastika: Antroponyma*. Praha: FF UK.
- NERUDA, Jan. 1885. *Povídky malostranské* [online]. 2. vyd. Praha: Eduard Valečka [cit. 7.2.2019]. Dostupné z: https://cs.wikisource.org/wiki/Povídky_malostranské

- POHORSKÝ, Miloš. 1961. Jan Neruda. In: POHORSKÝ, Miloš, ed. *Dějiny české literatury III. Literatura druhé poloviny devatenáctého století*. Praha: Nakladatelství ČSAV, 139–187.
- POPESCU, Ioan-Iovitz – ALTMANN, Gabriel – GRZYBEK, Peter – JAYARAM, Bijapu Dayaloo – KÖHLER, Reinhard – KRUPA, Viktor – MAČUTEK, Ján – PUSTET, Regina – UHLÍŘOVÁ, Ludmila – VIDYA, Matummai N. 2009. *Word frequency studies*. Berlin – New York: Mouton de Gruyter.
- POPESCU, Ioan-Iovitz. 2007. Text ranking by the weight of highly frequent words. In: GRZYBEK, Peter – KÖHLER, Reinhard, eds. *Exact methods in the study of language and text*. Berlin – New York: Mouton de Gruyter, 557–567.
- POPESCU, Ioan-Iovitz – ALTMANN, Gabriel. 2011. Thematic Concentration in Texts. In: KELIH, Emmerich. – LEVICKIJ, Victor – MATSKULYAK, Yuliya, eds. *Issues in Quantitative Linguistics, svazek 2*. Lüdenscheid: RAM-Verlag, 110–116.
- TREVINO, Andrea. 2016. Introduction to K-means Clustering. In: *datascience.com* [online]. 6.12. [cit. 7.2.2019]. Dostupné z: <https://www.datascience.com/blog/k-means-clustering>

Michal Místecký

Department of Czech Language

Faculty of Arts, University of Ostrava

702 00 Ostrava

Czech Republic

mmistecky@seznam.cz



Toto dílo lze užít v souladu s licenčními podmínkami Creative Commons BY-NC-ND 4.0 International (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>). Uvedené se nevztahuje na díla či prvky (např. obrazovou či fotografickou dokumentaci), které jsou v díle užity na základě smluvní licence nebo výjimky či omezení příslušných práv.