

Analýza tematické koncentrace textu: komparace publicistiky Ladislava Jehličky a Karla Čapka*

JANA DAVIDOVÁ GLOGAROVÁ – JAROSLAV DAVID – RADEK ČECH

Analysis of the thematic concentration of texts: A comparison of journalistic texts by Ladislav Jehlička and Karel Čapek

ABSTRACT: This study analyzes the thematic characteristics of journalistic texts written by the Czech Catholic intellectual and journalist Ladislav Jehlička (1916–1996) and the writer, journalist and representative of pre-war democracy Karel Čapek (1890–1938). The main aim of the article is to illustrate how Jehlička's pre-war journalism does not correspond to what has been generally accepted by the majority of literary critics – that it is a mere manifestation of his sympathy for nationalism, Nazism, and Fascism. The methodology used was, based on the quantitative characteristics of the text, the analysis of the thematic concentration. The study detected unexpected similarities between both authors; specifically, it revealed that the texts by Jehlička and Čapek are focused on the same problems, e.g. young people and their future, the role of political parties and the circumstances determining a nation's existence.

Key words: quantitative linguistic analysis, thematic concentration, Czech Catholic literature, journalism, democracy, nationalism, ideology, Ladislav Jehlička, Karel Čapek

Klíčová slova: kvantitativnělingvistická analýza, tematická koncentrace, česká katolická literatura, publicistika, demokracie, nacionalismus, ideologie, Ladislav Jehlička, Karel Čapek

1. Úvod

Snad pouze s výjimkou dadaistických textů, případně projevů lidí s mentálním postižením (např. Wernickovou afázií), se naprostá většina jazykových projevů týká nějakého tématu či souboru témat (v nejširším slova smyslu). Běžně se dá mluvit o hlavním tématu (tématech) rozhovoru, knihy, básně, případně o tématech vedlejších. Intuitivně víme, že např. odborné texty jsou tematicky vyhraněnější než beletrie, že ten či onen člověk „měl co říct“, zatímco jiný „vlastně nic neřekl“. Jak však překročit hranici vágního a intuitivního hodnocení týkajícího se tematické charakteristiky textů? Jednou z možností je metoda měření tzv. tematické koncentrace textu, představená Popescem (2007), dále rozpracovaná Popescem et al. (2009), Popescem a Altmannem (2011), Čechem et al. (v tisku) a aplikovaná Wilsonem (2009) a Čechem (v tisku). Jedná se o metodu, jejímž prostřednictvím je možné v textu detekovat tzv. tematická slova, změřit jejich tematickou váhu a nakonec kvantifikovat tematickou koncentraci celého textu. Rozdíl mezi tematickými koncentracemi jednotlivých textů je následně možné testovat pomocí statistických testů (viz část 2).

Metodu analýzy tematické koncentrace textu budeme aplikovat na články katolického publicisty Ladislava Jehličky (1916–1996) z konce 30. let 20. století, které byly

* Text příspěvku vznikl jako součást řešení grantového úkolu GA ČR č. P406/11/0268 Historická sémantika a projektu SGS Ostravské univerzity v Ostravě č. 9/FF/2011 Literatura v širších kulturních kontextech. Za pomoc s automatickým zpracováním textu děkujeme Zdeňkovi Žabokrtskému.

zaměřeny k aktuálním kulturně-společenským a politickým otázkám a které jsou literární kritikou považovány za „kontroverzní“, vypjatě nacionalistické s tendencí k antisemitismu a fašismu. Právě problematičnost Jehličkových raných textů stála u zrodu myšlenky pokusit se porovnat jejich tradiční výklady s výsledky analýzy tematické koncentrace. Metodu analýzy tematické koncentrace jsme navíc použili i pro zkoumání publicistických textů Karla Čapka (1890–1938), jenž představuje autora stejného období, kterému však hodnocení literárních historiků přisuzují zcela odlišné charakteristiky než Jehličkovi. Vystupuje v nich jako představitel demokratického a humanistického proudu české literatury, jako bojovník proti všem formám totalitarismu.

Cílem našeho článku je tedy shrnout dosavadní hodnocení díla Ladislava Jehličky a Karla Čapka (část 3), následně je porovnat s výsledky analýzy tematické koncentrace a zjistit, které z tradičních charakteristik jejich textů z 30. let byly touto metodou potvrzeny a které nikoliv. Srovnáním výsledků analýzy publicistiky těchto dvou autorů ze stejného období, především z druhé poloviny 30. let a v Jehličkově případě také z počátku 40. let, budeme zjišťovat, zda se jejich odlišná ideologická stanoviska odrazila i ve volbě témat, rozuměj ve skladbě tematických slov (část 4).

Přihlédneme-li k tradičnímu čtení publicistiky Ladislava Jehličky a Karla Čapka a k Čechově (v tisku) analýze projevů československých a českých prezidentů, ve kterých se projevuje vztah mezi tematickou koncentrací a ideologií (projevy demokratických prezidentů vykazovaly signifikantně nižší hodnoty tematické koncentrace než projevy prezidentů komunistických), předpokládáme, že Čapkova demokraticky orientovaná publicistika by měla být méně tematicky koncentrovaná než publicistika Ladislava Jehličky. Tento předpoklad bude statisticky testován (část 4.2).

2. Metodologie a jazykový materiál

Měření tematické koncentrace je založeno na analýze frekvenční struktury textu, konkrétně vychází z vlastností tzv. *h*-bodu (srov. Hirsch, 2005; Popescu et al., 2009): pokud seřadíme slova (v našem případě lemmata) daného textu podle frekvence v sestupném pořadí, hodnota *h*-bodu je určena místem, kde se pořadí slova rovná jeho frekvenci, tj.

$$(1) \quad r = f(r),$$

kde *r* je pořadí slova a *f*(*r*) frekvence slova v daném pořadí. Pokud se ve frekvenční distribuci žádná taková hodnota nevyskytuje, pak se *h*-bod vypočítá následujícím způsobem:

$$(2) \quad h = \frac{f(i)j - f(j)i}{j - i + f(i) - f(j)},$$

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*). Pro ilustraci uvedme výpočet *h*-bodu ve frekvenční distribuci lemmat článku Ladislava Jehličky Katoličtí studenti mluví k národu, viz tab. 1.

Tab. 1: Frekvence lemmat v článku Ladislava Jehličky *Katoličtí studenti mluví k národu*, tučně označená lemmata reprezentují tematická slova

pořadí	lemma	frekvence
1	<i>a</i>	39
2	<i>být</i>	26
3	<i>katolický</i>	26
4	<i>ten</i>	23
5	<i>v</i>	23
6	<i>se</i>	21
7	<i>student</i>	17
8	<i>tento</i>	15
9	<i>který</i>	12
10	<i>z</i>	11
11	<i>studentstvo</i>	8
12	<i>právě</i>	8
13	<i>k</i>	8
14	<i>o</i>	8
15	<i>práce</i>	8

Vzhledem k tomu, že v tab. 1 se u žádného lemmatu pořadí nerovná frekvenci, $r \neq f(r)$, použijeme vzorec (2) a získáváme

$$h_{\text{Katoličtí studenti mluví k národu}} = \frac{11 \cdot 11 - 8 \cdot 10}{11 - 10 + 11 - 8} = 10,25.$$

Popescu (2007) a Popescu et al. (2009) ukázali, že h -bod je možné interpretovat jako neostrou hranici mezi synsémantickými a autosémantickými slovy. Autosémantická slova, jejichž pořadí je nižší než h -bod (a tudíž se vyskytují v „oblasti“ synsémantik), jsou slova, která lze díky jejich frekvenčním charakteristikám považovat za slova vyjadřující hlavní téma textu. Vzhledem k tomu, že nositelem tématu jsou obvykle substantiva, budeme za tematická slova považovat právě substantiva a jejich predikáty prvního řádu, tj. adjektiva a verba. Sledujeme-li slova v tab. 1, vidíme, že se zde vyskytují dvě takto definovaná tematická slova, jejichž pořadí je nižší než hodnota h -bodu, a to slova *katolický* a *student*.

Pro porovnání tematické váhy jednotlivých slov i celých textů je nutné hodnotu tematické koncentrace kvantifikovat. Při stanovení indexu tematické koncentrace slova, prezentované Popescem et al. (2009), se vychází jednak z jeho frekvence, jednak z jeho pořadí. Na první pohled je jasné, že čím je hodnota pořadí daného slova menší, tím je vyšší jeho tematická váha. Vyjdeme-li z výše uvedeného předpokladu, že tematická koncentrace textu je reprezentována tematickými slovy nad h -bodem (označme pořadí těchto slov symbolem r'), je možné tematickou váhu charakterizovat jako vzdálenost mezi h -bodem a pořadím slova nad h -bodem vynásobeným jeho frekvencí, tj. $(h-r') \cdot f(r')$. Takto stanovenou tematickou váhu normalizujeme tak, že váhu každého tematického slova vydělíme sumou vah všech slov nad h -bodem a nejvyšší frekvencí lemmatu v textu $f(1)$. Sumu všech vah vypočítáme

$$(3) \sum_{r=1}^h (h-r) = h(h) - \sum_{r=1}^h r = h^2 - \frac{h(h+1)}{2} = \frac{h(h-1)}{2}.$$

Vydělíme-li tematickou váhu slova, tj. $(h-r') \cdot f(r')$, touto sumou, můžeme stanovit index tematické váhy slova TV

$$(4) \quad TV = 2 \frac{(h-r')f(r')}{h(h-1)f(1)}.$$

Pro slovo *katolický* z tab. 1 je tak hodnota indexu tematické váhy tohoto slova

$$TV_{\text{katolický}} = 2 \frac{(10,25 - 2,5)26}{10,25(10,25 - 1)39} = 0,108987.$$

Vzhledem k tomu, že v tab. 1 jsou dvě slova s frekvencí $f = 26$, přičemž se jedná o druhé a třetí slovo v pořadí, je hodnota r' dána průměrem pořadí slov se stejnou frekvencí, tj. $r'_{\text{katolický}} = (2+3)/2 = 2,5$. Tematická koncentrace TK celého textu je pak dána sumou hodnot tematických koncentrací jednotlivých tematických slov, tj.

$$(5) \quad TK_{\text{text}} = 2 \sum_{r'=1}^T \frac{(h-r')f(r')}{h(h-1)f(1)},$$

kde T je počet tematických slov nad h -bodem. V případě článku *Katoličtí studenti mluví k národu* je jeho hodnota tematické koncentrace následující

$$\begin{aligned} TK_{\text{Katoličtí studenti mluví k národu}} &= TV_{\text{katolický}} + TV_{\text{student}} = \\ &= 2 \frac{(10,25 - 2,5)26}{10,25(10,25 - 1)39} + 2 \frac{(10,25 - 7)17}{10,25(10,25 - 1)39} = 0,138871. \end{aligned}$$

Popescu a Altmann (2011) odvodili vzorec pro výpočet variance tematické koncentrace, což je nutné pro statistické testování rozdílů hodnot tematické koncentrace jednotlivých textů;

$$(6) \quad \text{VAR}(TK) = \left(\frac{2}{h(h-1)f(1)} \right)^2 \cdot \left(\sum_{r'=1}^T f(r') \right) \cdot m_{2,r'},$$

kde $m_{2,r'}$ je rozptyl (druhý centrální moment) tematických slov nad h -bodem, tj.

$$(7) \quad m_{2,r'} = \frac{\sum_{r'=1}^T (r' - m_{1,r'})^2 f(r')}{\sum_{r'=1}^T f(r')},$$

kde $m_{1,r'}$ je první centrální moment, tj.

$$(8) \quad m_{1,r'} = \frac{\sum r' \cdot f(r')}{f(r')}.$$

Pro tab. 1 je výpočet variance následující: hodnota prvního centrálního momentu je

$$m_{1,r'} = \frac{2,5(26) + 7(17)}{26 + 17} = 4,27907;$$

na základě toho dostáváme hodnotu rozptylu (druhý centrální moment)

$$m_{2,r'} = \frac{(2,5 - 4,27907)^2 26 + (7 - 4,27907)^2 17}{26 + 17} = 4,84073.$$

Hodnota variance tematické koncentrace textu Katoličtí studenti mluví k národu (tab. 1) je

$$\begin{aligned} \text{VAR}(TK_{\text{Katoličtí studenti mluví k národu}}) &= \\ &= \left(\frac{2}{10,25(10,25 - 1)39} \right)^2 \cdot (26 + 17) \cdot 4,84073 = 0,0000609. \end{aligned}$$

Nyní je již možné pomocí asymptotického u -testu testovat rozdíly mezi jednotlivými texty,

$$(9) \quad |u| = \frac{TK_1 - TK_2}{\sqrt{\text{VAR}(TK_1) + \text{VAR}(TK_2)}}.$$

Pro ilustraci porovnejme texty Ladislava Jehličky Katoličtí studenti mluví k národu ($TK = 0,138871$; $\text{VAR}(TK) = 0,0000609$) a Poučné konce tzv. německého aktivismu ($TK = 0,045588$, $\text{VAR}(TK) = 0,00000552$). Na základě vzorce (9) získáváme

$$|u| = \frac{0,138871 - 0,045588}{\sqrt{0,0000609 + 0,0000055}} = 11,45,$$

což znamená, že mezi tematickou koncentrací obou textů je signifikantní rozdíl (na hladině významnosti $\alpha = 0,05$ je rozdíl signifikantní, jestliže $u > 1,96$).

Pokud chceme porovnávat průměrné hodnoty tematické koncentrace, do jmenovatele vzorce (9) dosadíme namísto $\text{VAR}(TK)$ hodnotu podílu rozptylu průměrů tematické koncentrace s^2 a počtu měření n , tj.

$$(10) \quad |u| = \frac{TK_1 - TK_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

V naší analýze bylo použito 95 publicistických textů Ladislava Jehličky z let 1936–1942 (Jehlička – publicistika, 1936–1942),¹ které byly publikovány v časopisech Akord, Jitro, Obnova² a Řád a 95 publicistických textů Karla Čapka z let 1936–1938 publikovaných v sebraných spisech (Čapek, 1991). Byly vybrány pouze ty texty, které se vyjadřují k aktuálním dobovým politickým a sociálním otázkám. Všechny texty byly automaticky lemmatizovány.

3. Hodnocení Ladislava Jehličky a Karla Čapka v české literární historii

Publicistika Ladislava Jehličky v současnosti stojí na okraji zájmu české literární historie. S připomenutím Jehličkovy tvorby se v literárněhistorických příručkách setkáme spíše sporadicky, a to zejména v souhrnných pojednáních o katolické literatuře

¹ Texty byly vybrány ze souboru čítajícího 150 časopiseckých článků, který vytvořila Jana Davidová Glogarová sběrem a následným přepisem Jehličkových textů v letech 2010–2011. Do budoucna počítáme s jeho zpřístupněním formou kritické edice.

² Časopis Obnova byl od září roku 1938 cenzurován, což postihlo i texty Ladislava Jehličky. Avšak vzhledem k nevelkému rozsahu cenzurních zásahů do Jehličkových článků se nedomníváme, že by tato skutečnost mohla zásadním způsobem ovlivnit výsledky kvantitativní analýzy.

(Hanuš, 2005; Med, 2010; Putna, 2010) či v drobnějších medailonech esejistického charakteru, psaných nejčastěji namísto předmluv k výborům Jehličkových textů (Fučík, 1995; Kautman, 2011; Stankovič, 2000; Stankovič, 2011). Ze všech období Jehličkovy tvorby³ je jeho raná publicistika, uveřejňovaná ve druhé polovině 30. let a na počátku 40. let v časopisech Akord, Jitro, Obnova a Řád, relativně nejméně zmapována. Snížený zájem o Jehličkovy rané práce je způsoben nejen tím, že nebyly dosud edičně zpřístupněny, ale zejména charakterem těchto textů, jimž je obvykle připisován radikální katolicismus (Kautman, 2011, s. 534; Putna, 2010, s. 706; Stankovič, 2000, s. 544), vypjatý nacionalismus a autorovo tíhnutí k totalitarismu (Med, 2010, s. 286). V literárněhistorickém kontextu tak bývají rané Jehličkovy práce hodnoceny jako „nejproblematičtější období jeho publicistiky“ (Stankovič, 2000, s. 544), mající velmi blízko k ideologii italského fašismu (Stankovič, 2000, s. 544) a nacismu (Putna, 2010, s. 708).

Můžeme shrnout, že čtení raných textů Ladislava Jehličky, o jehož popis jsme se pokusili již dříve (Davidová Glogarová, 2011), se v mnohém shoduje se staršími výklady jeho díla. Autorova tvorba z 30. let je poznamenána jeho silně prožívanou katolickou vírou, která prostupuje téměř všemi jeho texty. Nejranější texty, především z roku 1936 a z první poloviny roku 1937, jsou zaměřeny k obecným otázkám společenského dění, jako je problematika školství, otázka budoucnosti mladých lidí, role katolicismu v současném světě, obdiv k osobnostem literárního a kulturního života. Autorovu silně prokatolickou orientaci, s níž komentuje a hodnotí okolní dění, můžeme považovat za jejich jediné ideové a „ideologické“ zatížení. Druhou skupinu Jehličkových článků tvoří politické komentáře, které se z pohledu znalosti kontextu konce 30. let jeví o poznání problematičtější. Jedná se o nacionalisticky laděné texty se sklony k radikální pravici, fašismu, nacismu a antisemitismu.

Ideového protichůdce Ladislava Jehličky, alespoň v hodnocení literárních historiků, představuje Karel Čapek (Bradbrooková, 2006; Buriánek, 1988; Klíma, 2003; Opelík, 2008; Pohorský, 1991).⁴ Zatímco Ladislav Jehlička je zařazován k militantním katolickým autorům, stejně jako Jaroslav Durych, Rudolf Ina Malý, Karel Schwarzenberg, Václav Renč aj., jimž bývá vytýkána přílišná názorová radikalita, kriticismus či pozitivní vztah k nacionalismu a totalitním režimům, Karel Čapek je hodnocen jako osobnost utvářející českou podobu demokracie (Klíma, 2003, s. 6; Opelík, 2008, s. 166; Pohorský, 1991, s. 642) a odmítající všechny projevy nacionalismu (Buriánek, 1988, s. 254; Klíma, 2003, s. 13; Pohorský, 1991, s. 648). V literární historii bývá rovněž připomínáno, že právě Karel Čapek byl za své názory často atakován některými představiteli katolické publicistiky (Pohorský, 1991, s. 648).

Soudobá literární historie řadí Karla Čapka do proudu té části meziválečné české literatury, která myšlenkově souzněla s masarykovským duchem první republiky, brá-

³ Nejčastěji se hovoří o těchto periodách: 1. meziválečná a válečná publicistika (1918–1945); 2. poválečná publicistika (1945–1948); 3. publicistika z 60. let; 4. samizdatová a exilová publicistika ze 70. a 80. let (Chlábec – Petruželková, 2010; Putna, 2010, s. 706–716).

⁴ Česká literární historie se v minulosti detailněji zabývala srovnáním dvojice Karel Čapek a Jaroslav Durych (Med, 2010; Putna, 2003). V případě těchto dvou osobností byla ideová a názorová rozdílnost zdůrazněna také komplikovanými osobními vztahy.

nila ideje názorové plurality i parlamentní demokracie (Med, 2010, s. 17–21; Opelík, 2008, s. 166–175). V tomto směru je Čapek dokonce považován za „reprezentanta a státotvornou osobnost tehdejší republiky“ (Pohorský, 1991, s. 642). V souvislosti s jeho publicistikou z druhé poloviny 30. let bývají nejčastěji zmiňovány analýzy fungování české společnosti, kritika intrikování politických stran i nástupu totalitních ideologií spojených s postupným rozkladem obecně humanitních hodnot (Pohorský, 1991, s. 7, 9). Z tohoto stručného výčtu je zřejmé, že publicistice Karla Čapka jsou přiřazovány zcela odlišné charakteristiky než tvorbě Ladislava Jehličky ze stejného období. V hodnocení obou autorů můžeme nalézt pouze jisté základní tematické shody. Oba publicisté nejčastěji uveřejňovali politicky laděné komentáře spojené s analýzami domácího i mezinárodního dění.

4. Výsledky analýzy tematické koncentrace

4.1. Tematická slova: srovnání politické publicistiky Ladislava Jehličky a Karla Čapka

Celkovou tematickou „zaměřenost“ obou autorů jsme sledovali následujícím způsobem. V každém z textů jsme nejprve určili tematická slova a změřili jejich tematickou váhu (TV). Následně jsme u každého lemmatu sečetli všechny tematické váhy a vydělili jsme je počtem všech analyzovaných článků. Tímto způsobem jsme zjistili průměrnou tematickou váhu daného lemmatu v celém korpusu publicistických textů každého autora. Pro přehlednost je tato hodnota vynásobena 10000 (srov. tab. 2 a 3).

Tab. 2: Tematická koncentrace publicistiky Ladislava Jehličky

lemma	ΣTV	průměrná TV na 1 text TV/95	průměrná TV (10000)	počet textů, ve kterých se lemma vyskytlo
<i>katolický</i>	0,314540	0,00349489	34,95	4
<i>strana</i>	0,168530	0,00187256	18,73	6
<i>člověk</i>	0,156789	0,00165041	16,50	3
<i>český</i>	0,142600	0,00158444	15,84	9
<i>Polsko</i>	0,109405	0,00121561	12,16	3
<i>slovenský</i>	0,084714	0,00094127	9,41	3
<i>propaganda</i>	0,083333	0,00092592	9,26	1
<i>svět</i>	0,077572	0,00086191	8,62	1
<i>německý</i>	0,075683	0,00084092	8,41	3
<i>hospodářství</i>	0,068323	0,00075914	7,59	1
<i>světový</i>	0,062745	0,00069717	6,97	1
<i>století</i>	0,055814	0,00062016	6,20	1
<i>národní</i>	0,054348	0,00060387	6,04	1
<i>kolega</i>	0,047894	0,00053216	5,32	1
<i>syndikát</i>	0,047619	0,00052910	5,29	1
<i>doba</i>	0,046992	0,00052213	5,22	1
<i>student</i>	0,046891	0,00052101	5,21	2
<i>španělský</i>	0,038095	0,00042328	4,23	1
<i>Uhlíř</i>	0,038095	0,00042328	4,23	1

<i>Foerster</i>	0,037694	0,00041882	4,19	1
<i>dítě</i>	0,036058	0,00040064	4,01	1
<i>sedlák</i>	0,027950	0,00031056	3,11	1
<i>politika</i>	0,026881	0,00029868	2,99	1
<i>francouzský</i>	0,026638	0,00029598	2,96	1
<i>Německo</i>	0,024015	0,00026683	2,67	3
<i>volební</i>	0,022222	0,00024691	2,47	1
<i>kniha</i>	0,020979	0,00023310	2,33	1
<i>malý</i>	0,020979	0,00023310	2,33	1
<i>lidský</i>	0,020915	0,00023239	2,32	1
<i>pan</i>	0,020779	0,00023088	2,31	1
<i>země</i>	0,020192	0,00022436	2,24	2
<i>Pekař</i>	0,018495	0,00020550	2,06	1
<i>druhý</i>	0,016049	0,00017832	1,78	1
<i>duch</i>	0,015873	0,00017637	1,76	1
<i>mladý</i>	0,015693	0,00017437	1,74	2
<i>život</i>	0,014499	0,00016110	1,61	2
<i>Guardini</i>	0,012821	0,00014246	1,42	1
<i>velký</i>	0,012472	0,00013858	1,39	1
<i>dílo</i>	0,012422	0,00013802	1,38	1
<i>povstání</i>	0,012171	0,00013523	1,35	1
<i>právo</i>	0,012121	0,00013468	1,35	1
<i>Itálie</i>	0,011738	0,00013042	1,30	2
<i>práce</i>	0,007982	0,00008869	0,89	1
<i>ústředí</i>	0,007982	0,00008869	0,89	1
<i>nový</i>	0,007143	0,00007937	0,79	1
<i>Praha</i>	0,006119	0,00006799	0,68	1
<i>politický</i>	0,005760	0,00006400	0,64	1
<i>škola</i>	0,005342	0,00005936	0,59	1
<i>Francie</i>	0,004651	0,00005168	0,52	1
<i>smlouva</i>	0,004329	0,00004810	0,48	1
<i>stavovský</i>	0,004329	0,00004810	0,48	1
<i>Španělsko</i>	0,004242	0,00004713	0,47	1
<i>Slovensko</i>	0,002584	0,00002871	0,29	1
<i>básník</i>	0,002469	0,00002743	0,27	1
<i>sociální</i>	0,002268	0,00002520	0,25	1
<i>víra</i>	0,002042	0,00002269	0,23	1

Z hodnot uvedených v tab. 2 vyplývá, že k nejsilnějším tematickým slovům Jehličkovy publicistiky z 30. let patřily výrazy spojené s katolictvím (*katolický*), pojmenování odkazující k zájmu o člověka (*člověk*), zejména o výchovu a budoucnost mladé generace (*student, dítě, mladý, život*), a obecná označení spojená s politickým a sociálním děním, a to jak českým, tak mezinárodním (*strana, český, Polsko, slovenský, propaganda, syndikát, svět, německý, hospodářský, světový*).

Shoda výsledků tematické analýzy s tradičním literárněhistorickým čtením se tak projevila především na úrovni připomínaného autora vypjatého katolicismu. Zastou-

pení slov souvisejících s politickými tématy odkazuje na obecné zaměření obsahové skladby Jehličkovy publicistiky a na silný zájem o politické dění spíše než na ideovou orientaci autora. Nacionalismus, tradičně připomínaný jako určující rys jeho textů, můžeme odkrývat za častým výskytem výrazů *český* a *národní*, avšak stejně tak jej lze interpretovat ve vztahu k autorově pozornosti věnované aktuálním problémům české státní existence.

V souboru tematických slov však překvapivě zcela absentují výrazy spojené s připomínanými Jehličkovými sympatiemi k fašismu, nacismu nebo antisemitismu. Pomíne-li vlastní jména evropských států, a to nejen těch, v nichž nabývala na intenzitě některá z pravicově radikálních ideologií (*Polsko, Německo, Itálie, Francie, Španělsko, Slovensko*), či přídavná jména od nich utvořená (*slovenský, německý, španělský, francouzský*), a nezohledníme-li rovněž hodnotu tematické váhy výrazu *propaganda* (9,26), vyskytujícího se však pouze v jednom textu, tematická slova spojená s konkrétními ideologiemi téměř chybí.

Tab. 3: Tematická koncentrace publicistiky Karla Čapka

lemma	ΣTV	průměrná TV na 1 text TV/95	průměrná TV (10000)	počet textů, ve kterých se lemma vyskytlo
<i>národ</i>	0,576056	0,00606375	60,64	10
<i>člověk</i>	0,301707	0,00317586	31,76	8
<i>dušička</i>	0,24	0,00252632	25,26	1
<i>hlas</i>	0,24	0,00252632	25,26	1
<i>lidský</i>	0,237808	0,00250324	25,03	2
<i>stanovisko</i>	0,235294	0,00247678	24,77	1
<i>dítě</i>	0,214986	0,00226301	22,63	2
<i>nový</i>	0,196726	0,00207080	20,71	2
<i>sovětský</i>	0,183333	0,00192982	19,30	1
<i>svátek</i>	0,179272	0,00188707	18,87	1
<i>Kč</i>	0,178571	0,00187969	18,80	1
<i>vývoj</i>	0,13104	0,00137937	13,79	1
<i>statek</i>	0,127273	0,00133972	13,40	1
<i>řád</i>	0,119954	0,00126267	12,63	1
<i>vítězství</i>	0,111111	0,00116959	11,70	1
<i>klid</i>	0,092593	0,00097466	9,75	1
<i>lež</i>	0,090909	0,00095694	9,57	1
<i>ústava</i>	0,085714	0,00090225	9,02	1
<i>krásný</i>	0,083333	0,00087719	8,77	1
<i>politický</i>	0,081818	0,00086124	8,61	1
<i>aršík</i>	0,07326	0,00077116	7,71	1
<i>vidět</i>	0,062766	0,00066069	6,61	1
<i>vysoký</i>	0,051451	0,00054159	5,42	1
<i>ruka</i>	0,049107	0,00051692	5,17	1
<i>bitva</i>	0,046154	0,00048583	4,86	1
<i>matka</i>	0,039216	0,00041280	4,13	1
<i>Kramář</i>	0,037037	0,00038986	3,90	1

<i>Anglie</i>	0,033613	0,00035382	3,54	1
<i>ukázat</i>	0,033613	0,00035382	3,54	1
<i>propaganda</i>	0,033333	0,00035087	3,51	1
<i>německý</i>	0,032468	0,00034177	3,42	1
<i>řici</i>	0,031373	0,00033024	3,30	1
<i>škola</i>	0,030948	0,00032577	3,26	1
<i>chtít</i>	0,028463	0,00029961	3,00	2
<i>výstava</i>	0,027273	0,00028708	2,87	1
<i>věrný</i>	0,025	0,00026316	2,63	1
<i>očista</i>	0,024845	0,00026153	2,62	1
<i>víra</i>	0,02381	0,00025063	2,51	1
<i>směřovat</i>	0,022932	0,00024139	2,41	1
<i>autonomie</i>	0,021663	0,00022803	2,28	1
<i>italský</i>	0,021164	0,00022278	2,23	1
<i>jít</i>	0,020906	0,00022006	2,20	1
<i>mínění</i>	0,020647	0,00021734	2,17	1
<i>veřejný</i>	0,020647	0,00021734	2,17	1
<i>mír</i>	0,019697	0,00020734	2,07	1
<i>strom</i>	0,018699	0,00019683	1,97	2
<i>svět</i>	0,015971	0,00016812	1,68	1
<i>země</i>	0,015023	0,00015814	1,58	1
<i>válka</i>	0,012388	0,00013040	1,30	2
<i>kritický</i>	0,011294	0,00011888	1,19	1
<i>telefon</i>	0,010588	0,00011145	1,11	1
<i>rok</i>	0,009932	0,00010455	1,05	1
<i>mladý</i>	0,008766	0,00009227	0,92	1
<i>kandidát</i>	0,008361	0,00008801	0,88	1
<i>turistický</i>	0,005102	0,00005371	0,54	1
<i>změnit</i>	0,003361	0,00003538	0,35	1

Skladba tematických slov v publicistice Karla Čapka (tab. 3) ukazuje, že k nejsilnějším patří výrazy spojené s národním děním (*národ*) a se zájmem o člověka (*člověk*). Autorovu pozornost věnovanou člověku, nikoliv tedy zaměřeni pouze k politické situaci, naznačují rovněž další tematická slova z okruhu lidského bytí, v jednom případě i ve formě deminutiva (*dušička, hlas, lidský, dítě, krásný, ruka, matka*).

Tato zjištění ilustrují především v literární historii obligátně zdůrazňovaný Čapkův humanismus a zájem o člověka. Politická témata jsou rovněž zastoupena (*ústava, politický, země, válka*), ovšem můžeme říci, že v menší míře než u Ladislava Jehličky. Je zajímavé, že stejně jako u Jehličky téměř chybí výrazy spojené s konkrétními totalitními ideologiemi, rovněž u Karla Čapka nenalezneme tematické slovo, jež by přímo odkazovalo k jeho demokratické orientaci. Naopak, s jeho ideovým protichůdcem Ladislavem Jehličkou jej spojuje zastoupení výrazů souvisejících s konkrétní politickou situací na české i mezinárodní scéně (*sovětský, politický, Kramář, Anglie, propaganda, německý, italský*). Sledujeme-li skladbu tematických slov z tohoto okruhu, můžeme pouze konstatovat, že na rozdíl od Jehličky se u Čapka objevují rovněž tematická slo-

va odkazující ke státům přímo vystupujícím proti zájmům pravicových ideologií, viz výrazy *sovětský*, *Anglie*.

Nejen průměrná hodnota jednotlivých tematických slov, ale také její usouvztažnění k počtu článků, v nichž se tato slova vyskytují, přináší další zajímavé poznatky. V publicistice Karla Čapka do popředí vystupují již zmiňovaná tematická slova *národ* (10 článků) a *člověk* (8 článků), zatímco ostatní tematická slova se vyskytují vždy jen v jednom, případně dvou autorových člancích. Vzhledem k literárnímu hodnocení může být tato skutečnost překvapivá zejména v případě výrazu *národ*, který bychom a priori předpokládali spíše u nacionálně orientovaného Ladislava Jehličky. Z tohoto pohledu můžeme v případě Ladislava Jehličky k takto silným tematickým slovům přiřadit zejména výrazy *katolický*, *strana*, *člověk* a *český*. Především výraz *český*, vyskytující se jako tematické slovo v devíti Jehličkových textech, může potvrzovat jeho nacionalismus. Obecně však lze říci, že ani tento způsob výkladu analýzy tematické koncentrace publicistiky obou autorů neodhalil jejich odlišná ideová východiska. Naopak, ukázal na shody v tematickém zaměření obou autorů, tedy zájem o národní otázku a obecně lidskou problematiku.

4.2. Průměrná hodnota tematické koncentrace

V úvodu jsme vyslovili hypotézu, podle níž by měla být průměrná hodnota tematické koncentrace Jehličkových textů vyšší než v případě textů Čapkových. Vycházíme přitom z tradičního hodnocení díla Ladislava Jehličky, podle kterého je v jeho publicistice daného období patrný silný vliv ideologií, ať je to již výše zmiňovaný radikální katolicismus (Kautman, 2011, s. 534; Putna, 2010, s. 706; Stankovič, 2000, s. 544), vypjatý nacionalismus či totalitarismus (Rataj, 1993, s. 21; Rataj, 1997, s. 36; Med, 2010, s. 286). Předpokládali jsme tedy, že ideologičnost Jehličkova psaní by se měla projevit ve vyšších hodnotách tematické koncentrace. Náš předpoklad se ovšem nepotvrdil, naopak, vyšší průměrnou hodnotu tematické koncentrace vykazují texty Karla Čapka (viz tab. 4).

Tab. 4: Porovnání průměrných hodnot tematické koncentrace u Jehličky a Čapka

	průměrná hodnota TK	$s^2(TK)$	$s^2(TK)/n$
Jehlička	0,02515747	0,001153789	0,00001282
Čapek	0,04903547	0,007453216	0,00007845

Porovnáním průměrné hodnoty tematické koncentrace obou autorů pomocí asymptotického u -testu zjišťujeme, že po dosažení hodnot z tabulky 4 do vzorce (10) dostáváme

$$|u| = \frac{0,02515747 - 0,04903547}{\sqrt{0,00001282 + 0,00007845}} = 2,5,$$

což znamená, že mezi tematickou koncentrací obou autorů je signifikantní rozdíl, protože vypočítaná hodnota u je vyšší než 1,96. Tento výsledek lze dát do souvislosti

s výsledky deskriptivní části naší analýzy (srov. část 4.1). Promítla se v něm skutečnost, že v publicistice Ladislava Jehličky s výjimkou slova *katolický* nenajdeme tematická slova vyjadřující ideologické postoje. Zdá se tedy, že přestože při tradičním čtení z některých Jehličkových textů cítíme autorovy sympatie k radikálním pravicovým ideologiím,⁵ tato skutečnost se výraznějším způsobem neprojevuje na charakteru jeho autorského stylu, tj. ve volbě tematických slov. Naopak, publicistika Karla Čapka, u něhož bychom tendenci k ideologickému myšlení hledali jen stěží, je tematicky mnohem sevřenější a koncentrovanější než publicistika Ladislava Jehličky. Z tohoto pohledu byla naše původní hypotéza, že by tematická analýza mohla odhalit ideologické rysy textu, ve smyslu předpokladu, že čím větší ideologická zaměřenost, tím vyšší hodnota tematické koncentrace, v případě porovnání publicistických textů Ladislava Jehličky a Karla Čapka vyvrácena.

5. Závěry

Závěrem tedy můžeme konstatovat, že z tradičních literárněhistorických charakteristik obou autorů se potvrdil zejména Jehličkův katolicismus a Čapkův humanismus. U obou autorů je patrný zájem o aktuální dobovou politickou situaci. U Čapka se toto zaměření projevuje v menší míře než u Jehličky, ale zato jeho publicistika vyjadřuje větší sílu tematické koncentrovanosti (výraz *národ* s průměrnou hodnotou *TV(10000)* = 60,64 a výskytem v 10 člancích). Je zajímavé, a do značné míry překvapující, že se u Ladislava Jehličky explicitně neobjevily tematicky relevantní výrazy odkazující k jeho inklinaci k fašismu, nacismu či antisemitismu, který mu bývá přisuzován literárněhistorickými hodnoceními; stejně tak však v Čapkových textech nebyly přímé a jasné odkazy k demokracii. Obecně se dá konstatovat, že skladba tematických slov obou autorů vykazuje mnohé podobnosti, což můžeme vysvětlit tím, že se oba dva pohybovali v oblasti publicistiky konce 30. let 20. století a že je spojoval společný okruh témat.

Hypotéza, že by se Jehličkova inklinace k nedemokratickým postojům (vypjatý nacionalismus, totalitarismus) měla projevit ve vyšší míře průměrné hodnoty tematické koncentrace než u Karla Čapka, byla vyvrácena. Ukázalo se naopak, že signifikantně vyšší hodnotu průměrné tematické koncentrace mají právě Čapkovy texty. Tuto skutečnost lze dle našeho soudu vykládat vyhocenou dobovou atmosférou konce 30. let. Čapek i Jehlička tvořili v rámci stejného diskurzu české politické publicistiky, v němž se mísila témata ohrožení, vypjatého nacionalismu i víry v demokracii či zájmu o osud

⁵ Srovnej např. článek Budou se ubíjet mladí lidé? (Národní obnova, 2, 1938, č. 42, s. 2), v němž Ladislav Jehlička píše: „A konečně si v Německu pomohli dokonalým řešením otázky židovské. Tento problém musí být vyřešen i u nás, a to radikálně a rázně. Píšeme o něm na jiném místě. Zde jen připomínáme, že v těchto dnech sdělil Spolek čsl. mediků, že počet židovských mediků a lékařů dosahuje u nás výše 40 procent. To je tedy číslice přímo obhludná, uvážíme-li, že např. medicí-Židé se rekrutují výlučně z bohatých rodin. Numerus clausus (omezení počtu židovských studentů na přirozené procento židovského obyvatelstva, 2–4 procenta) je tu věcí, která přímo řve o okamžité provedení, jež by ovšem zasáhlo ještě nynější fakultní ročníky. Podobná situace, jako u mediků, bude jistě i na právech.“

českého člověka bez ohledu na skutečnost, ke kterému ideovému proudu jednotliví autoři patřili. Zdá se, že atmosféra názorové plurality ovlivňovala jednotlivé publicisty natolik, že se následně projevila i v charakteru jejich autorského stylu. Tak i publicistika autora, jenž je hodnocen jako sympatizant radikální pravice, neobsahuje z pohledu analýzy tematické koncentrace rysy, jež by přímo odkazovaly k fašismu, nacismu či antisemitismu. Signifikantně vyšší hodnoty tematické koncentrace se naopak objevují u autora, který je představitelem prvorepublikové demokratické tradice. Zde vyvstávají otázky: Tvoří opravdu Jehličkova předválečná publicistika homogenní celek ovlivněný pravicovými ideologiemi, nebo v ní lze i z tohoto pohledu hledat jisté posuny, či dokonce na první pohled neočekávané kvality? Jaký vliv sehrává u obou autorů reakce na ohrožení republiky? Jakou roli lze přičítat významu autorského stylu? Částečné odpovědi na tyto otázky přinesou až studie zaměřené např. na srovnání hodnot tematické koncentrace textů 30. let s propagandistickou publicistikou 50. let či srovnání tematických analýz jednotlivých období u daných autorů (např. Čapkovy texty z 20. a 30. let).

Překvapivost zjištěných výsledků tematické analýzy nás ani v nejmenším nevede k myšlence o nutnosti zavrnutí tradičního hodnocení literární historie jako nedůvěryhodného. Přesto však otevírá nové obzory a naznačuje možná východiska i metodologické postupy v oblasti zkoumání autorského stylu či vlivu dobového kontextu na publicistickou tvorbu, které mohou být do jisté míry inspirativní i pro literární vědu či historii.

PRAMENY

- ČAPEK, K. (1991): *Od člověka k člověku III* (= Spisy, 16). Ed. M. Chlěbcová, E. Macek, M. Pohorský. Praha: Československý spisovatel. V elektronické verzi dostupné na [www stránkách Městské knihovny v Praze](http://www.mlp.cz/karelcapek): <<http://www.mlp.cz/karelcapek>>.
- JEHLIČKA – PUBLICISTIKA (1936–1942): Články Ladislava Jehličky v časopisech Akord, Řád, Jitro a Obnova z let 1936–1942.

LITERATURA

- BRADBROOKOVÁ, B. (2006): *Karel Čapek: Hledání pravdy, poctivosti a pokory*. Praha: Academia.
- BURIÁNEK, F. (1988): *Karel Čapek*. Praha: Československý spisovatel.
- ČECH, R. (v tisku): Language and ideology: Quantitative thematic analysis of New Year speeches given by Czechoslovak and Czech presidents (1949–2011). *Quality & Quantity*.
- ČECH, R. – POPESCU, I. I. – ALTMANN, G. (v tisku): Methods of analysis of a thematic concentration of the text. *Czech and Slovak Linguistic Review*, 1 (2).
- DAVIDOVÁ GLOGAROVÁ, J. (2011): Místo Ladislava Jehličky v české katolické publicistice ve 30. letech 20. století. In: E. Příhodová (ed.), *Kontexty a inšpirácie katolíckej moderny*. Ružomberok: Verbum – vydavateľstvo KU, s. 66–89.
- FUČIK, B. (1995): Jeden z nás v bouři. In: V. Binar – M. Trávníček (eds.), *Setkávání a mýjení*. Praha: Melantrich, s. 235–237.
- HANUŠ, J. (2005): *Malý slovník osobností českého katolicismu 20. století*. Brno: CDK.
- HIRSCH, J. E. (2005): An index to quantify an individual's research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (46), s. 16569–16572.

- CHLÍBEC, B. – PETRUŽELKOVÁ, A. (2010): [text na přebalu knihy]. In: L. Jehlička, *Křik koruny svatováclavské*. Ed. B. Chlíbec, A. Petruželková. Praha: Torst.
- KAUTMAN, F. (2011): Láda Jehla: Několik črt k lidskému profilu Ladislava Jehličky. In: L. Jehlička, *Křik koruny svatováclavské*. Ed. B. Chlíbec, A. Petruželková. Praha: Torst, s. 533–535.
- KLÍMA, I. (2003): Karel Čapek – politický novinář. In: I. Klíma (ed.), *Karel Čapek: O demokracii, novinách a českých poměrech*. Praha: Academia, s. 5–13.
- MED, J. (2010): *Literární život ve stínu Mnichova (1938–1939)*. Praha: Academia.
- OPELÍK, J. (2008): *Čtrnáctero prací o Karlu Čapkovi a ještě jedna o Josefu Čapkovi jako přivažek*. Praha: Torst.
- POHORSKÝ, M. (1991): Čapkův optimismus v letech krize. In: K. Čapek, *Od člověka k člověku III* (= Spisy, 16). Ed. M. Chlíbcová, E. Macek, M. Pohorský. Praha: Československý spisovatel, s. 641–656.
- POPESCU, I.-I. (2007): Text ranking by the weight of highly frequent words. In: P. Grzybek – R. Köhler (eds.), *Exact Methods in the Study of Language and Text: Dedicated to Gabriel Altmann on the Occasion of his 75th Birthday*. Berlin – New York: Mouton de Gruyter, s. 555–565.
- POPESCU, I.-I. – ALTMANN, G. – GRZYBEK, P. – JAYARAM, B. D. – KÖHLER, R. – KRUPA, V. – MAČUTEK, J. – PUSTET, R. – UHLÍŘOVÁ, L. – VIDYA, M. N. (2009): *Word Frequency Studies*. Berlin – New York: Mouton de Gruyter.
- POPESCU, I.-I. – ALTMANN, G. (2011): Thematic concentration in texts. In: E. Kelih – V. Levickij – Y. Matskuyak (eds.), *Issues in Quantitative Linguistics*, 2. Lüdenscheid: RAM, s. 110–116.
- PUTNA, M. C. (2003): *Jaroslav Durych*. Praha: Torst.
- PUTNA, M. C. (2010): *Česká katolická literatura v kontextech 1918–1945*. Praha: Torst.
- RATAJ, J. (1993): Krize českého politického myšlení za druhé republiky. In: J. Rataj (ed.), *Z druhé republiky, I, II*. Praha: Historický ústav Armády České republiky – Památník odboje, s. 7–154.
- RATAJ, J. (1997): *O autoritativní národní stát: Ideologické proměny české politiky v druhé republice 1938–1939*. Praha: Karolinum.
- STANKOVIČ, A. (2000): Bořitel moderních českých historických a politických mýtů. In: M. Drápala (ed.), *Na ztracené vartě západu: Antologie české nesocialistické publicistiky z let 1945–1948*. Praha: Prostor, s. 473–480.
- STANKOVIČ, A. (2011): Odešel veselý zpátečník. In: L. Jehlička, *Křik koruny svatováclavské*. Ed. B. Chlíbec, A. Petruželková. Praha: Torst, s. 536–539.
- WILSON, A. (2009): Vocabulary richness and thematic concentration in internet fetish fantasies and literary short stories. *Glottology*, 2 (2), s. 97–107.

Katedra české literatury a literární vědy FF OU

Reální 5, 701 03 Ostrava

<jana.davidova@osu.cz>

Katedra českého jazyka FF OU

Reální 5, 701 03 Ostrava

<jaroslav.david@osu.cz>

<radek.cech@osu.cz>