

## POETIKA STATISTICKÉHO TESTOVÁNÍ HYPOTÉZ

Plecháč, Petr – Kolár, Robert. *Kapitoly z korpusové versologie*. Praha : Akropolis, 2017, 135 s.

Michal Škrabal  
michal.skrabal@ff.cuni.cz



Začněme nikoliv předkládanou knihou, ale předchozí monografií této autorské dvojice (a také Jakuba Říhy jakožto třetího člena versologického týmu ÚČL AV ČR). Ve vstupní kapitole *Úvodu do teorie verše*<sup>1</sup> je versologie definována jako „samostatn[á] disciplín[a] stojící na pomezí literární vědy a lingvistiky“ (s. 7). Z toho plynou i jistá specifika, vydělující tuto literárněvědnou disciplínu z množiny jiných podoborů literární vědy. K jednomu z nich bezpochyby náleží i její metodologicky přirozená inklinace k empiričnosti. Jen málokterá literárněvědná disciplína dokáže natolik funkčně korelovat s korpusově-kvantitativními metodami jako právě versologie. Nejde přitom o žádnou novinku: versologové ochotně využili možnosti, jež s sebou přinesly počítače (v tomto ohledu se autoři hrdě hlásí k výzkumu Miroslava Červenky a Květy Sgallové, tuzemských průkopníků využití výpočetní techniky v humanitních vědách). Leč snahy o formalizaci a větší exaktnost lze ve versologii vyzorovat již mnohem dřív, v éře předpočítačové, a to již od počátku 20. století (ruší formalisté).

Explicitním užitím přívlastku *korpusový* — notabene na místě nejexponovanějším, v titulu — se autoři přihlásili nejen ke svým předchůdcům a vzorům (v českém kontextu je pro *předkorpusovou* versologii výjimečná úloha M. Červenky, mj. učitele R. Kolára), ale rovněž ke korpusu jakožto pozoruhodnému nástroji textové analýzy. Korpusy bývají v povědomí odborné veřejnosti připisovány zejména lingvistice, ve skutečnosti se však neomezují zdaleka jen na ni: nabízejí se všem (nepředpojatým) humanitním vědcům, pouze čekají, až bude jejich potenciál v tom kterém oboru objeven a zužitkován. Tím spíše vítám recenzovanou monografii jako krok správným směrem. Ve své loňské recenzi<sup>2</sup> jsem chválil Richarda Změlíka coby dovedného iniciátora a účastníka mezioborového dialogu mezi literárními vědci a lingvisty; krátce nato můžeme dopřát sluchu také tandemu Plecháč-Kolár. Rovněž jejich hlas je v tomto diskurzu zvukný a hodnotný.

Pravda, je třeba přichystat se na zvýšenou míru specifických metodologických kritérií a postupů, které jsou v jiných literárněteoretických studiích neobvyklé. Uznávají to i sami autoři v úvodu knihy kladouce na čtenáře „nemalé nároky (pro někoho může být obtížné vpravit se do versologické problematiky, pro někoho zase do četných výpočtů a grafů)“ (s. 11). Mohu je však hned zkraje ujistit, že jejich obava je lichá a skutečně „náročnost nepadá na vrub nejasně formulovaných otázek, nepřehledné kompozice nebo nefunkčního jazyka“ (ibid.). Jejich výklad je jasný a logický, výstavba textu dobře strukturovaná a vyvážená. A zvýšená náročnost na recipienta? Ta je legitimní: zvykejme si na to, že podobné metody práce i prezentace jejich vý-

1 Ibrahim — Plecháč — Říha 2013.

2 Škrabal 2016.



sledků se stávají běžným instrumentářem dnešního (přínejmenším) versologa — jde totiž o nutnou daň za možnost (či snad privilegium?) pracovat s daty v objemu přesahujícím časové možnosti jednotlivcovy, ba celých týmů.

Ve stručném<sup>3</sup> úvodu je nastíněno pozadí vydání monografie. Nejen jí dal vzniknout pětiletý grant GA ČR (2011–2015), orientovaný primárně na zkoumání českého verše 19. století. Na jeden grantový projekt se urodil úctyhodný počet výsledků: především Korpus českého verše<sup>4</sup>, materiálová základna pro tento výzkum, a k němu *ad hoc* vytvořený srovnávací korpus prozaických textů ProKop. Mimo tyto dva korpusy vzniklo množství online nástrojů volně dostupných na webových stránkách [www.versologie.cz](http://www.versologie.cz). Rovněž tak byla zveřejněna zdrojová data, „čímž nejen činíme svou práci transparentní, ale především nabízíme všem, kdo o to mají zájem, aby si na jejich základě napsali knihu vlastní“ (s. 13). I v tomto ohledu — co do ověřitelnosti a replikovatelnosti bádání — může práce versologického týmu posloužit za příklad mladým a začínajícím badatelům. Dílčí studie publikované autory v daném časovém údobí pak do značné míry shrnuje a završuje tato monografie, po zbytek recenze se budu věnovat už výhradně jí.

Samotná kniha se skládá z pěti částí, dále členěných do jednotlivých kapitol. V prvních čtyřech oddílech se autoři zaměřují na vybrané rytmické charakteristiky verše, přičemž nezačínáme tradičně od nejmenších jednotek k těm vyšším, ale pohybujeme se opačným směrem: od makrotříd v podobě celých versifikačních systémů přes základní sylabotónická metra a rozměry až po jevy suprasegmentální (slovo, text, interval, rytmický celek, incipit, klauzule) a segmentální (vokalická kvantita, předložková spojení) včetně morfologické roviny verše. Poslední, pátá, kapitola se ubírá opět k vyššímu a abstraktnějšímu celku: pisatelé se v ní, mj. v návaznosti na M. Červenku, snaží definovat tzv. rytmický styl poetického textu, jakožto „výsled[ek] souhry metrických, jazykových a estetických (poetických) norem“ (s. 13). Využívají k tomu vícerozměrné statistiky opírající se o tzv. míru Delta (s. 105), schopnou kvantifikovat mezitextovou podobnost.

Konkrétně si vzali čtveřici básníků — Máchu, Heyduka, Zeyera a Gellnera —, jejichž tvorba se z hlediska rytmického výrazně odlišuje (autoři vybrali incipit a klauzuli mužského jambu a tyto dvě proměnné jim tvorbu všech čtyř spolehlivě oddělily). Po přidání dalších dvou tvůrců (Kvapila a Klášterského) si už však s dvojrozměrným prostorem (resp. grafem) nevystačíme: je potřeba přijít s další proměnnou, v tomto případě s podílem jednoslabičných R-celků v monosylabických incipitech (s. 108, 56–58); spolu s ní však vstupuje do hry i třetí rozměr, v němž jsou jednotlivé sbírky reprezentovány vektory. A nemusíme zdaleka setrvat jen v třídimenzionálním prostoru. Každá následující proměnná, nesoucí s sebou ovšem nutně další rozměr, přispívá k přesnějšímu vydělení *x*-terých básnických individualit. Vícerozměrné grafy však narážejí na naše běžné kognitivní bariéry. Autoři se proto snaží zobrazit vzajem-

3 Ke kladům recenzovaného díla patří mj. i to, že netrpí mnohomluvností a podobně jako *Úvod do teorie verše* je napsáno koncentrovaným a hutným jazykem.

4 Jedná se o korpus české poezie 19. a počátku 20. století, automatickými nástroji lemmatizovaný a anotovaný, přičemž fonetickou, metrickou a strofickou anotaci (kromě toho je tu ještě anotace morfologická) zajistil *ad hoc* vytvořený systém Květa. Autoři deklarují úspěšnost tohoto systému v rozmezí 95–99 % (s. 10).



nou podobnost/odlišnost mezi jednotlivými sbírkami — aniž by se omezovali počtem proměnných — prostřednictvím shlukové analýzy (s. 105nn.). Nejčastěji se tak děje za pomoci tzv. dendrogramu (s. 112–113), který jednotlivé autory shlukuje do nadosobních celků, jež bývají tradičně chápány jako básnické školy. Bude jistě zajímavé zjistit, zda tradiční klasifikace, tak jak ji známe ze středoškolských hodin literatury<sup>5</sup>, odpovídá i výsledkům korpusově-kvantitativní analýzy. Vzhledem k subjektivnosti vymezení takovýchto tříd (oproti např. stanovení celku básnického díla konkrétní tvůrčí osobnosti) se shluková metoda jeví být vhodným korektivem intuitivně vytvořeného abstraktu: může jej — s oporou v objektivních a změřitelných charakteristikách a dostatečně robustním množstvím textového materiálu — nejen verifikovat, ale též vést k jeho reklasifikaci.<sup>6</sup> Do úvahy je však třeba vzít také diachronní aspekt, postihující vývoj autorského rytmického stylu (a jeho potenciální příslušnost k více-rym básnickým školám v průběhu času). Jak elegantně, byť jen pomocí dvojrozměrné časové osy, lze tento vývoj zobrazit, ukazují autoři názorně na s. 117.

Mimoto může rytmický styl jednotlivých autorů fungovat ještě jako (nevědomé) poznávací znamení, pomyslný „otisk autorova prstu“ vedoucí k jeho odhalení — v tomto ohledu je shlukovou analýzu možno využít k určování autorství.<sup>7</sup>

Nerad bych, aby poslední odstavce vzbuzovaly dojem, že první čtyři kapitoly knihy co do informační hutnosti a vědecké objektivnosti za tou poslední zaostávají. Naopak, leitmotiv v podobě ověřování konkrétních versologických hypotéz, tradovaných z dřívějška či nově formulovaných přímo autory, se vine napříč celou knihou. Tento „důsledný přechod od kvantitativního popisu zkoumaných jevů ke statistickému testování hypotéz“<sup>8</sup> (s. 12, srov. s. 37) je umožněn díky rozsáhlému korpusovému materiálu a dnešním technickým možnostem. Sami autoři tuto část svého výzkumu v úvodní kapitole označují jako „progresivní“ (s. 12), za jeden z přínosů celé knihy pak považují „formulování otázek tak, aby je bylo možné buď vyvrátit, nebo potvrdit“ (s. 13).

Tvůrčím způsobem rozvíjejí výzkum svých předchůdců, polemizují se staršími tezemi a hypotézami, na něž je nahlíženo čerstvým prizmatem. Nové poznatky jsou přesvědčivě formulovány, např.: „[...] versologie by se neměla zabývat distribucí přízvuků (to náleží spíše teorii recitace), ale distribucí objektivně pozorovatelných jazykových jevů, které pravděpodobnost výskytu přízvuku ovlivňují“ (s. 46). Příkladem takového jevu je vzápětí vyložený (a pregnantně definovaný) rytmický celek (R-celek), chápáný jako „objektivně definovatelný protějšek mezipřízvukového intervalu“ (s. 47).

Na závěr chci zmínit, ba pochválit grafickou podobu díla. Tvůrci vsadili na typografii, jež se jim osvědčila minule, výrazně však přibylo grafů a vzorců. Obzvlášť

5 V daném časovém období (od konce 18. stol. do konce první třetiny 20. stol.) vydělují autoři 10 takovýchto celků: thámovci, puchmajerovci, preromantici, romantici, májovci, ruchovci, lumírovci, dekadenti, Katolická moderna, proletářská poezie (s. 12).

6 Podobná analýza je aktuálně testována kolektivem pracovníků ÚČNK na českém jazykovém materiálu mj. s cílem automatické vnitrotextové žánrové klasifikace textů (Cvrček a kol. [V tisku]).

7 Srov. Plecháč — Flaišman 2017.

8 Celkově jich je 21, z nich 17 se potvrdilo, čtyři autoři vyvrátili (s. 123–124).



šikovná je práce s jednou jedinou barvou (posledně petrolejová modř, zde trávová zeleň) — i ta stačila k dostatečnému kontrastu s černou (případně šedou), skvěle se vyjímá v grafech a stylově „vytučňuje“ podstatné pasáže knihy. Absence překlepů a podobných nešvarů, jinde hojných, svědčí o svědomité a poctivé redakci (Jakub Říha).

Z názvů obou monografií je patrná jistá ostýchavost a respekt před úkolem, jenž oba badatele čeká a který je věru nelehký a nezáviděníhodný: od *kapitol*, respektive *úvodu* dospět k velkému, syntetickému dílu. Oba pro to nicméně mají ty nejlepší předpoklady: dobrý vklad od svých učitelů, dostatek materiálu, ale též nápadů a vědeckého entuziasmu. A v neposlední řadě i smysl pro humor (ukázky č. 10 a 14 na s. 51 a 53, do nichž se potutelně zvětšili — a mám takové tušení, že jinde i své kolegy z pracoviště). Nechybí jim ani odvaha překročit úzký okruh tuzemské vědy, prezentovat své výsledky na mezinárodním poli a své metody testovat rovněž na cizojazyčném materiálu.

## LITERATURA

Cvrček, Václav a kol. „Variabilita češtiny: multidimenzionální analýza“. [V tisku].

Ibrahim, Robert — Plecháč, Petr — Říha, Jakub.

Úvod do teorie verše. Praha : Akropolis 2013.

Plecháč, Petr — Flaišman, Jiří. „Problém Barák-Neruda z pohledu současné stylometrie“.

Česká literatura, 2017, roč. 65, č. 5, s. 743–769.

Škrabal, Michal. „Jan Čep z kvantitativní perspektivy“. Česká literatura, 2016, roč. 63, č. 4, s. 605–609.