

**From Punch Cards to Modern Corpora: Quantification and Statistics
as Tools for Literary Studies?**

Traditionally, methods of quantitative and corpus analysis have chiefly been applied in linguistics. Only in recent years have (mainly foreign) studies tested their possibilities in the sphere of literary studies. The main purpose of such an approach is to provide literary studies with an exact tool, one based on formally analysed verbal material and not burdened with subjective introspective aspects. The principle of quantitative-corpus analysis does not only comprise the schematised taxonomy of a verbal corpus – it also presupposes its categorical statistical analysis, which is capable of manifesting (representing) absolute structural values and qualities.

Keywords: Quantitative Analysis, Corpus Analysis, Statistical Analysis, Taxonomy, Verbal Corpus, Literary Studies

Mgr. Richard Změlík, Ph.D.
Uniwersytet Wrocławski
Instytut Filologii Słowiańskiej
Pocztowa 9
53-313 Wrocław
richard.zmelik@uwr.edu.pl

OD DĚRNÝCH ŠTÍTKŮ K MODERNÍM KORPUSŮM: KVANTIFIKACE A STATISTIKA JAKO NÁSTROJE LITERÁRNÍ VĚDY?

RICHARD ZMĚLÍK

1.

Literárněvědné myšlení prošlo za poslední desetiletí řadou změn, jejichž výsledkem je nemalé množství rozdílných směrů, konceptů, metodologických přístupů a kritérií. Zejména v posledním desetiletí lze zaznamenat výraznější uplatnění kvantitativních a korpusových metod také v prostředí literární vědy, především literární historie. Bez nároku na absolutizaci konstatování si dovolíme tvrdit, že ikonickým příkladem takového přístupu je práce Franca Morettiho, jenž v roce 2005 publikoval významnou vědeckou monografii, která teprve nedávno byla přeložena do češtiny pod názvem *Grafy, mapy, stromy: abstraktní modely literární historie* (2014).¹ V ní autor představil tři různé způsoby, kterými lze kvantifikaci a formální analýzu dat, jimiž jsou myšleny literární texty, smysluplně využívat v literární vědě. Prosazování takového přístupu, který si koncem 20. století získával oporu především v lingvistice, také v literární vědě bylo poměrně komplikované, neboť naráželo na odpor tradičních představ o tom, jakými metodami je umělecké dílo možné zkoumat. Přesto Morettiho práce (a v důsledku zdaleka nejen jeho) významně iniciovala odbornou diskuzi, jež je dnes nedílnou součástí nově se prosazujícího směru v humanitním bádání, o kterém se hovoří jako o tzv. Digital Humanities.²

Aktuální situace v této disciplíně, zejména pak množství odborných prací, představuje široké spektrum různě orientovaných a definovaných problémů, mezi kterými má jedno z dominantních postavení také otázka atribuce textů. Naším úkolem není obsáhnout veškeré možnosti, které jmenovaná disciplína literární vědě nabízí, ale spíše na vybraných přístupech z českého i zahraničního kontextu demonstrovat rámcové vývojové etapy,

/1/ Zde Moretti definuje svoji metodu, kterou označuje jako distant reading: „*Distant reading* (distanční čtení), jak jsem toto pojetí již jednou napůl v žertu, napůl ovšem vážně nazval, je přístupem, pro nějž vzdálenost neznamena překážku, nýbrž specifickou formu poznání. Vzdálenost sice pomíjí jednotlivé detaily, zato však odhaluje jejich širší propojení, jejich vztahy, vzorce a formy.“ (Moretti 2014: 7–8) Nejnovější Morettiho publikací, ve které uvedenou problematiku dále rozpracovává, je kniha s názvem *Distant reading* (2013).

/2/ Digital Humanities zahrnuje široké spektrum využití digitálních technologií v humanitních vědách. V případě umělecké literatury se užívá pojem Digital Literary Studies. Ovšem i pod tímto označením se kromě jmenovaných kvantitativních metod v literární vědě objevuje rozsáhlejší oblast spojující uměleckou literaturu a digitální média. V českém literárněvědném prostředí je aktuálně diskutována problematika literatury prezentované prostřednictvím elektronických médií, zejména internetu (viz příspěvky šestého čísla časopisu *Česká literatura* z roku 2015).

kterým myšlení o literatuře založené na kvantitativních a statistických metodách procházelo.

2.

V humanitních vědách orientovaných na jazyk je kvantitativní a korpusová analýza tradičně spojována s lingvistikou. Kvantitativní lingvistika jako samostatný obor se začala výrazně prosazovat v souvislosti s rozvojem počítačové technologie zejména od počátku druhé poloviny 20. století. Lingvisté brzy dokázali využít možností rychle se rozvíjejícího moderního oboru – kybernetiky, a to zejména v oblasti strojového zpracování lexikonu. První pokusy o počítačem sestavené slovníky využívaly tzv. děrnoštítkovou metodu. V českém prostředí je takovým příkladem autorský slovník Jitky Štindlové *Konkordanční a frekvenční index k Slezským písním Petra Bezruče* (1969). Podobné práce vznikaly i v zahraničí, např. autorský slovník Leslieho Hancocka *Word Index to James Joyce's Portrait of the Artist* (1967) nebo *A Concordance to the Plays and Prefaces of Bernard Shaw* (1971) Earla D. Bevana. Výzkumy tohoto typu se pochopitelně zpočátku prováděly na literárních textech, které v předkorpusem období byly spolehlivým zdrojem jazykového materiálu. Výzkum však byl primárně orientován lingvisticky a jeho úkolem bylo na základě exaktní strojové kvantifikace demonstrovat konkrétní obecné charakteristiky a zákonitosti lexikonu, mezi které náleží hodnoty jako bohatství, rozptyl (disperze), koncentrace slovníku nebo vztahy mezi frekvencí lexikálních jednotek a jejich kvalitativní charakteristikou a distribucí v rámci jednotlivých frekvenčních pásem apod.

Uvedené otázky na materiálu české prózy systematicky zkoumala především Marie Těšitelová, která spolu s Josefem V. Bečkou a Josefem Jelínkem v roce 1961 vydala *Frekvenční slovník češtiny*. Základní metodologické postupy spojené s kvantitativní analýzou lexikonu Těšitelová formulovala v publikaci *Otázky lexikální statistiky* (1974), kde výzkum založila rovněž na analýze konkrétních literárně uměleckých textů. Určitý potenciál, jež kvantitativní metody nabízí literární vědě, konstatoval u nás Lubomír Doležel již v roce 1965 ve studii *Pražská škola a statistická teorie básnického jazyka*. Doleželova argumentace směřuje zejména k otázkám funkčního využití statistiky při zkoumání textové normovanosti a aktualizace. Formální výsledky analýzy lze aplikovat na rozlišování básnického a ne-básnického jazyka nebo využít pro stanovení míry konstantnosti a variabilnosti jazykových prostředků v uměleckých textech (viz Doležel 1965: 105–106). Doleželovy úvahy se přirozeně opírají o strukturalistické základy a mimo jiné rovněž dokládají úzký vztah mezi strukturální metodologií a kvantitativní analýzou:

Pražská teorie básnického jazyka byla již charakterizována jako teorie statistické povahy. Ačkoli tato charakteristika je v podstatě správná, přece nelze

zapomínat, že v pražských pracích nalézáme rovněž velmi cenné analýzy kvalitativně strukturních vztahů v básnickém textu. [...] Na druhé straně musím zdůraznit, že pražská teorie básnického jazyka má charakter prestatistický, protože zde nebyl aplikován ani ve vymezení, ani v analýze básnického jazyka skutečný aparát matematické statistiky. Hlavní věty teorie jsou statistické svým obsahem, nikoli svou formulací.

(Tamtéž: 103–104)

Zejména poslední tvrzení je důležité, neboť odkazuje k tomu, jakým způsobem byly kvantitativní metody reflektovány v raných strukturalistických pracích. Doležel upozorňuje na rozdíl mezi kvantifikací ve smyslu vlastní metody, která se řídí nejen specifickými postupy, ale determinuje výslednou povahu analýzy, a kvantifikací ve významu pouhého doplnění metodologicky ovšem jinak zaměřeného výzkumu. V druhém případě slouží statistické informace jako příloha nekvantitativního výzkumu.

Ačkoli uvedené otázky, a zejména pak výsledky práce Marie Těšitelové vybízely k testování kvantitativních postupů v kontextu literární vědy, k podobným pokusům nedošlo. Veškeré formule tak zůstaly pouze na teoretické rovině, což dokládá i studie Miroslava Klivara *Význam kvantitativní lingvistiky pro literární teorii* (1984), ve které autor de facto opakuje výsledky M. Těšitelové. Pouze na několika úvodních místech Klivar konstatuje některé velmi obecné skutečnosti:

Prokazatelné výsledky kvantitativní lingvistiky naznačují řadu podnětů pro literární vědu. Kvantitativní lingvistika byla definována jako vědecký obor, jehož předmětem je zjišťovat statistické zákonitosti jazykových jevů a jejich vztahů a získávat statistické údaje o jednotlivých jevech a na jejich a na jejich základě charakterizovat analyzované celky. [...] Z hledisek teorie literatury jde pak o další nadsystém, v němž lingvistika je jedním ze subsystémů vědecké soustavy výzkumných oborů a metod podřízených literárně teoretickému „supersystému“. (Klivar 1984: 24)

I když Klivar sympatizuje s možností aplikovat kvantitativní postupy do literární vědy, současně si uvědomuje hranice těchto možností. Především se jedná o skutečnost, že kvantitativní analýza se bezprostředně dotýká jazykové roviny. I přes nepopiratelný fakt, že jazyková vrstva tvoří s „vyššími“ oblastmi (tematika, motivika apod.) strukturní výstavby uměleckého díla funkčně koherentní jednotu, podle Klivara vzniká otázka, nakolik bude kvantifikace schopna referovat také k těmto oblastem:

Na druhé straně si uvědomujeme určité relativní hranice, v nichž se kvantitativní lingvistika pohybuje a které nemůže překročit. Za prvé je to jednak její základní zakotvení v lingvistické oblasti, z čehož plyne její vztah k širším nad-

systémům v teorii literatury zcela bezprostředně a jasně. Za druhé je to její těžiště v kvantitativní stránce metod výzkumu, jak jsme viděli sice v jednotě s kvalitativním hlediskem (již ve výběru materiálu pro výzkum), ale v jednotě kvantity a kvality je těžiště v kvantitativní stránce vztahu této „strojové“ lingvistiky, jak bývá také někdy nazývána. Za třetí je z hlediska teorie literatury zjevné, že uvedená specifika jsou jen dílčí v celkové koncepci tří vrstev uměleckého díla – hmotné, významové a obsahové. Je tu řada dalších specifik, například ve významové vrstvě vztahy času vnitřního a vnějšího aj., které se projevují ve významové vrstvě jinak než ve hmotné, jsou tu hodnoty ideologické, filozofické, složitý souhrn hodnot mimoestetických, které nesouvisejí se zákonitostmi zkoumanými kvantitativní lingvistikou.

(Tamtéž: 34)

Klivarovy námitky jsou pochopitelné; zcela odpovídají situaci, ve které se kvantifikace (uměleckého) lexikonu nacházela v 60. a 70. letech 20. století. Práce M. Těšitelové, ze kterých Klivar vychází, přes veškerý potenciál, jenž nabízí pro domýšlení a rozvedení v prostředí literární vědy, zůstávají v prvé řadě lingvistickými analýzami soustředěnými na problematiku lexikonu. Klivar si uvědomuje, co kvantitativní analýza lexikonu literární vědě nabízí, avšak nevytváří funkční překlenutí mezi oběma disciplínami. K tomu dochází teprve koncem 20. století.

3.

Ve druhé polovině 20. století byly otázky týkající se možností kvantitativní analýzy v literární vědě promýšleny spíše hypoteticky, její hlavní vývoj byl orientován zcela jiným směrem. Od 60. let se v literární teorii výrazně prosazuje naratologie, která přichází s vlastními otázkami a problémy. Krize strukturálního myšlení, se kterým je kvantifikace úzce spojena, se reflektuje například v dekonstrukci. Současná literární teorie disponuje rozsáhlou škálou tematických okruhů, jimiž věnuje pozornost – gender studies, nový historismus, nová cenzura, reprezentace paměti, postkoloniální studia nebo výrazný příklon k otázkám prostoru, jež nabývají na aktuálnosti zejména od konce 80. let.³ Spaciálně orientované bádání dnes představuje jeden z výrazných směrů (nejen) literárněvědného myšlení, jež postupně opouštělo vyhraněnou oblast metodologických diskuzí a přecházelo směrem k otázkám spíše ideovým či ideologickým a interdisciplinárním. Ačkoli se na různých úrovních lingvistika a literární věda stále setkávaly, obě disciplíny se vyvíjely samostatně, což je na druhou stranu jedním z hlavních znaků svébytnosti vědních oborů.

/3/ Viz Bachmann-Medick, 2006: 286n.

Kvantitativní a formální analýza založená na strojovém algoritmu se znovu stala aktuální na počátku 90. let, kdy začínají vznikat moderní jazykové korpusy.⁴ Teprve nedávné výzkumy potvrzují, že literární věda a lingvistika ve spojení s moderní technologií mohou tvořit užší funkční celek. Tato možnost spočívá v dnešních moderních elektronických jazykových korpusech. Spojení výsledků, které nabízí korpusová a moderní kvantitativní lingvistika, s literárněvědným bádáním se ovšem ne vždy setkává s porozuměním. Jedním z kritiků spojování kvantitativně-korpusové lingvistiky a literární vědy je i holandský literární teoretik Willie van Peer, který ve studii *Quantitative Studies of Literature. A Critique and an Outlook* (1989) konstatuje: „I take it that quantitative studies of literature do nothing less than that: trying to describe as accurately as possible the various linguistic elements and configurations one encounters in literary works of art.“ (Peer 1989: 302) Na druhé straně se právě kvantitativní i korpusová analýza ve světě čím dál častěji využívá v kontextu literárněvědného bádání. Jak poznamenává David L. Hoover, možnosti tohoto přístupu se stále rozšiřují: „Quantitative approaches are most naturally associated with questions of authorship and style, but they can also be used to investigate larger interpretive issues like plot, theme, genre, period, tone, and modality.“ (Hoover 2008: 517) V čem spočívá tato názorová dichotomie, kdy na jedné straně je odmítán kvantitativně-korpusový přístup v literární vědě a na straně druhé, zejména v zahraniční literatuře, vidíme stále rostoucí úsilí o vzájemné propojení. Van Peerův názor významně koresponduje s tím, co uvedl již Klivar, že totiž literární systém je složitým komplexem různých strukturních a významových úrovní, které nelze redukovat na společný základ, kterým je jazyková rovina.

Tuto skutečnost nejen že nelze popřít, ale naopak je třeba ji zdůraznit. Literární umělecké dílo je komplexním jevem, který kromě řady imanentních aspektů zahrnuje i širší mimoliterární oblasti. Především se jedná o otázky spojené s dobovým a kulturním kontextem díla, recepční aspekty apod. Literární dílo nelze vyčerpát z hlediska jediné metodologie; jeho povaha komplexního kulturního jevu, jenž je navíc nositelem složitých a mnohdy variabilních významových kvalit, dovoluje řadu dokonce protichůdných interpretací či metodologických postupů. Od žádné literárněvědné metody nelze proto právem očekávat konečné a jediné relevantní řešení celé řady složitých otázek, jež s problematikou literárního díla souvisí. Smysl plurality (nejen) literárněvědných diskurzů spočívá v tom, že umožňuje nezbytnou kritickou konfrontaci mezi jednotlivými přístupy, která stojí rovněž na pozadí jejich vlastního odborného vývoje. Otázky kladené napříč metodologiemi nejenže problematizují samotný předmět zkoumání, ale také rozvíjí každý z jednotlivých teoretických přístupů. Nejinak je tomu i v tomto případě. Očekávat od kvantitativní či korpusové analýzy uměleckého sloves-

/4/ Český národní korpus byl založen v roce 1994.

ného díla, že odpoví na otázky spojené s komplexní povahou literárního díla, je základem fatálního nedorozumění. Stejně tak je mylné očekávat od kvantitativní analýzy výsledky, ke kterým není disponována, nebo se domnívat, že kvantitativní či korpusová metoda a její pracovní postupy mohou nahradit vlastní literárněvědnou analýzu a interpretaci. Ostatně ani v lingvistice, která s korpusy pracuje, neplatí, že by samotné korpusové výsledky dostačovaly k relevantnímu vysvětlení příslušných jevů. Je třeba zdůraznit, že korpus, stejně jako metoda a postupy statistické a kvantitativní analýzy, slouží především v úloze nástrojů, jejichž potenciál spočívá v tom, že v kontextu humanitních věd umožňují objektivizovat vybranou problematiku. Je rovněž mylné se domnívat, že formální analýza jazykové vrstvy uměleckého díla se týká výhradně a pouze této vrstvy.

Ačkoli platí, že explicitně se kvantitativně-korpusová analýza vztahuje k jazykové složce uměleckého díla, která ostatně jako jediná je empiricky dostupná našemu zkoumání (viz Schmid 2004: 47), nemusí referovat výhradně pouze k ní. Samotné kvantitativní i korpusové techniky náleží obecně pod statistiku jazyka a záleží v prvé řadě na výzkumné strategii, kterou se budou řídit. Zde je na místě znovu upozornit na Doleželův výrok, kterým rozlišil dvojí způsob zacházení se statistickými metodami. Van Peerova kritika pokládá kvantifikaci za pouhé počítání množství výskytů určitého prvku, nicméně kvantita je rovněž odrazem strukturního systému a jeho organizace (srov. Čermák 2010: 244). Velká výhoda kvantitativní a korpusové analýzy spočívá v tom, že dovoluje objektivně registrovat zákonitosti a charakter konkrétního strukturního systému, což se pochopitelně týká především jazykové vrstvy. Vyšší strukturní oblasti, kam náleží například narativní mody a strategie, aspekty z oblasti teorie fikčních světů apod., sice nejsou bezprostředně kvantifikovatelné, nicméně s jazykovou vrstvou vytváří funkčně provázaný celek. Navíc kvantitu nelze nikdy myslet bez kvalitativních kritérií, jimiž v případě literárního díla jsou typy jazykových prvků a jejich sémantika. Van Peer svoji kritiku zakládá na argumentu, že různé roviny slovesného díla nelze stejně úspěšně kvantifikovat. Čím výše postoupíme na úrovni strukturní abstrakce, tím menší bude úspěšnost kvantifikace jejích komponent. To je ovšem zcela nepopíratelné. Na druhé straně kvantitativně-korpusový přístup neusiluje o absolutní kvantifikaci všech složek uměleckého díla, ale pouze těch, které jsou úspěšně kvantifikovatelné. Slovem úspěšně se rozumí, že počítat lze pouze takové prvky, které lze relativně přesně a objektivně identifikovat a vymezit s ohledem na formální procedury taxonomizace, což je i případ slov (lexémů). To však automaticky nepředpokládá, že výzkum založený na kvantifikaci lexémů uměleckého díla se musí nutně omezit výhradně na danou oblast. Struktura uměleckého díla je komplexní organizovanou jednotou, ve které jsou vzájemně propojeny nejen prvky na téže strukturní úrovni, ale i jednotlivé strukturní vrstvy. Nelze například myslet sémantiku fikčního světa a jeho entit bez funkční komplementarity

jazykových, narativních, kompozičních, tematických, motivických a celé řady dalších aspektů, které sice lze teoreticky izolovat a zkoumat relativně samostatně, avšak ve výsledku vždy vytváří homogenní celek, organickou strukturu uměleckého díla.

Řekli jsme, že je rovněž otázkou pracovní strategie, za jakým účelem budou kvantitativně-korpusové metody využívány. Formální analýza umožňuje zkoumat nejen prostý počet výskytu izolované jednotky, např. slova, ale současné korpusové nástroje dovolují také formalizovat kontexty (kolokace), jež jsou důležité pro specifikaci významu. Každá jednotka se nejen vřazuje do svého kontextového okolí (syntagmatický aspekt), ale disponuje i paradigmatickou, což například představuje její příslušnost pod určitou sémantickou třídou. Charakteristiku sémantických tříd lze realizovat s ohledem na lingvistickou, ale i literárněvědnou klasifikaci. Ve výsledku tedy dochází k propojení několika formálně analyzovatelných vrstev a oblastí – kvalitativní aspekt je určen typem jednotky (nemusí se jednat pouze o slovo, ale například o kolokaci), jejím kontextem a paradigmatem, tzn. příslušností do nadřazených sémantických tříd a skupin, kvantitativní aspekt specifikuje příznakovost užití dané jednotky v systému. Vždy ovšem platí, že relevantních výsledků se lze dobat teprve srovnáváním dílčích i komplexnějších jevů a teprve po náležité interpretaci statistických a korpusových údajů, která musí probíhat v rámci konkrétní lingvistické nebo literárněvědné metodologie. Zejména vyšší jednotky, tzn. sémantické třídy a jejich vztahy odkazují k dalším strukturním činitelům v rámci výstavbové strategie uměleckého textu, především k oblasti narativních technik a strategií, kompozice, tematiky apod. Samozřejmě, že zde se již nelze spoléhat výhradně na samotné výsledky kvantitativně-korpusové analýzy, neboť se ocitáme v komplikovaně kvantifikovatelných oblastech. Ke slovu tak přistupuje kritická extrapolace, jejímž korektivem je relevantní metodologie, což v případě literární vědy představují zejména ty postupy, které jsou epistemologicky koherentní s kvantitativně-korpusovou analýzou, tzn. strukturálně orientovaná interpretace.

Výstupní informace uvedené formální analýzy spočívají především v absolutních hodnotách, tedy nemohou substituuovat individuální analyticko-interpretační výkon. Absolutní hodnoty jsou však přesně tím, co je třeba od kvantitativně-korpusové analýzy očekávat, v čem spočívá její základní význam a přínos. Tyto informace jsou validní právě svojí objektivitou a „plošnou“ referencí k fundamentální rovině strukturního systému uměleckého textu.

V současné době je vyvíjeno nemalé úsilí taxonomizovat i vyšší segmenty uměleckého textu, např. narativní promluvy, přímé řeči postav nebo dialogy v dramatech, jak konstatuje David L. Hoover:

Words themselves, as the smallest clearly meaningful units, are the most frequently counted items, and syntactic categories (noun, verb, infinitive, su-

perlative) are also often of interest, as are word *n*-grams (sequences) and collocations (words that occur near each other). Thematic or semantic categories (angry words, words related to time), while more difficult to count, have the advantage of being clearly relevant to interpretation, and automated semantic analysis may reduce the effort involved. Phrases, clauses, syntactic patterns, and sentences have often been counted, as have sequences or subcategories of them (prepositional phrases, subordinate clauses, passive sentences). Many of the items above are also used as measure of the lengths of their items: word length in characters, sentence or clause length in letter or words, text length in words, sentences, paragraphs, and so forth. Nonlinguistic textual division ranging from small units like lines and couplets to large structural units like paragraphs, stanzas, scene, acts, and characters (including subcategories like first-person and third-person narrators, and characters divided by age, ethnicity, nationality, class, and gender), and plot elements (marriages, deaths, journeys, subplots). [...] Counting may be limited to the dialogue or narration of a text, to one or more speakers or narrators, or to specific passages. (Hoover 2008: 519–520 a 520).

V tomto ohledu je pozoruhodná také studie Larryho L. Stewarta „Charles Brockden Brown: Quantitative Analysis and Literary Interpretation“ (2003), ve které autor kvantifikuje rozdíly mezi jednotlivými narativními promluvy jedné Brockdenovy novely. Stewartova zjištění učiněná na základě kvantifikací vybraných textových (jazykových) markerů⁵ potvrzují shody mezi zdánlivě rozdílnými vypravěči. Kromě toho je Stewartova studie dobrým příkladem toho, jak lze kvantitativních analytických metod smysluplně a funkčně využít v literárněvědné práci.

Kvantitativní zkoumání soustředěné na umělecké texty má čím dál větší tendenci se prosazovat v zahraniční odborné literatuře, jak o tom svědčí mnohé články v časopise *Literary and Linguistic Computing*⁶ nebo ve sbornících, např. *A Companion to Digital Humanities*,⁷ *Digital Literary Studies – Corpus Approaches to Poetry, Prose, and Drama* (2014), kde je mimo jiné k dispozici studie zabývající se stylovou charakteristikou řečových promluv v dramatu. Postupně se této problematice začaly věnovat i samostatné monografie, např. rozsáhlá práce Jana Christopera Meistera *Computing Action (Narratologia)*, monografie Roberta Franzosiho *Quantitative Narrative Analysis* (2010)⁸ nebo připravovaná monografie *Quantitative Analysis of Poetic Texts* Ioana Iovitze

/5/ Viz Stewart 2003, s. 132–133

/6/ Teprve nedávno přejmenováno na *Digital Scholarship in the Humanities* (<http://dsh.oxfordjournals.org/>).

/7/ Mnohé z nich jsou dostupné na internetu.

/8/ Viz heslo *Computational Narratology* v *Living Handbook of Narratology* (<http://www.lhn.uni-hamburg.de/article/computational-narratology>).

Popescu, Mihaiely Lupey, Diony Tatar a Gabriela Altmanna. Výčet prací na dané téma zdaleka nekončí.⁹

Kvantitativně-korpusová analýza tím, že pořádá jevy na základě jejich kvantitativní charakteristiky, umožňuje objektivizovat zkoumání strukturních vlastností konkrétního systému, ve kterém jsou tyto jevy uspořádány a organizovány. V případě analýzy literárního díla to jistě neznamená, že se tím vyčerpávají veškeré jeho vlastnosti. Podstatné ovšem je, že pomocí formální taxonomizace a algoritmizace lze relativně objektivně a exaktně zkoumat fundamentální rovinu uměleckého díla, která s vyššími rovinami vytváří celistvost strukturně organizované jednoty.

/9/ Na kvantitativních a korpusových metodách je v českém prostředí založen projekt Korpus českého verše, který je od roku 2012 realizován versologickým týmem na Ústavu pro českou literaturu AV ČR, v. v. i. (<http://www.versologie.cz/>).

LITERATURA

- BEVAN, Dean E.
1971 *A Concordance to the Plays and Prefaces of Bernard Shaw* (Detroit – Michigan: Gale Research Company)
- ČERMÁK, František
2010 *Lexikon a sémantika* (Praha: Lidové noviny)
- Digital Literary Studies*
- Digital Scholarship in the Humanities*
- DOLEŽEL, Lubomír
1965 „Pražská škola a statistická teorie básnického jazyka“; *Česká literatura* 13, č. 2, s. 110–113
- HANCOCK, Leslie
1967 *Word Index to James Joyce's Portrait of the Artist* (London – Amsterdam: Southern Illinois University Press)
- HOOVER, David L.
2008 „Quantitative Analysis and Literary Studies: History, Goals, and Theoretical Foundation“; in *A Companion to Digital Literary Studies* (Oxford: Blackwell), s. 517–533
- KLIVAR, Miroslav
1984 „Význam kvantitativní lingvistiky pro literární teorii.“; in *Literárněvědný sborník Památníku národního písemnictví* (Praha: Památník národního písemnictví), s. 24–52
- MEDICK, Doris B.
2006 *Cultural Turns. Neuorientierungen in den Kulturwissenschaften* (Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag)
- MEISTER, Jan Ch.
2003 *Computing Action – Narratologia* (Berlin – New York: Walter de Gruyter)
- MORETTI, Franco
2014 *Grafy, mapy, stromy. Abstraktní modely literární historie* (Praha: Karolinum)
- PEER, van Willie
1989 „Quantitative Studies of Literature. A Critique and an Outlook“; *Computers and the Humanities* 23, s. 301–307
- SCHMID, Wolf
2004 *Narativní transformace*; přel. Petr Málek (Brno: Ústav pro českou literaturu AV ČR)
- STEWART, Larry L.
2003 „Charles Brockden Brown: Quantitative Analysis and Literary Interpretation“; *Literary and Linguistic Computing* 18, č. 2, s. 129–138

ŠTINDLOVÁ, Jitka

1969 *Konkordanční a frekvenční index k Slezským písním Petra Bezruče* (Praha – Ostrava: Socialistická akademie v Ostravě)

TĚŠITELOVÁ, Marie

1974 *Otázky lexikální statistiky* (Praha: Academia)