

CATMA – Digitálny nástroj na anotáciu, analýzu a vizualizáciu literárnych textov

Miroslav Zumrík

ZUMRÍK, M.: CATMA – a Digital Tool Used to Annotate, Analyze and Visualize Literary Texts

SLOVENSKÁ LITERATÚRA 67, 2020, No. 6, p. 614 – 629

DOI: <https://doi.org/10.31577/slovlit.2020.67.6.7>

Key words: digital literary studies, digital humanities, digital annotation of literary texts, narrative analysis, method of literary scientific analysis, philosophy of technology

The paper reflects on a freely accessible program to annotate, analyze and visualize literary texts called CATMA, which is developed by the narratological team led by Jan Christoph Meister in Hamburg, which follows the tradition of the German research into the area of digital or quantitative narratology. The design of the annotation program develops the conceptual apparatus of classical narratology and these (e.g. Genettian types of anisochrony) are implemented in digitally supported analysis and interpretation of semantics in the genre of short story. Furthermore, annotation and visualization do not replace quantitative analysis, what they do is prepare for a literary scientist a kind of a cognitive map which may help them notice potentially interesting (ir)regularities of texture of the work (grammatical, morphological, semantic features). The CATMA tool thus opens possibilities not only for quantitative study of authors' corpora but also for analysis of motifs which are in the sense of narrative semantics of fictitious worlds as formulated by Lubomír Doležel intentionally expressed by means of distribution of linguistic signs.

Kľúčové slová: digitálne literárne štúdiá, digitálne humanitné vedy, digitálna anotácia literárnych textov, naratívna analýza, metóda literárnovednej analýzy, filozofia technológie

Prístup k analýze jednotlivých literárnych diel, literárnych korpusov či literárneho diskurzu vo všeobecnosti označovaný ako digitálne literárne štúdiá (*Digital Literary Studies*) tvorí v súčasnom literárnovednom diskurze čoraz početnejšiu a rôznorodejšiu množinu výskumných výstupov, projektov, ale aj výskumných infraštruktúr, pričom dôraz sa stále viac kladie na živý interdisciplinárny dialóg a výskum multimodálnych korpusov.¹ Rôznorodosť a pluralitný charakter korpusových a digitálnych zdrojov a prístupov sa zrkadlí napríklad už v štruktúre encyklopedického diela *A Companion to Digital Literary Studies*:² tri časti príručky sa postupne venujú tradíciám (*traditions*) namiesto tradície, takzvaným textuálnym (*textualities*) namiesto textuality a metodológiám (*methodologies*) namiesto metodológie. V tomto množstve informácií a podnetov je ľahké stratiť orientáciu či prehliadnúť potenciálne inšpiratívny prístup k výskumu literárneho diskurzu a uvažovania o ňom. V príspevku sa preto zameriam na konkrétnu podobu tohto rôznorodého výskumu, počítačový program CATMA (*Computer Aided Text Markup and Analysis*) určený, ako napovedá už jeho názov, na počítačovo podporovanú anotáciu (značkovanie), vizualizáciu a analýzu literárnych textov. Mojim cieľom bude reflektovať tento počítačový nástroj z metodologického hľadiska, poukázať na jeho inšpiračný potenciál a – keďže ide o voľne dostupný online program – jeho využiteľnosť pri počítačovo podporovanej analýze aj v rámci slovenskej literárnej vedy, teda pri analýze slovenských literárnych korpusov, či už pôvodných, alebo prekladových.

Digitálne literárne štúdiá, samy osebe rôznorodé, tvoria súčasť ešte komplexnejšej oblasti označovanej ako digitálne humanitné vedy (*Digital Humanities*). Priam exponenciálny nárast dát, metadát a v nich ukrytých štruktúr, informácií a potenciálnych poznatkov v tejto oblasti potom dopĺňa jednak metodologická i filozofická sebareflexia digitálnych humanitných vied, jednak badať úsilie o komunikáciu s „tradičnejšími“ prístupmi k filologickej a literárnovednej analýze a interpretácii. „Tradičnejší“ prístup možno s istou dávkou zjednodušenia extrapolovať ako ten kvalitatívny, zatiaľ čo ten „modernejší“ sa spája s kvantitatívnym prístupom.

Úvodzovkami pri týchto prívlastkoch chcem naznačiť, že oba prístupy sa podľa môjho názoru v skutočnosti dopĺňajú, tvoria vlastne rozdielne akcenty záujmu o literárne dielo. Kým prvý, „tradičnejší“ prístup zdôrazňuje potrebu analyzovať dielo či diela v ich konkrétnosti a jedinečnej literárnej kvalite, ten druhý, „digitálny“, poukazuje na potrebu kvantifikácie kvalitatívnych tvrdení o literárnom diskurze. Takéto tvrdenia sa môžu týkať údajov napríklad pri skúmaní literárnej produkcie v nejakom období alebo u nejakého autora s cieľom danú literárnu produkciu štylisticky charakterizovať či „formálne historizovať“.³ Kvantitatívny prístup sa v rámci textovej analýzy konkrétneho diela môže využiť napríklad pri skúmaní autorských poetík podľa výskytu a distribúcie použitých jazykových

1 ČERMÁK, František: *Korpus a korpusová lingvistika*. Praha: Karolinum, 2017, s. 68 a 133.

2 SIEMENS, Ray – SCHREIBMAN, Susan (eds.): *A Companion to Digital Literary Studies*. Oxford: Blackwell, 2008. Dostupné na: <http://www.digitalhumanities.org/companionDLS/>

3 WERNER, Lukas: Zeit im Text – Möglichkeiten der formalen Historisierung. Zur Relation von Zeit und Ich-Erzähler in Grimmselhausens *Simplicissimus* und Schnabels Insel Felsenburg. In: WEIXLER, Antonius – WERNER, Lukas (Hrsg.): *Zeiten erzählen. Ansätze – Aspekte – Analysen. Narratologia 48*. Berlin – Boston: De Gruyter, 2015, s. 84.

616 výrazov. V príspevku chcem preto poukázať na to, že kvantitatívne a kvalitatívne akcenty sa navzájom nevyklučujú, skôr naopak, nástroje kvantitatívnej analýzy, akým je aj program CATMA, predstavujú pomôcku, ktorá uľahčuje literárnovednú analýzu a interpretáciu.⁴ V oblasti korpusovej lingvistiky, a teda počítačovo podporovanej analýze jazyka sa podobný metodologický posun vyjadruje napríklad ako pohyb „od introspekcie ku korpusovo podloženým, ‚informovaným‘ tvrdeniam“ (*from introspective to corpus-informed judgments*)⁵ či ako explorácia „ciest k dobre podloženým a spoľahlivým tvrdeniam o reči“ (*Wege zu wohl dokumentierten und verlässlichen Aussagen über Sprache*).⁶ Analogicky možno uvažovať o pohybe ku kvantifikovateľným, korpusovo podloženým tvrdeniam o konkrétnych literárnych textoch, autorských či žánrových a historických korpusoch. Na základe kvantitatívnych zistení, získaných aj vďaka kvantitatívnym nástrojom, by bolo možné presnejšie formulovať tvrdenia o výrazových a štylistických príznakoch týchto textov, autorov či celých období.

V rámci oblasti digitálnych humanitných vied (DH) má naznačená teoretická sebareflexia a komunikácia s ďalšími prístupmi k literárnemu diskurzu rôznorodú podobu. Čo sa týka metodologickej i terminologickej reflexie základných postupov a konceptov, registrovať možno úsilie o objasnenie týchto východísk napríklad v podobe pôvodných prehľadových prác o rôznych definíciách a postupoch DH⁷ alebo tiež v podobe kultúrneho transferu, sprístupňovania, konkrétne prekladu zahraničných prác,⁸ ako aj zvyšovania povedomia o nových prístupoch, tak ako sa o to pokúša či mohla by sa pokúšať napríklad jazykoveda vo vzťahu k verejnosti.⁹ Úsilie o komunikáciu zástupcov komunity DH s ďalšími vedeckými komunitami a verejnosťou má zasa podobu úvah a iniciatív zameraných na možnosti počítačovo *podporovanej*, nie však počítačom *suprovanej* literárnej vedy¹⁰ a tiež zameraných na budovanie digitálnych výskumných infraštruktúr a platforiem.¹¹

Rezervovanosť časti literárnovednej komunity vo vzťahu k prístupom digitálnej literárnej vedy možno pritom s istou dávkou zjednodušenia vyjadriť ako obavu, že sa kvalitatívne hodnotenie literárnych motívov a bytostnej sémantickej nejednoznačnosti literárnych textov nahrádza akousi mechanickou transformáciou a rozkladom týchto textov na čísla, a to v prostredí počítačových programov

4 SCHÜCH, Lena: Computerphilologische Analyse. „Tagging in a huge meadow of time“ – Analysen der Zeit in Erzähltexten mit Hilfe des Programms CATMA. In: Weixler – Werner, *Zeiten erzählen...*, c. d., s. 28.

5 DESAGULIER, Guillaume: *Corpus Linguistics and Statistics with R. Introduction to Quantitative Methods in Linguistics*. Cham : Springer International Publishing, 2018, s. 1.

6 EICHINGER, Ludwig M.: Linguisten brauchen Korpora und Korpora Linguisten: Wege zu wohl dokumentierten und verlässlichen Aussagen über Sprache. In: KALLMEYER, Werner – ZIFONUN, Gisela (eds.): *Sprachkorpora: Datenmengen und Erkenntnisfortschritt*. Berlin – New York : Walter de Gruyter, 2007, s. 1.

7 GOGORA, Andrej: Digital humanities a filozofia – problém digitálneho výskumu. In: *World Literature Studies*, roč. 8, 2016, č. 3, s. 104 – 114.

8 BURDICKOVÁ, Anne – DRUCKEROVÁ, Johanna – LUNENFELD, Peter – PRESNER, Todd – SCHNAPP, Jeffrey: *Digital Humanities*. Prel. David Vichnar. Praha : Academia, 2019.

9 DOLNÍK, Juraj: *Jazyk v sociálnej kultúre*. Bratislava : Veda, 2017.

10 DEBNÁR, Marek: O esejistickom štýle z kvantitatívneho hľadiska alebo o možnostiach využitia korpusov v širšom humanitnom výskume. In: ŠIMKOVÁ, Mária – LEVICKÁ, Jana – DEBNÁR, Marek: *Dynamické javy v súčasnej slovenčine a jej výskume*. Bratislava : Veda, 2018, s. 165 – 192; ZMĚLÍK, Richard: *Kvantitatívne-korpusová analýza a literárni veda. Model a realizace autorského korpusu a slovníku Jana Čepa v kontextu zahraniční a české autorské lexikografie*. Olomouc : Univerzita Palackého, 2015.

11 Na Slovensku ide o Centrum digitálnych humanitných vied na Univerzite Konštantína Filozofa v Nitre.

ako interpretačne necitlivých „čiernych skriniek“. Takáto obava je pochopiteľná, ak by motiváciou digitálnych literárnych štúdií skutočne bolo presvedčenie, že špecifickú kvalitu a významovú mnohoznačnosť literárneho diskurzu možno bezo zvyšku rozložiť na jej výrazové jednotky a skúmanie literárneho diskurzu možno nahradiť kvantitatívnym skúmaním jazykových znakov ako materiálnych nosičov literárnosti. V takom prípade by teda neplatila holistická téza, podľa ktorej celok je „viac než súčet častí“. ¹² Podobné presvedčenie sa však u zástupcov digitálne podporovaného prístupu k literatúre zväčša neobjavuje. Zaujímavé je však v tejto súvislosti zistenie, že rezervovanosť voči inštrumentálne, a teda technologicky podporovanej analýze dát prinajmenšom v istom čase nepretrvávala iba v oblasti humanitných, tradične hermeneuticky orientovaných vied, ale aj v takej exaktnej empirickej vede, akou je fyzika. Tak to aspoň konštatuje Irving Wladawsky-Berger v rozhovore s názvom *Why Do We Need Data Science When We've Had Statistics for Centuries?* (2014). ¹³ Podľa filozofov technológie ¹⁴ skôr platí, že technológia a technologické artefakty sú v nejakej podobe „všadeprítomné“ (*ubiquitous*), ¹⁵ a to či už v podobe fyzických predmetov a nástrojov (knihy, písacie potreby, okuliare), alebo digitálnych predmetov (počítačové programy, databázy, virtuálne repositáre, digitalizované zbierky).

Naznačenú širokú a rôznorodú oblasť metodológie, respektíve metodológií DH a jej podoblasti, digitálnych literárnych štúdií (DLS), sa pokúša rozčleniť a usporiadať Jan Horstmann (2020) v podobe grafu, pričom jeho os x tvorí kontinuálna, tzv. metodologická škála (*methodologische Skala*). Na jeden jej koniec (bližšie k stredu súradnicovej sústavy) umiestňuje J. Horstmann pól dištancie, na druhý koniec pól blízkosti. Os y tvorí rovnako kontinuálna škála veľkosti analyzovaných korpusov ako objektov – je to tzv. objektová škála (*Objekt-Skala*). Táto os siaha jednoducho od „malých“ po „veľké“ korpusy. ¹⁶ Metodologickú rôznorodosť prístupov v oblasti DLS znázorňuje klesajúca krivka, v rámci ktorej J. Horstmann rozlišuje tri skupiny výskumných projektov. Dištančné prístupy uskutočňované

12 SVÍTEK, Miroslav: *Víc než součet částí: Systémový pohled na proces lidského poznání*. Praha: Academia, 2013.

13 I. Wladawsky-Berger v rozhovore uvádza: „I still remember a meeting in 1965 with a very eminent physicist from whom I was taking a graduate course. He asked me what I planned to do research on for my degree, and I told him that I was already working with Prof. Roothaan on atomic and molecular calculations. He just said that good theoretical physics should require no more than pencil and paper, rather than these elaborate new computers. In his mind, this wasn't real physics.“ Dostupné na: <https://blog.irvingwb.com/blog/2014/04/why-do-we-need-data-science-when-weve-had-statistics-for-centuries>

14 Možno spomenúť napr. Marka Coeckelbergha, ktorý sa živo zaujíma o filozoficko-etické otázky spojené s prítomnosťou moderných technológií (biotechnológií, robotiky, génového inžinierstva) v našom živote a takisto z hľadiska filozofie technológie vykladá koncept jazykových hier L. Wittgensteina, ktorého označuje výslovné za „filozofa technológie“ (COECKELBERGH, Mark: *Using Words and Things. Language and Philosophy of Technology*. New York/Oxon: Routledge, 2017; COECKELBERGH, Mark – FUNK, Michael: Wittgenstein as a Philosopher of Technology: Tool Use, Forms of Life, Technique, and a Transcendental Argument. In: *Human Studies*, roč. 41, 2018, č. 2, s. 165 – 191).

15 INTRONA, Lucas: Phenomenological Approaches to Ethics and Information Technology. In: ZALTA, Edward N. (ed.): *The Stanford Encyclopedia of Philosophy*. 2017. Dostupné na: <<https://plato.stanford.edu/archives/fall2017/entries/ethics-it-phenomenology/>>

16 HORSTMANN, Jan: Textvisualisierung: Epistemik des Bildlichen im Digitalen. In: HUBER, Martin – KRÄMER, Sybille – PIAS, Claus (Hrsg.): *Wovon sprechen wir, wenn wir von Digitalisierung sprechen? Gehalte und Revisionen zentraler Begriffe des Digitalen. Digitalität in den Geisteswissenschaften*. Bayreuth: DFG-geförderte Symposienreihe Universität Bayreuth, 2020, s. 159.

618 na tých najväčších korpusoch označuje ako „megadištančné čítanie“ (*Mega Distant Reading*) a zaraďuje sem skúmanie štruktúr, ako sú sekvencie hlások a slov, takzvané n-gramy či využitie programov na vektorovú reprezentáciu sémantiky slov a viet.¹⁷ Do prostrednej oblasti krivky umiestňuje J. Horstmann paradigmu dištančného čítania, pôvodne od Franca Morettiho, do ktorej zaraďuje štylometriu ako disciplínu aplikovanej jazykovedy,¹⁸ úlohu rozpoznávania pomenovaných entít (*Named Entity Recognition*) ako jednu z úloh umelej inteligencie a tiež analýzu sentimentu či analýzu sociálnych sietí. Najviac napravo na osi x Horstmann zobrazuje metódy *close reading*, a teda postupy zamerané na rozsahovo malé korpusy, ktoré sa skúmajú z najväčšej „blízkosti“: sem zaraďuje kolaboratívnu anotáciu týchto malých korpusov až manuálnu anotáciu jednotlivých textov. Klesajúca krivka má pritom znázorňovať postupne sa zmenšujúci rozsah analyzovaných textov, predmetov, ktoré sa literárny vedec s využitím počítačových programov pokúša vizualizovať, a tak uchopiť. Vychádzajúc z Horstmannovej krivky patrí program CATMA práve do poslednej opisanej skupiny postupov a predstavuje teda počítačovo podporované pokračovanie v tradícii *close reading*.

Funkcionalita a dizajn programu CATMA

Počítačový program na anotáciu, vizualizáciu a analýzu textov CATMA¹⁹ vyvíjajú od roku 2008 nemeckí naratológovia z Univerzity v Hamburgu ako reimplemenciu pôvodného programu TACT (*Textual Analysis Computing Tools*). Ústrednou osobnosťou hamburského naratologického tímu je Jan Christoph Meister, ktorý sa počítačovým spracovaním naratívneho textu (tzv. počítačovou naratológiou, *computational narratology*) zaoberá dlhodobo. V tejto oblasti možno spomenúť jeho rozsiahlu monografiu,²⁰ v ktorej sa pokúsil odhaliť návratnú, invariantnú štruktúru fikčných udalostí v cykle noviel od J. W. Goetheho, alebo príspevok, v ktorom sa zamýšľa nad „komputabilitou“, „vypočítateľnosťou“ (*Berechenbarkeit*) naratívnych textov, a teda vlastne nad otázkou, či sa môže počítač „naučiť“ generovať zmysluplné naratívne texty.²¹

Program CATMA jeho tvorcovia priebežne obnovujú, v súčasnosti je k dispozícii jeho šiesta verzia, pričom autori projektu sa, ako naznačím ďalej, jednak obracajú na používateľov zo širokej odbornej i neodbornej verejnosti, jednak sa pokúšajú o integráciu doterajších funkcionalít programu (najmä o prepájanie

17 Napr. VAN CRANENBURGH, Andreas – VAN DALEN-OSKAM, Karina – VAN ZUNDERT, Joris: Vector space explorations of literary language. In: *Lang Resources & Evaluation*, roč. 53, 2019, s. 625 – 650; DESAGULIER, Guillaume: Can word vectors help corpus linguists? In: *Studia Neophilologica*, roč. 91, 2019, č. 2, s. 1 – 21; REPLINGER, Michael – BEINBORN, Lisa – ZUIDEM, Willem: Vector-space models of words and sentences. In: *Nieuw Archief voor Wiskunde (NAW journal)*, roč. 5/19, 2018, č. 3, s. 167 – 174.

18 Štylometria sa tradične využíva napríklad pri určovaní autorstva (*authorship attribution*), a to prostredníctvom skúmania charakteristického radenia výrazových prostriedkov (napr. funkčných slov, akým je sponové sloveso byť), ktoré holisticky vytvárajú ťažko falzifikovateľný, tzv. autorský signál. Preto sa v praktickej oblasti využíva aj na odhalovanie plagiatorstva.

19 Dostupný na [www.catma.de](https://catma.de). Softvérové a technologické charakteristiky dostupné na: <https://catma.de/documentation/technology-and-versions/>

20 MEISTER, Jan Christoph: *Computing Action. A Narratological Approach*. Narratologia 2. Boston – Berlin: De Gruyter, 2003.

21 MEISTER, Jan Christoph: Kann man das Erzählen berechenbar machen? In: MÜLLER, Corinna – SCHEIDGEN, Irina (Hrsg.): *Erzählen, Archivieren, Beschreiben*. Marburg: Schriftenreihe der Gesellschaft für Medienwissenschaft 5, 2007, s. 19 – 39.

anotácie a analýzy). Zároveň badať úsilie autorov o rozširovanie programových funkcionalít s cieľom vytvoriť infraštruktúrne podmienky pre „digitálnu heuristiku naratívu“ (*Digital Heuristics of Narrative*) a počítačovej naratológie v rámci projektov CATMA i heureCLÉA.²²

Možno povedať, že autori projektu takisto venujú pozornosť teoretickej sebareflexii. Tú tvoria na jednej strane príspevky venované teoretickým až filozofickým premisám a implikáciám počítačovej aplikácie,²³ na druhej strane sa usilujú dokumentovať históriu projektu CATMA i predchádzajúcich projektov.²⁴ Zaujímavá je metodologická sebareflexia spolupracovníkov projektu, ktorí vnímajú vznikajúcu digitálnu platformu literárnych analýz ako metodologické novum so značnými možnosťami ďalšieho rozvoja.²⁵ Orientácia na používateľa vyplýva z priebežného publikovania tutoriálov i anotačných manuálov,²⁶ a tiež z dôrazu na projektovú dokumentáciu,²⁷ kde sa uvádzajú dosiahnuté výsledky, identifikované problémy a možné výskumné otázky a úlohy v budúcnosti.

Vychádzajúc z novšieho spracovania anotačných zásad²⁸ možno proces užívateľského spracovania literárnych textov v programe opísať približne nasledujúcim spôsobom. Užívateľ (alebo skupina používateľov) pracuje s uchovanou a prostredníctvom používateľského účtu sprístupnenou elektronickou podobou literárneho textu či celého korpusu textov vo formáte XML (*extensible mark-up language*). Tento formát umožňuje kódovať jednotlivé textové štruktúry, ako sú nadpisy, repliky, odseky a podobne. Do týchto štruktúr zasa anotátor pomocou grafického rozhrania programu CATMA a nástrojov vnáša preddefinované alebo vlastné definované značky (*tagy*). Značkovat môže jednotlivé výskyty explicitných časových markerov v texte, ako sú napríklad dátumy, alebo relatívnych časových údajov (výrazy ako *predtým*, *najprv*, *a potom*). Grafické rozhranie programu pritom umožňuje vyznačiť nielen lineárne usporiadané textové štruktúry (nadpisy, odseky, repliky), ale aj prerušované či distribuované štruktúry (napríklad vložené uvádzacie vety replík). Anotátor tiež môže značkovat naratívne javy, ktoré už

22 <http://www.heureclea.de>

23 GIUS, Evelyn – JACKE, Janina: The Hermeneutic Profit of Annotation. On Preventing and Fostering Disagreement in Literary Analysis. In: *International Journal of Humanities and Arts Computing*, roč. 11, 2017, č. 2, s. 233 – 254; MEISTER, Jan Christoph: From TACT to CATMA or A mindful approach to text annotation and analysis. In: NYHAN, Julianne – ROCKWELL, Geoffrey – SINCLAIR, Stéfan (eds.): *On Making in the Digital Humanities: Essays on the Scholarship of Digital Humanities*. Development in Honour of John Bradley. 2020. Dostupné na: https://jcmeister.de/downloads/texts/Meister_2020-TACT-to-CATMA.pdf; Meister, Kann man das Erzählen berechenbar machen?, c. d.

24 LANCASHIRE, Ian – BRADLEY, John – McCARTY, Willard – STAIRS, Michael – WOOLDRIDGE, T. R.: *Using TACT with Electronic Texts: A Guide to Text-Analysis Computing Tools, Version 2.1 for MS-DOS and PC DOS*. New York: Modern Language Association of America, 1996; MOULIN, Claudine: Am Rande der Blätter. Gebrauchsspuren, Glossen und Annotationen in Handschriften und Büchern aus kulturhistorischer Perspektive. In: *Quarto. Zeitschrift des Schweizerischen Literaturarchivs*, 30/31, 2010, s. 19 – 26.

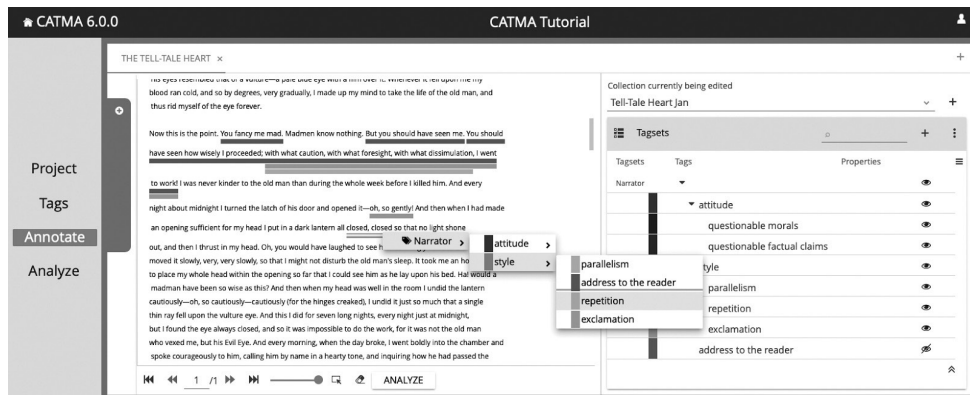
25 „Die Interaktionsmöglichkeiten digitaler Textvisualisierung erschaffen einen solchen prozessualen kognitiven Bewegungsraum; ein Bereich, in dem es noch erhebliches Entwicklungspotential gibt“ (Horstmann, Textvisualisierung: Epistemik des Bildlichen im Digitalen, c. d., s. 170).

26 HORSTMANN, Jan: Manuelle Annotation mit CATMA. In: *forTEXT. Literatur digital erforschen*. Dostupné na: <https://fortext.net/routinen/lerneinheiten/manuelle-annotation-mit-catma>. 2019.

27 BÖGEL, Thomas – GERTZ, Michael – GIUS, Evelyn – JACKE, Janina – MEISTER, Jan Christoph – PETRIS, Marco – STRÖTGEN, Jannik: Collaborative Text Annotation Meets Machine Learning: heureCLÉA, a Digital Heuristic of Narrative. In: *DHCommons Journal 1*, July 2015. Dostupné na: DOI: 10.5281/zenodo.3240591

28 Horstmann, Manuelle Annotation mit CATMA, c. d.

620 predpokladajú istú prácu s textom, porozumenie či interpretáciu. Takými javmi sú napríklad genetovské kategórie poriadku, trvania a frekvencie a v rámci kategórií napríklad podoby anizochronie (prolepsa či analepsa). J. Horstmann tiež v príspevku uvádza možnosť, aby sa užívateľ (anotátor) zamerlal na anotovanie a analýzu spôsobov reprodukcie reči (priama, nepriama, polopriama reč).



Ukážka používateľského rozhrania programu pri analýze poviedky *Tell-Tale Heart* od E. A. Poea²⁹

Manuálna anotácia literárnych textov tak napriek virtuálnemu prostrediu počítačového programu pripomína fyzické podčiarkovanie vybraných pasáží v knihe. František Čermák v tejto súvislosti analogicky opisuje rozsiahle korpusové zdroje a nástroje využívané v korpusovom skúmaní jazyka ako moderné obdoby kartotecných lístkov.³⁰ Z digitálneho charakteru skúmaného predmetu (elektronického textu) aj prostredia výskumu (počítačového programu) však v porovnaní s fyzickým anotovaním textu vyplývajú pridané možnosti, napríklad možnosti jednoduchých opráv anotácií či viacnásobných anotácií, ich štatistického vyhodnotenia, vizualizácie, a tiež možnosti kolaboratívneho a komparatívneho anotovania literárnych textov, ktoré by vo fyzickom prostredí narážalo na limity materiálneho nosiča, knihy či na limity ľudskej kognície (pamäti). Anotátor tiež môže definovať takzvané vlastnosti značiek, napríklad určitosť, pomocou ktorej anotátor do textu pridáva nielen metaúdaj o identifikovanom naratívnom jave (napríklad prolepse), ale aj meta-metaúdaj o miere istoty, s akou naratívny jav určil ako prolepsu. Táto možnosť je výhodná napríklad pri hodnotení významu motívu, keďže významová polyvalentnosť literárneho textu môže v čitateľovi vyvolať otázku, či je daný motív v texte kľúčový, alebo skôr satelitný, alebo si čitateľ-anotátor môže položiť otázku, kadiaľ vedú hranice nejakej udalosti (aký je jej dosah), či je udalosť vo fikčnom svete skutočná a pod. Tu tiež vďaka virtuálnemu charakteru pracovného prostredia možno porovnávať riešenia viacerých anotátorov súčasne.

Program CATMA tak z metodologického hľadiska predstavuje pokus o vytvorenie intersubjektívne dostupného predmetu, elektronického a elektronicky

29 Tutoriál vrátane ukážok dostupný na: <https://catma.de/how-to/tutorials/analyze-and-visualize/>

30 Korpusy sú podľa neho „zhruba řečeno moderním elektronickým ekvivalentem starých a rozsahem omezených manuálních excerpt z různých (většinou psaných) zdrojů (v podobě kartotecných lístků“ (ČERMÁK, František: *Korpus a korpusová lingvistika*. Praha: Karolinum, 2017, s. 13).

upravovateľného textu, a zároveň vytvorenie digitálnej platformy, rozhrania, ktoré umožňuje ďalšie spracovanie elektronického textu, jeho obohatenie o vrstvy metadát, a tak vlastne vytváranie akýchsi „kníh v knihách“. Vytvára sa tiež platforma pre literárnovednú diskusiu o anotovaných dielach i o samotných anotáciách, vizualizáciách a analýzach. Vedci združení okolo projektu pritom čiastkovo realizujú aj možnosti automatickej, nielen manuálnej anotácie literárnych korpusov.³¹ Autori projektu zasa deklarujú „nedogmatickosť“ (*undogmatic*) svojho počítačového nástroja, ktorý okrem preddefinovaných súborov značiek (*tagsetov*) umožňuje aj definíciu vlastných, prakticky ľubovoľných kategórií a taxonómií. Túto črtu programového dizajnu možno vnímať v súvislosti s tvrdením F. Čermáka, podľa ktorého sú každá taxonómia a každý systém kategórií do istej miery idiosynkratické, črtá sa v nich teoretická pozícia ich autora.³²

Uplatnenie programu CATMA na literárnovednú analýzu

Využitie programu CATMA na konkrétnom texte ukazuje Lena Schüchová (2015) na literárnovednej analýze poviedky *A Rose for Emily* (1930) od Williama Faulknera. Naratívnu sémantiku tohto diela odhaľuje pomocou kategórií klasickej naratológie, tak ako ich pri naratívnej analýze uplatňoval G. Genette. Na jednej strane ide o kategórie a značky vzťahujúce sa na fikčný čas v dielach, a to jednak rôzne gramatické časy (*tenses*), jednak časové údaje (*dates*). Pri časových údajoch sa rozlišujú nepriame reprezentácie času (*indirect-time-representation*)³³ a priame reprezentácie času (*direct-time-representation*).³⁴ Na druhej strane ide o kategórie vzťahu medzi sujetom a fabulou: usporiadanie (*order*), trvanie (*duration*) a frekvencia (*frequency*); v rámci usporiadania sa rozlišuje chronológia (*chronology*), anachronia (*anachrony*) a achronia (*achrony*); v rámci trvania izochronia (*isochrony*) a anizochronia (*anisochrony*); v rámci frekvencie sa rozlišujú singulatívne, repetitívne a iteratívne pasáže. Okrem toho sa fabula delí na jednotlivé segmenty.³⁵ Uvedené kategórie a druhy naratívnych javov sa ďalej delia v zhode s Genettovým systémom kategórií.

Na základe vykonanej manuálnej anotácie (vyznačenia, virtuálneho „podčiarknutia“) výskytov uvedených textových štruktúr potom anotátor môže postúpiť k vizualizácii, schéme vzťahu medzi jednotlivými segmentmi textu na úrovni sujetu a na úrovni fabuly.³⁶ Táto schéma napovedá, že hoci sa časová štruktúra Faulknerovej poviedky zvykne hodnotiť ako zložitá a náročná na vnímanie, schéma vzťahu medzi udalosťami vo fikčnom svete v ich prirodzenom poriadku na jednej strane a radením týchto udalostí v texte poviedky na druhej

31 Monograficky BRUNNER, Annelen: *Automatische Erkennung Von Redewiedergabe: Ein Beitrag Zur Quantitativen Narratologie. Narratologia 47*. Berlin – Boston : De Gruyter, 2015.

32 Čermák, Korpus a korpusová lingvistika, c. d., s. 37.

33 A v rámci nich tzv. sezóna (*seasonal-time*), udalosť (*event-related-time*), generácia (*generational-time*), plynutie času (*time passing*) (SCHÜCH, Lena: Computerphilologische Analyse. „Tagging in a huge meadow of time“ – Analysen der Zeit in Erzähltexten mit Hilfe des Programms CATMA. In: WEIXLER, Antonius – WERNER, Lukas (Hrsg.): *Zeiten erzählen. Ansätze – Aspekte – Analysen. Narratologia 48*. Berlin – Boston : De Gruyter, 2015, s. 36).

34 A v rámci nich tzv. časový bod (*time-point*), časový interval (*time-span*) a iteratívny výraz (*iterative expression*). Časový bod a časový interval sa ďalej delia na nevzťažné a vzťažné body/intervaly (*non-relative/relative*), a tieto opäť na ďalšie poddruhy informácií o fikčnom čase (tamže).

35 Tamže, s. 37.

36 Tamže, s. 43.

622 strane je pomerne jednoduchá: obe roviny sú v zrkadlovo obrátenom vzťahu, pričom osou tohto „obrátene“ je motív zápachu vychádzajúceho z domu hlavnej postavy, pani Emily. Motív zápachu, ako sa ukazuje na základe vizualizácie fabuly i sujetu, má v oboch rovinách rovnakú pozíciu: nachádza sa približne uprostred fikčnej chronológie a tiež uprostred literárneho textu.³⁷ Už formálne sa tak naznačuje možná interpretačná závažnosť uvedeného motívu.

Indície pri dešifrovaní naratívnej sémantiky textu však možno čerpať aj z ďalšej funkcionality programu CATMA, distribučnej analýzy daných výrazov v rámci textu. Tak možno zistiť, že manuálne anotované druhy gramatických časov (v angličtine napríklad predminulý či perfektum) sa vyskytujú v lineárne usporiadanom texte (zaznačenom na osi x) v rozličnej miere (zaznačenej na osi y).³⁸ Distribučné heterogenity daného výrazového prostriedku (napríklad nerovnomerná distribúcia časových údajov alebo striedanie rozprávačov v texte) môžu naznačiť prítomnosť interpretačne zaťažených miest, podobne ako to bolo v prípade nápadne zrkadlovej štruktúry vo vzťahu medzi fabulou a sujetom. Tieto možnosti analýz potom autorku uvedeného príspevku privádzajú až k poetologickým úvahám o filozofickom modeli času, ktorý vo Faulknerovskom fikčnom svete neplynú zakaždým rovnako rýchlo a možno ho, využitím napríklad zhustenia časových údajov, z perspektívy čitateľa aj z perspektívy postáv „prijbrzdíť“.³⁹

Naznačené kategórie naratívnej analýzy využíva v prostredí programu CATMA aj Evelyn Giusová,⁴⁰ ktorá svoju monografiu považuje za príspevok k „digitálnej naratológii“. Zameriava sa pritom na analýzu korpusu hovorených textov, vytvorených z rozhovorov s respondentmi. Témou rozhovorov sú medziludské konflikty respondentov na ich pracoviskách. Tento empirický materiál autorka analyzuje, podobne ako J. Meister,⁴¹ s cieľom objaviť invariantné štruktúry v rozprávaní o medziludských konfliktoch, pričom anotačný systém programu CATMA zároveň podrobne predstavuje.⁴²

Fakt, že tvorcovia programu zvolili ako preddefinovaný systém značiek (*tagset*) do veľkej miery prevzatý z genettovskej naratológie, naznačuje, že sa zároveň s proklamovanou „nedogmatickosťou“ a pluralitným prístupom prinajmenšom do istej miery hlásia k odkazu klasickej štrukturalistickej naratológie. Kvantitatívne skúmanie distribúcie jazykových javov pritom v literárnej vede má dlhú tradíciu napríklad v oblasti skúmania verša, distribúcie rytmických stôp.⁴³

Pokusy o odhalenie akejsi komputability naratívneho textu, o zodpovedanie otázky, či je naratívny text „vypočítateľný“ (*berechenbar*),⁴⁴ ktoré vidieť v Meisterovej monografii (2003) o Goetheho novelách, ale aj v monografii E. Giusovej

37 Tamže.

38 Tamže, s. 41, 45, 47.

39 Tamže, s. 48.

40 GIUS, Evelyn: *Erzählen über Konflikte. Ein Beitrag zur digitalen Narratologie. Narratologia* 46. Berlin – Boston : De Gruyter, 2015.

41 MEISTER, Jan Christoph: *Computing Action. A Narratological Approach. Narratologia* 2. Berlin – Boston : De Gruyter, 2003.

42 Gius, c. d., s. 70 – 97.

43 KOLÁR, Robert – PLECHÁČ, Petr: *Kapitoly z korpusové versologie*. Praha : Akropolis, 2017.

44 S odkazom na názov článku MEISTER, Jan Christoph: Kann man das Erzählen berechenbar machen? In: MÜLLER, Corinna – SCHEIDGEN, Irina (Hrsg.): *Erzählen, Archivieren, Beschreiben*. Marburg : Schriftenreihe der Gesellschaft für Medienwissenschaft 5, 2007, s. 19 – 39.

o analýze konfliktov na pracovisku (2015), pritom možno na jednej strane vnímať ako pokračovanie úsilia o odhalenie greimasovskej naratívnej gramatiky či proppovskej naratívnej morfológie. Tvorcovia programu CATMA, ako vyplýva aj z dizajnu anotátorských značiek, si však na druhej strane uvedomujú i istú vágnosť naratologických kategórií⁴⁵ a s ňou spojenú interpretačnú otvorenosť literárnych textov.⁴⁶ Táto otvorenosť znamená, že už stanovenie hraníc jednotlivkej fikčnej udalosti či významu nejakého motívu je otázkou interpretácie. Reflektovanie mnohoznačnosti naratívneho textu a možnosti počítačového spracovania tejto mnohoznačnosti sa zasa ozývajú v príspevku o tom, či je možné počítače naučiť, aby „pochoptili“ význam a hranice fikčnej udalosti.⁴⁷

Nadväznosť programu CATMA na myšlienky klasickej naratológie vidím tiež v istej paralele medzi anotačnými postupmi v programe a modelom analýzy naratívnej sémantiky od Lubomíra Doležela, ktorý sa sám hlásil k živému odkazu štrukturalizmu a Pražského lingvistického krúžku.⁴⁸ Anotačné postupy uplatňované v programe CATMA možno potom vnímať ako digitálnu obdobu analýzy povrchovej štruktúry literárneho textu. Túto povrchovú štruktúru označuje L. Doležel ako *textúru*, čo je „presná forma výrazu, pôvodní znění, v němž se motiv objevuje v literárním textu“.⁴⁹ Motív roztržitosti tak, napríklad, môže byť vyjadrený nielen tematickým obrazom, extenzionálne, ale aj prostredníctvom štylistických príznakov textu, formálne či intenzionálne, napríklad prerušovanými replikami postavy. Analyzované jazykové štruktúry v literárnom texte tak vytvárajú takzvanú intenzionálnu štruktúru, ktorá je „plně určena texturou fikčního textu“.⁵⁰ L. Doležel uvádza príklad *Robinsona Crusoa* od Daniela Defoea. Fikčný svet románu sa tu z hľadiska extenzionálnych prvkov (napríklad postáv) prostredníctvom textúry a intenzionálnej štruktúry delí na dve časti. Tie sa vytvárajú tak, že časť postáv sa označuje vlastnými menami (Robinson, Piatok), časť sa označuje jedinečnými opismi (môj otec, portugalský kapitán). Spôsob, akým sú v texte identifikované postavy, tak fikčný svet rozdeľuje na akési dve sféry. Podobne sa prostredníctvom textúry člení fikčný svet v *Procese* od Franza Kafku. L. Doležel v tomto prípade vytvára sústrednú kruhovú schému, kde sa uprostred nachádza postava K., v bližšom okolí postavy slečny Bürstnerovej, advokáta Hulda a Lenky, v širšom okolí postavy riaditeľa banky a sudcu.⁵¹ Kým hlavná postava románu je paradoxne intenzionálne (spôsobom danosti) identifikovaná iba iniciálou, užší okruh jej známych vystupuje pod priezviskom, širší okruh postáv sa identifikuje prostredníctvom svojho povolania.

45 BÖGEL, Thomas – GERTZ, Michael – GIUS, Evelyn – JACKE, Janina – MEISTER, Jan Christoph – PETRIS, Marco – STRÖTGEN, Jannik: Collaborative Text Annotation Meets Machine Learning: heureCLÉA, a Digital Heuristic of Narrative. In: *DH Commons Journal* 1. July 2015. Dostupné na: DOI: 10.5281/zenodo.3240591, s. 21 – 24.

46 GIUS, Evelyn – JACKE, Janina: Informatik und Hermeneutik. Zum Mehrwert interdisziplinärer Textanalyse. In: BAUM, Constanze – STÄCKER, Thomas (eds.): *Grenzen und Möglichkeiten der Digital Humanities*. 2014. ZfdG-Sonderband 1. Dostupné na: http://zfdg.de/sb001_006, s. 130.

47 DUNN, Stuart – SCHUMACHER, Mareike: Explaining Events to Computers: Critical Quantification, Multiplicity and Narratives in Cultural Heritage. In: *Digital Humanities Quarterly*, roč. 10, 2016, č. 3, s. 1 – 21.

48 DOLEŽEL, Lubomír: *Narativní způsoby v české literatuře*. Příbram : Pistorius & Olšanská, 2014, s. 9.

49 DOLEŽEL, Lubomír: *Heterocosmica. Fikce a možné světy*. Praha : Karolinum, 2003, s. 49.

50 Tamže, s. 146.

51 Tamže, s. 148.

Schéma rozdelenia fikčných postáv podľa spôsobu ich označovania v texte predstavuje podobu vizualizácie naratívnej štruktúry, tak ako túto štruktúru možno skúmať v prostredí programu CATMA. Metodologický základ programu preto tiež vnímam v súvislosti s tézami vtelenej kognície (*embodied cognition*). Kognícia ako „ekologický“ proces⁵² vzniká v interakcii subjektu s okolitými predmetmi, vrátane vlastného tela subjektu a jeho zmyslov. Technologické pomôcky pritom možno považovať za kvantitatívne i kvalitatívne extenzie zmyslov. V prípade korpusovej lingvistiky sa spomína metafora korpusu ako mikroskopu či teleskopu,⁵³ ktorý jazykovedcovi umožňuje odhaliť veci, aké by inak nemal možnosť vidieť. V prípade anotácií a vizualizácií v programe CATMA ide analogicky o akúsi rozšírenú a intersubjektívne dostupnú a čitateľskú pamäť, o druh mnemotechnickej pomôcky.⁵⁴ Program CATMA zároveň predstavuje podobu rozhrania (*interface*), ktoré sa reflektuje aj v našom geografickom priestore.⁵⁵ Koncept rozhrania charakterizuje Jakub Albert Ferenc ako „stýčný bod, kde se potkávají člověk a technologie“ a ktorý má „vždy za úkol vyrovnávat nekompatibility mezi člověkem a technologií, a to tím způsobem, že vždy podle nějaké strategie schovává komplexitu technologie i světa a nabízí člověku jen ty funkce a možnosti, které potřebuje“.⁵⁶ Vizualizácia naratívnych štruktúr tak predstavuje účinnú formu redukcie dimenzionality, mnohorozmernosti, akou sa vyznačuje nielen sama príroda,⁵⁷ ale prirodzene aj technológia, keďže ľudský subjekt je schopný efektívne prijímať, identifikovať iba ohraničené množstvo podnetov, podobne ako má kapacitné limity jeho pamäť.

V tejto súvislosti sa na jednej strane otvára filozofická otázka prezentácie a reprezentácie sveta v jazyku, otázka „jazykov reprezentácie“.⁵⁸ Prítomnosť okolitého sveta sa na druhej strane z fenomenologického hľadiska javí ako proces, ktorý tvoria stopy tohto sveta z minulosti. Tento proces „často nejsme schopni sledovat v celém jeho trvání. Setkáváme se buďto s částí (nebo částmi) aktuálně probíhajícího procesu, anebo nějakou částí (částmi) procesu už uplynulého. Tato fragmentárnost zjevování procesu nemusí znamenat, že nám proces jako celek uniká. Je určitým způsobem spolupřítomný, neboť je s těmi částmi, s nimiž se setkáváme, označován“.⁵⁹ Počítačovo podporovanú naratívnu analýzu a vizualizáciu v programe CATMA, a teda počítačovo podporované objavovanie potenciálne významných štruktúr v naratívnom texte, potom možno interpretovať ako konštrukciu istých intersubjektívne dostupných „stôp“ čitateľskej skúsenosti s textom.

52 KUSÁ, Daniela (ed.): *Ekológia ľudských kognícií*. Bratislava : Veda, 2011.

53 JAROŠOVÁ, Alexandra: Špirála poznania alebo funkcionalizmus (v lexikológii) cez optiku dvoch kongresových jubileí. In: *Slovenská reč*, roč. 83, 2018, č. 2, s. 133.

54 „The sciences and the technologies have visual aids to memory generated by formal languages and graphics, and now they are being enhanced exponentially by computers“ (YOOS, George E.: *Simplifying Complexity. Rhetoric and the Social Politics of Dealing with Ignorance*. Warsaw – Berlin : de Gruyter Open 2015, s. 4).

55 SUWARA, Bogumila (ed.): *{(staré a nové) rozhrania / *interfejsy* / [literatúry]}*. Bratislava : Ústav svetovej literatúry SAV, 2014.

56 FERENC, Jakub Albert: *Věda o interface v teorii rozšířené kognice na pozadí McLuhanovy mediální teorie*. Olomouc : Univerzita Palackého, 2019. Dostupné na: <https://www.academia.edu/3851738>

57 MELOUN, Milan – MILITKÝ, Jiří – HILL, Martin: *Statistická analýza vícerozměrných dat v příkladech*. Praha : Academia, 2012, s. 17.

58 VEBEROVÁ, Veronika – BÍLEK, Petr A. – PAPOUŠEK, Vladimír – SKALICKÝ, David (eds.): *Jazyky reprezentace*. Praha : Akropolis, 2012.

59 AJVAZ, Michal – PAUKNEROVÁ, Karolína: *Spor o proces a událost*. Praha : Pavel Mervart, 2019, s. 19.

Takto skonštruovaný objekt má rôznu podobu: funkcionality programu CATMA umožňujú vytvoriť a vizualizovať 1. frekvenčný zoznam najčastejších slov v texte, 2. distribúciu nejakého jazykového javu v lineárnom priebehu textu alebo 3. vytvoriť tzv. dvojité strom (*double tree*), ktorým sa zobrazuje spájateľnosť vybraného výrazu v texte s predchádzajúcimi a nasledujúcimi výrazmi v okolí vybraného výrazu, čo umožňuje charakterizovať preferencie autora pri výbere výrazov, jeho štýl. Takéto technologicky podporované spracovanie literárneho textu vnímam v súvislosti s reflexiami Martina Heideggera o povahe modernej vedy, ako ich predstavil v spise *Věk obrazu sveta*. Takzvaná „bytnosť“ modernej vedy podľa neho spočíva v „rozvrhovaní okrskov súcna“: „Má-li být rozvržený okrsek okrskem předmětů, pak je třeba nechat jej vystoupit v celé jeho vrstevnatosti a spleitosti.“⁶⁰ Vytváranie anotácií a vizualizácií naratívnych textov v nadväznosti na túto myšlienku interpretujem ako činnosť, ktorej zmyslom je nechať vystúpiť štruktúry naratívneho textu v ich vrstevnatosti a spleitosti. Konceptuálna metafora vrstevnatosti a spleitosti sa potom v programe CATMA realizuje ako „vrstva“ pridaných anotátorských značiek, metadát. Konceptuálna metafora „vrstvy“ zasa odkazuje na niektoré filozofické koncepty: leibnizovský koncept záhybu⁶¹ i na koncept spleti,⁶² ale odkazuje aj na empirické postupy súčasnej vedy, na metódy štatistickej analýzy viacrozmerých dát. Cieľom týchto metód, ako je napríklad analýza hlavných komponentov (*Principal Component Analysis*), je totiž „zpracovat data tak, aby se zřetelně indikoval model a tak odkryl skrytý jev“.⁶³ Skrytý jav možno odhaliť ako „strukturu a vazby ve znacích“⁶⁴ alebo ako „strukturu a vazby v objektech“⁶⁵ a takýmito znakmi a objektmi sú v prípade programu CATMA práve výrazové prostriedky, jazykové výrazy tvoriace doleželovskú textúru. Ich charakteristika a distribučné pravidelnosti či nepravidelnosti tejto textúry potom literárnemu vedcovi ponúkajú možnosť, aby skonštruoval istý model naratívneho textu. Tento model je na jednej strane už výsledkom istej interpretácie a na druhej strane je východiskom ďalších interpretácií, čo považujem za zaujímavé rozšírenie inštrumentária literárnovedného výskumu.

Záver

Kvantitatívne prístupy k výskumu v humanitných vedách majú tradíciu aj na Slovensku, napríklad v prácach o matematickej lingvistiky od Jána Horeckého. Záujem o kvantifikáciu výskumných otázok badať prirodzene i v prácach klasika slovenskej štylistiky, Jozefa Mistríka. V súčasnosti zasa treba vyzdvihnúť možnosti, aké pre kvantitatívny výskum nielen literárneho jazyka poskytujú rozsiahle i špecializované zdroje Slovenského národného korpusu.⁶⁶ Súčasné, ale i prípadne rozšírené, ciele budované autorské korpusy slovenských autorov a literárnych období by

60 HEIDEGGER, Martin: *Věk obrazu světa*. Praha : Oikoymenh, 2013, s. 13.

61 DELEUZE, Gilles: *Záhyb*. Praha : Hermann & synové, 2015.

62 NITSCHKE, Martin: *Metodická přednost spleti. Transzitivně-topologický model fenomenologie*. Praha : Togga, 2016.

63 Meloun – Militký – Hill, c. d., s. 69.

64 Tamže, s. 55.

65 Tamže.

66 Prehľad veľkej časti dostupných korpusov zo zdrojov Slovenského národného korpusu pozri na: <https://korpus.sk/structure1.html>

626 tak vďaka uplatňovanej čiastočne automatickej i manuálnej anotácii mohli tvoriť solidnú materiálovú bázu literárnovedných výstupov, podobne ako na základe autorských korpusov vznikol slovník autorských štýlov Karla Čapka,⁶⁷ Bohumila Hrabala⁶⁸ a Jana Čepa.⁶⁹ Analýzy zamerané na kvantitatívne skúmanie štýlu slovenských autorov však existujú už teraz, pričom z novších možno spomenúť výskum polysémie rýmovaných slov v Sládkovičových básňach *Marína* a *Morava*.⁷⁰ Na existujúce predpoklady a tradíciu kvantitatívneho výskumu slovenských diel z poetologického hľadiska možno nadväzovať aj prostredníctvom analýzy vybraných textových jednotiek (slovesných časov, tematických slov, spôsobov reprodukcie reči, autorských idiolektov a iných) v prostredí programu CATMA. Zaujímavé by mohli byť aj nadväzujúce analýzy rozsiahlejších korpusov, ktoré Jan Horstmann zaradil k prístupu tzv. dištančného čítania.⁷¹ Ako relevantné výskumné otázky sa črtajú napríklad analýzy staršej slovenskej literatúry, prípadne časopiseckých zdrojov či žánrovej literatúry mimo kánonu. Túto rozsiahlu, hoci sčasti menej vnímanú oblasť literárnej produkcie označuje Margaret Cohenová ako „veľkú neznámu“ či „neprečítanú literatúru“ (*the great unread*).⁷² Hoci autorsky proklamovanú nedogmatickosť programu CATMA treba metodologicky reflektovať, o čo som sa pokúsil v príspevku, dúfam zároveň, že o používateľskej ústretovosti programu sa budú môcť presvedčiť aj slovenskí literárni vedci.

Literatúra

- AJVAZ, Michal – PAUKNEROVÁ, Karolína: *Spor o proces a událosť*. Praha : Pavel Mervart, 2019.
- BÖGEL, Thomas – GERTZ, Michael – GIUS, Evelyn – JACKE, Janina – MEISTER, Jan Christoph – PETRIS, Marco – STRÖTGEN, Jannik: Collaborative Text Annotation Meets Machine Learning: heureCLÉA, a Digital Heuristic of Narrative. In: *DHCommons Journal 1*, July 2015. Dostupné na: DOI: 10.5281/zenodo.3240591
- BRUNNER, Annelen: *Automatische Erkennung Von Redewiedergabe: Ein Beitrag Zur Quantitativen Narratologie*. *Narratologia* 47. Berlin – Boston : De Gruyter, 2015.
- BURDICKOVÁ, Anne – Druckerová, Johanna – Lunenfeld, Peter – Presner, Todd – Schnapp, Jeffrey: *Digital Humanities*. Prel. David Vichnar. Praha : Academia, 2019.
- COECKELBERGH, Mark – FUNK, Michael: Wittgenstein as a Philosopher of Technology: Tool Use, Forms of Life, Technique, and a Transcendental Argument. In: *Human Studies*, roč. 41, 2018, č. 2, s. 165 – 191.
- COECKELBERGH, Mark: *Using Words and Things. Language and Philosophy of Technology*. New York/Oxon : Routledge, 2017.

67 ČERMÁK, František: *Slovník Karla Čapka*. Praha : Lidové noviny, 2009.

68 CVRČEK, Václav – ČERMÁK, František: *Slovník Bohumila Hrabala*. Praha : Lidové noviny, 2009.

69 ZMĚLÍK, Richard: *Kvantitativně-korpusová analýza a literární věda. Model a realizace autorského korpusu a slovníku Jana Čepa v kontextu zahraniční a české autorské lexikografie*. Olomouc : Univerzita Palackého, 2015.

70 KOLENČIKOVÁ, Natália – MÍŠTECKÝ, Michal – ALTMANN, Gabriel: Polysemy of Rhyme Words: A Case Study of Two Slovak Poems. In: *Glottometrics*, roč. 49, 2020, s. 13 – 31.

71 Horstmann, Textvisualisierung: Epistemik des Bildlichen im Digitalen, c. d.

72 COHEN, Margaret: *The Sentimental Education of the Novel*. Princeton (NJ) : Princeton University Press, 2002.

- COHEN, Margaret: *The Sentimental Education of the Novel*. Princeton (NJ) : Princeton University Press, 2002.
- CVRČEK, Václav – ČERMÁK, František: *Slovník Bohumila Hrabala*. Praha : Lidové noviny, 2009.
- ČERMÁK, František: *Korpus a korpusová lingvistika*. Praha : Karolinum, 2017.
- ČERMÁK, František: *Slovník Karla Čapka*. Praha : Lidové noviny, 2009.
- DEBNÁR, Marek: O esejistickom štýle z kvantitatívneho hľadiska alebo o možnostiach využitia korpusov v širšom humanitnom výskume. In: ŠIMKOVÁ, Mária – LEVICKÁ, Jana – DEBNÁR, Marek: *Dynamické javy v súčasnej slovenčine a jej výskume*. Bratislava : Veda, 2018, s. 165 – 192.
- DELEUZE, Gilles: *Záhyb*. Praha : Hermann & synové, 2015.
- DESAGULIER, Guillaume: Can word vectors help corpus linguists? In: *Studia Neophilologica*, roč. 91, 2019, č. 2, s. 1 – 21.
- DESAGULIER, Guillaume: *Corpus Linguistics and Statistics with R. Introduction to Quantitative Methods in Linguistics*. Cham : Springer International Publishing, 2018.
- DOLEŽEL, Lubomír: *Heterocosmica. Fikce a možné světy*. Praha : Karolinum, 2003.
- DOLEŽEL, Lubomír: *Narativní způsoby v české literatuře*. Příbram : Pistorius & Olšanská, 2014.
- DOLNÍK, Juraj: *Jazyk v sociálnej kultúre*. Bratislava : Veda, 2017.
- DUNN, Stuart – SCHUMACHER, Mareike: Explaining Events to Computers: Critical Quantification, Multiplicity and Narratives in Cultural Heritage. In: *Digital Humanities Quarterly*, roč. 10, 2016, č. 3, s. 1 – 21.
- EICHINGER, Ludwig M.: Linguisten brauchen Korpora und Korpora Linguisten: Wege zu wohl dokumentierten und verlässlichen Aussagen über Sprache. In: KALLMEYER, Werner – ZIFONUN, Gisela (eds.): *Sprachkorpora: Datenmengen und Erkenntnisfortschritt*. Berlin – New York : Walter de Gruyter, 2007, s. 1 – 8.
- FERENC, Jakub Albert: *Věda o interface v teorii rozšířené kognice na pozadí McLuhanovy mediální teorie*. Olomouc : Univerzita Palackého, 2019. Dostupné na: <https://www.academia.edu/3851738>
- GIUS, Evelyn: *Erzählen über Konflikte. Ein Beitrag zur digitalen Narratologie*. *Narratologia* 46. Berlin – Boston : De Gruyter, 2015.
- GIUS, Evelyn – JACKE, Janina: Informatik und Hermeneutik. Zum Mehrwert interdisziplinärer Textanalyse. In: BAUM, Constanze – STÄCKER, Thomas (eds.): *Grenzen und Möglichkeiten der Digital Humanities*. ZfdG-Sonderband 1. 2014. Dostupné na: http://zfdg.de/sb001_006
- GIUS, Evelyn – JACKE, Janina: The Hermeneutic Profit of Annotation. On Preventing and Fostering Disagreement in Literary Analysis. In: *International Journal of Humanities and Arts Computing*, roč. 11, 2017, č. 2, s. 233 – 254.
- GOGORA, Andrej: Digital humanities a filozofia – problém digitálneho výskumu. In: *World Literature Studies*, roč. 8, 2016, č. 3, s. 104 – 114.
- HEIDEGGER, Martin: *Věk obrazu světa*. Praha : Oikoymenh, 2013.
- HORSTMANN, Jan: Textvisualisierung: Epistemik des Bildlichen im Digitalen. In: HUBER, Martin – KRÄMER, Sybille – PIAS, Claus (Hrsg.): *Wovon sprechen wir, wenn wir von Digitalisierung sprechen? Gehalte und Revisionen zentraler Begriffe des Digitalen. Digitalität in den Geisteswissenschaften*. Bayreuth : DFG-geförderte Symposienreihe Universität Bayreuth, 2020, s. 154 – 173.
- JAROŠOVÁ, Alexandra: Špirála poznania alebo funkcionalizmus a kognitivismus (v lexikológii) cez optiku dvoch kongresových jubileí. In: *Slovenská reč*, roč. 83, 2018, č. 2, s. 123 – 142.
- KOLÁR, Robert – PLECHÁČ, Petr: *Kapitoly z korpusové versologie*. Praha : Akropolis, 2017.

- 628 KOLENČÍKOVÁ, Natália – MÍSTECKÝ, Michal – ALTMANN, Gabriel: Polysemy of Rhyme Words: A Case Study of Two Slovak Poems. In: *Glottometrics*, roč. 49, 2020, s. 13 – 31.
- KUSÁ, Daniela (ed.): *Ekológia ľudských kognícií*. Bratislava : Veda, 2011.
- LANCASHIRE, Ian – BRADLEY, John – McCARTY, Willard – STAIRS, Michael – WOOLDRIDGE, T. R.: *Using TACT with Electronic Texts: A Guide to Text-Analysis Computing Tools, Version 2.1 for MS-DOS and PC DOS*. New York : Modern Language Association of America, 1996.
- MEISTER, Jan Christoph: From TACT to CATMA or A mindful approach to text annotation and analysis. In: NYHAN, Julianne – ROCKWELL, Geoffrey – SINCLAIR, Stéfan (eds.): *On Making in the Digital Humanities: Essays on the Scholarship of Digital Humanities Development in Honour of John Bradley*. 2020. Dostupné na: https://jmeister.de/downloads/texts/Meister_2020-TACT-to-CATMA.pdf
- MEISTER, Jan Christoph: *Computing Action. A Narratological Approach*. *Narratologia* 2. Berlin – Boston : De Gruyter, 2003.
- MEISTER, Jan Christoph: Kann man das Erzählen berechenbar machen? In: MÜLLER, Corinna – SCHEIDGEN, Irina (Hrsg.): *Erzählen, Archivieren, Beschreiben*. Marburg : Schriftenreihe der Gesellschaft für Medienwissenschaft 5, 2007, s. 19 – 39.
- MELOUN, Milan – MILITKÝ, Jiří – HILL, Martin: *Statistická analýza vícerozměrných dat v příkladech*. Praha : Academia, 2012.
- MOULIN, Claudine: Am Rande der Blätter. Gebrauchsspuren, Glossen und Annotationen in Handschriften und Büchern aus kulturhistorischer Perspektive. In: *Quarto. Zeitschrift des Schweizerischen Literaturarchivs*, 30/31, 2010, s. 19 – 26.
- NITSCHKE, Martin: *Metodická přednost spleti. Transzitivně-topologický model fenomenologie*. Praha : Togga, 2016.
- REPPLINGER, Michael – BEINBORN, Lisa – ZUIDEM, Willem: Vector-space models of words and sentences. In: *Nieuw Archief voor Wiskunde (NAW journal)*, roč. 5/19, 2018, č. 3, s. 167 – 174.
- SIEMENS, Ray – SCHREIBMAN, Susan (eds.): *A Companion to Digital Literary Studies*. Oxford : Blackwell, 2008. Dostupné na: <http://www.digitalhumanities.org/companionDLS/>
- SCHÜCH, Lena: Computerphilologische Analyse. „Tagging in a huge meadow of time“ – Analysen der Zeit in Erzähltexten mit Hilfe des Programms CATMA. In: WEIXLER, Antonius – WERNER, Lukas (Hrsg.): *Zeiten erzählen. Ansätze – Aspekte – Analysen*. *Narratologia* 48. Berlin – Boston : De Gruyter, 2015, s. 27 – 52.
- SUWARA, Bogumila (ed.): *{(staré a nové) rozhrania / *interfejsy* / [literatúry]}*. Bratislava : Ústav svetovej literatúry SAV, 2014.
- SVÍTEK, Miroslav: *Víc než součet částí: Systémový pohled na proces lidského poznání*. Praha : Academia, 2013.
- VAN CRANENBURGH, Andreas – VAN DALEN-OSKAM, Karina – VAN ZUNDERT, Joris: Vector space explorations of literary language. In: *Lang Resources & Evaluation*, roč. 53, 2019, s. 625 – 650.
- VEBEROVÁ, Veronika – BÍLEK, Petr A. – PAPOUŠEK, Vladimír – SKALICKÝ, David (eds.): *Jazyky reprezentace*. Praha : Akropolis, 2012.
- WERNER, Lukas: Zeit im Text – Möglichkeiten der formalen Historisierung. Zur Relation von Zeit und Ich-Erzähler in Grimmselhausens *Simplicissimus* und Schnabels *Insel Felsenburg*. In: WEIXLER, Antonius – WERNER, Lukas (Hrsg.): *Zeiten erzählen. Ansätze – Aspekte – Analysen*. *Narratologia* 48. Berlin – Boston : De Gruyter, 2015, s. 79 – 104.
- YOOS, George E.: *Simplifying Complexity. Rhetoric and the Social Politics of Dealing with Ignorance*. Warsaw – Berlin : de Gruyter Open, 2015.
- ZMĚLÍK, Richard: *Kvantitativně-korpusová analýza a literární věda. Model a realizace autorského korpusu a slovníku Jana Čepa v kontextu zahraniční a české autorské lexikografie*. Olomouc : Univerzita Palackého, 2015.

Elektronické zdroje

- GIUS, Evelyn – JACKE, Janina: *Zur Annotation narratologischer Kategorien der Zeit Guidelines zur Nutzung des CATMA-Tagsets*. 2016. Dostupné na: <http://heureclea.de/wp-content/uploads/2016/11/guidelinesV2.pdf>
- HORSTMANN, Jan: Manuelle Annotation mit CATMA. In: *forTEXT. Literatur digital erforschen*. Dostupné na: <https://fortext.net/routinen/lerneinheiten/manuelle-annotation-mit-catma>. 2019
- INTRONA, Lucas: Phenomenological Approaches to Ethics and Information Technology. In: ZALTA, Edward N. (ed.): *The Stanford Encyclopedia of Philosophy*. 2017. Dostupné na: <https://plato.stanford.edu/archives/fall2017/entries/ethics-it-phenomenology/>
- MÜLLER, Martin: „Scalable Reading“, *Scalable Reading* (blog). May 29, 2012. Dostupné na: https://scalablereading.northwestern.edu/?page_id=22
- Slovenský národný korpus. Dostupné na: <https://korpus.sk>