

Budovanie komplexných a reprezentatívnych digitálnych literárnych zbierok v rámci Európskej zbierky literárnych textov (ELTeC) na Slovensku a Ukrajine

Marek Debnár – Dmytro Jesypenko

DEBNÁR, M. – YESYPENKO, D.: Building of Complex and Representative Digital Literary Collections within the European Literary Text Collection (ELTeC) in Slovakia and in Ukraine
SLOVENSKÁ LITERATÚRA 67, 2020, No. 6, p. 630 – 638

DOI: <https://doi.org/10.31577/slovlit.2020.67.6.8>

Marek Debnár – ORCID ID: 0000-0002-2879-6878

Dmytro Yesypenko – ORCID ID: 0000-0003-4720-8370

Key words: European Literary Text Collection (ELTeC), distant reading, national literatures, literary history, Slovakia, Ukraine

The paper deals with problems of building representative digital literary collections of national literatures, in particular those of Slovak and Ukrainian literatures, within the international context of the project *European Literary Text Collection* (ELTeC). At the beginning text characterizes the requirements laid down on creation and representativeness of such a national multilingual literary collection, which is established as a result of the project COST – Distant Reading for European Literary History (1840 – 1920) and analyzes various factors related to literary production in Slovakia and Ukraine. On the one hand, it includes a discrepancy between stronger partners of the project (so-called great European literatures) and weaker partners (so-called small literatures), on the other hand, the factor of accessibility of digital sources corresponding to the requirements of a project of this kind.

Paradigma dištančného čítania, teda využívanie počítačových metód pri analýzach rozsiahlych zbierok literárnych textov, medzi ktoré patria kvantitatívna analýza, štylometria, rozpoznávanie autorstva či rozpoznávanie emócií, sa v posledných dvoch desaťročiach stále viac dostáva do centra záujmu v interdisciplinárnom humanitnom výskume. Pozornosť sa mu venuje nielen v kontexte digitálnych humanitných vied (tzv. *digital humanities*), ale vo všetkých odboroch, ktorých výskum sa primárne opiera o interpretáciu textov, tradične označovanú ako *close reading* (literárna veda, historiografia, filozofia, etika, náboženské štúdiá a ďalšie).¹

Na Slovensku a ani na Ukrajine, žiaľ, nedisponujeme dostatočne rozsiahlou, t. j. relatívne kompletnou a spracovanou digitálnou literárnou zbierkou, aké sú bežné v zahraničí, kde okrem kompletných zbierok klasikov národných literatúr (napr. Archív stredovekej severskej literatúry na Univerzite v Bergene, britský Brownov korpus či kompletne sprístupnené diela Williama Shakespeara a Jane Austenovej) môžeme nájsť aj veľké zbierky populárnej literatúry, akou je stanfordská FunFiction či desiatky autorských korpusov od Wittgensteinovho archívu v Nórsku, Mickiewiczovho v Poľsku alebo obdobných projektov v susedných Čechách (napr. Korpus Karla Čapka či Korpus českého verša, ako aj špecializované digitálne archívy periodickej tlače).²

Určitá nádej na zmenu sa objavila v roku 2019, keď sa Slovensko ako riadny člen a Ukrajina so štatútom „blízkeho suseda“ pripojili k medzinárodnému projektu COST Dištančné čítanie európskej literárnej histórie (1840 – 1920) (CA16204: Distant reading for European Literary History 1840 – 1920), ktorý vedú Christof Schöch (Univerzita Trier) a Maciej Eder (Poľská akadémia vied). Táto iniciatíva spája výskumné tímy z tridsiatich piatich krajín, čo predstavuje spolu viac ako sto účastníkov.³ Názov projektu odkazuje na nadnárodnú ideu

1 V tomto duchu prezentoval Franco Moretti aj svoj súčasný výskum na viacerých stretnutiach DR2 skupiny (DistantReading and Data-Driven Reseach in the History of Philosophy) na pôde Turinskej univerzity, ktorá je príkladom využitia korpusových metód práce s textom za hranicami jazykovedy a literatúry. Informácie o DR2 skupine dostupné na: <https://dr2blog.hcommons.org>. Na Slovensku tento prístup dlhodobo absentoval a objavil sa až v posledných rokoch, keď ho začali vo svojich prácach využívať autori hlásiaci sa k metódam a postupom digitálnych humanitných vied, napr. GOGORA, Andrej: The text mining of ethics and information technology. In: *Communications. Scientific Letters of the University of Zilina*, roč. 19, 2017, č. 1, s. 3 – 8; JUROVÁ, Jarmila – DEBNÁR, Marek: Interpretation of Values in the Epitaphs on the Tombstones of Old Jewish Cemetery in Bratislava, 2018. In: *Communications. Scientific Letters of the University of Zilina*, roč. 20, 2018, č. 1A, s. 71 – 79. V rokoch 2019 a 2020 vznikli aj prvé pôvodné slovenské monografie na tieto témy: DEBNÁR, Marek – GOGORA, Andrej: *Digitálne trendy v súčasných humanitných vedách*. Nitra : UKF, 2019; ZUMRÍK, Miroslav: *Filozofické aspekty korpusovej lingvistiky*. Bratislava : Veda, 2020.

2 V prípade ukrajinskej literatúry existuje pár výnimiek, napr. Ševčenkova portál (<http://kobzar.ua/>), ktorý obsahuje kompletnú zbierku čistých textov T. Ševčenka a skeny jeho rukopisov a tlačenej vydani. Na Slovensku zatiaľ nemáme takto vyčerpávajúco spracovaný autorský korpus.

3 Slovensko a Ukrajina sú krajiny, ktoré takpovediac naskočili na posledný vlak, keďže sa projekt začal ešte v roku 2017 a obe krajiny sa doň zapojili až ku koncu roka 2019 (autori tohto článku sú zároveň zástupcami týchto krajín v projekte). V prípade Slovenska išlo najmä o nedostatok personálnych kapacít pre takýto projekt a v prípade Ukrajiny ťažkosť súviseli so skutočnosťou, že Ukrajina nemá štatút „členskej krajiny COST“, ale „krajiny blízkeho suseda“ (<https://www.cost.eu/actions/CA16204/#tabs|Name:-management-committee>).

632 „európskej literárnej histórie“, ale pri bližšom skúmaní vyjde najavo, že sa skôr zameriava na konkrétne európske literatúry a vzťahy medzi nimi.

Úlohy a očakávané výstupy projektu sú podľa oficiálnej stránky projektu nasledovné:

1. vybudovanie viacjazyčnej Európskej zbierky literárnych textov (ELTeC), ktorá bude obsahovať okolo 2 500 kompletných románových textov v najmenej desiatich rôznych jazykoch, čo umožní otestovať výpočtové metódy a porovnať výsledky naprieč národnými tradíciami;
2. vytvoriť a zdieľať inovatívne nástroje a metódy textovej analýzy prispôbené viacjazyčnej literárnej tradícii Európy;
3. na základe výsledkov analýz prehodnotiť základné koncepty v literárnej teórii a histórii.⁴

Ciele projektu sú teda v súlade so základnou koncepciou dištančného čítania Franca Morettiho a jeho predstavami o zmene paradigmy nášho chápania literárnych dejín.⁵ Spoločným menovateľom prístupov a metód zjednotených pod pojmom dištančného čítania je využívanie elektronických textových korpusov v oblasti literárnej histórie. Cieľom COST iniciatívy, ktorá sa usiluje o vybudovanie zbierky ELTeC, je svojou multilingválnou a viacnárodnou platformou odhaliť skryté prepojenia národných literatúr v širších európskych súvislostiach, ako aj reagovať na kritikov koncepcie dištančného čítania, ktorí vyčítali tomuto prístupu anglocentrizmus.⁶ Voči týmto výčitkám sa síce ohradil aj sám Moretti,⁷ zbierka ELTeC je však v oveľa väčšej miere zameraná na podporu inkluzívnosti týchto skúmaní, vrátane jazykovej.

Jadrom projektu je vytvorenie viacjazyčnej zbierky ELTeC, pričom táto vstupná fáza zahŕňa prípravu digitálnych zbierok národných literatúr, ktoré tvoria jeho podskupiny. Príprava jednotlivých častí zbierky znamená súbor povinností, ktoré sa u rôznych národných tímov rôznia. V niektorých literatúrach existujú veľké alebo dokonca rozsiahle zbierky voľne dostupných textov, ktoré vznikli v rámci predchádzajúcich výskumných projektov alebo v rámci digitalizačných aktivít národných knižníc. V prípade mnohých ďalších literatúr je však získavanie textov pre ELTeC vážnou výzvou. V každej národnej literatúre spravidla existuje obrovské množstvo románov publikovaných v periodikách alebo vo forme samostatných vydání dostupných v knižniciach. Na rôznych internetových stránkach je tiež k dispozícii relatívne veľa digitalizovaných verzií týchto publikácií. Pokiaľ však ide o romány kanonizované a prevedené do formátu čistých textových súborov, teda pripravené na rýchle a relatívne pohodlné zaradenie do korpusov podobných korpusom ELTeC, ich počet je pomerne nízky.

Viacjazyčnosť limituje aj kritériá zostavovania zbierky ELTeC. Na rozdiel od korpusov národných literatúr, ktoré sú šité na mieru konkrétnej tradície, medzinárodný projekt sa musí riadiť kritériami vyváženej najmä preto, aby

4 Tento zoznam nie je úplný. Cieľom projektu je tiež podporiť odbornú prípravu a rozvoj začínajúcich riešiteľov, podporiť nové spoločné predkladanie návrhov do národných a európskych projektov a zabezpečiť vyvážené zastúpenie mužov a žien v počítačových vedách a príbuzných oblastiach (<https://www.distant-reading.net/about/objectives>).

5 MORETTI, Franco: Conjectures on World Literature. In: *New Left Review*, 2000, č. 1, s. 54 – 68.

6 ARAC, Jonathan: Anglo-globalism? In: *New Left Review*, 2002, č. 16, s. 35 – 45.

7 MORETTI, Franco: *Distant Reading*. London : Verso, 2013.

sa odstránila diskrepancia medzi veľkými a malými literatúrami. Je zrejmé, že slovenská literatúra z obdobia rokov 1840 – 1920 neobsahuje toľko literárnych diel ako literatúra nemecká alebo francúzska, a preto je aj výber jednotlivých titulov značne limitovaný.

Základné formálne kritériá pre diela zaradené do zbierky sú nasledovné:

Jazyk: pôvodné diela napísané v danom európskom jazyku, nie preklady.

Štýl textov je definovaný ako fiktívne prozaické rozprávania (projekt nedefinuje žáner románu z hľadiska literárnej teórie; väčšia pozornosť sa venuje vlastnostiam jednotlivých zložiek digitálnej zbierky na úrovni slova a spôsobu spracovania textov).

Časové obdobie: diela vydané knižne v rokoch 1840 – 1920.

Dĺžka textov: minimálna dĺžka jedného textu je desatisíc slov.

Publikovanie: v súčasnej fáze projektu sú preferované texty románov vydávané v knižných edíciách, je však možné zaradiť do zbierky aj romány vydávané časopisecky na pokračovanie.

Dostupnosť: primárne je zámerom zaradiť do zbierky len voľne dostupné digitalizované texty alebo texty nepodliehajúce autorským právam, ktoré bádatelia – riešitelia projektu sami zdigitalizujú.

Na Slovensku existuje niekoľko digitálnych zdrojov literárnych textov, od oficiálnych literárnych archívov knižníc (napr. digitálny archív Slovenskej národnej knižnice)⁸ cez verejné projekty postavené na dobrovoľníckej báze (napr. Zlatý fond denníka SME)⁹ až po neoficiálne súkromné internetové stránky, ktoré sa vyznačujú problematickými alebo nevysporiadanými vzťahmi k autorským právam. Každý z týchto zdrojov je problematický a na vytvorenie slovenskej verzie zbierky ELTeC použiteľný len v obmedzenej miere. Digitálne archívy knižníc disponujú väčšinou len skenmi bez alebo len s čiastočným (automatickým) optickým rozpoznaním znakov (OCR), čiže z jednotlivých stránok oskenovaných kníh je často potrebné extrahovať samotný text pomocou špecializovaných softwarov (Transkribus, FineReader). Verejné projekty, akým je napr. Zlatý fond denníka SME, obsahujú len selektívne diela, pochádzajúce z rôznych vydání (pričom často nejde o pôvodné texty, ale o texty v aktualizovanej jazykovej mutácii). Výsledný textový súbor rozsiahlejších próz je obyčajne výsledkom práce viacerých dobrovoľníkov a vyznačuje sa chybovosťou pri OCR. Tretia kategória sa zameriava prevažne na pirátske kópie populárnych literárnych diel a prekladov a týka sa len súčasnej slovenskej literatúry.

Prekvapením pri príprave ukrajinskej zbierky bolo, ako veľmi sa ukázali ako užitočné súkromné iniciatívy vytvárania elektronických knižníc.¹⁰ Z niektorých hľadísk nielenže nezaostávajú, ale dokonca predstihujú podobné zbierky pripravené veľkými inštitucionálnymi tímami národných knižníc. Je potrebné tiež poznamenať, že veľké množstvo publikácií ukrajinskej literatúry sa digitalizovalo v archívoch a knižniciach v Spojených štátoch amerických a Kanade.

8 <https://www.snk.sk/sk/zbierky-a-zdroje/zbierky-literarneho-archivu-snk.html>

9 <https://zlatyfond.sme.sk>

10 Napr. elektronické knižnice Chtyvo (<https://chtyvo.org.ua/help/biblioteka/pro-nas>) alebo Izbornyk (<http://litopys.org.ua/>). Je zaujímavé, že nájsť informácie o niektorých autoroch, ktorí iniciovali také rozsiahle (a svojim zámerom ušľachtilé) projekty, nie je ľahké. Anonymita môže súvisieť okrem iného so skutočnosťou, že publikácie nie vždy dodržiavajú autorsko-právnu legislatívu.

634 Neskôr sa veľa z nich objavilo na voľne prístupných stránkach, vrátane stránok pre ukrajinských používateľov.¹¹

Zostavovatelia čiastkových korpusov ELTeC, v tomto prípade tvorenom výberom sto diel z každej národnej literatúry v danom období (1840 – 1920), majú prístup k množstvu literárnych textov, či už v digitálnej alebo tlačenej podobe, ktoré je možné nazvať povinným čítaním. Tieto texty sa aktívne používajú vo výučbe a ich počet môže byť ešte väčší ako počet, ktorý vyžaduje zostavenie národného súboru pre ELTeC. To však ešte neznamená, že je možné vytvoriť zbierku jednoduchým preformátovaním povinného čítania podľa pravidiel ELTeC. Texty v zbierkach ELTeC musia spĺňať niekoľko kritérií, pričom kľúčovým sa ukazuje vyvážený pomer kanonických a nekanonických textov. „Povinné čítanie“ pozostáva z kanonických literárnych diel, a preto by zostavovatelia národných zbierok ELTeC mali vyhľadávať, skenovať, digitalizovať a kódovať aj množstvo ďalších nekanonických, menej známych textov.

Primárnym dôvodom je vyváženosť jednotlivých častí ELTeC, a to tak, aby tento základný výberový rámec bol reprezentatívnym celkom literatúry daného obdobia. Stanoviť pravidlá vyváženosti bolo pomerne ťažké, keďže jednotlivé európske literatúry nikdy neboli rovnako rozvinuté. Zatiaľ čo nemecká a francúzska literatúra udávali smerovanie vývinu literárneho diskurzu v Európe a zosobňovali to, čo v 19. storočí nazýval J. W. Goethe „svetovou literatúrou“,¹² menšie národy nedisponovali rovnako rozvinutou literárnou tradíciou. Pri tvorbe kritérií ELTeC bolo preto potrebné opierať sa o určitý teoretický model, teda o podobný základ, aký nachádzame vo vyváženosti všeobecných jazykových korpusov, ktoré poznáme z počítačovej lingvistiky. Tá rozlišuje medzi špecializovanými jazykovými korpusmi, akým môže byť napr. korpus medicínskych textov, a všeobecným korpusom, ktorý zastupuje jazyk ako celok. Hlavným účelom existencie všeobecných korpusov sú možnosti skúmania jazyka ako systému, t. j. opis jeho základných vlastností a poznávanie jeho reálneho fungovania. Dve podobné ambície nachádzame aj v digitálnych literárnych zbierkach. Ak chce nejaký literárny vedec študovať špecifiká poetického systému napr. v Shakespeareových hrách, využije na to špeciálny korpus, ktorý pozostáva zo Shakespeareových diel. Práve práca s ním, prípadne porovnávanie tohto korpusu s korpusom iného autora, mu umožní odhalíť, potvrdiť či vyvrátiť určité predpoklady, ku ktorým sa nie je možné dopracovať bežným čítaním.¹³ Pre takéto ciele sú zostavované a budované národné alebo špecializované digitálne literárne zbierky. Zámer vytvoriť viacjazyčnú európsku literárnu zbierku ako ELTeC však takéto parciálne využitie prekračuje, pretože zmyslom takejto zbierky je preskúmať systém literatúry podobným spôsobom,

11 K dispozícii je dokonca aj špeciálna elektronická knižnica obsahujúca skeny kníh publikovaných mimo Ukrajiny s názvom Diasporiana: <https://diasporiana.org.ua/>

12 K predstave svetovej literatúry, ako ju chápal J. W. Goethe v rozhovoroch s Eckermanom (1827) a o dvadsať rokov neskôr K. Marx a F. Engels v Manifeste komunistickej strany (1848), sa F. Moretti stavia kriticky. Opisuje ju ako súdobé intelektuálne gesto limitované priestorom západnej Európy, presnejšie „nemeckými filológmi pracujúcimi s francúzskou literatúrou“ (MORETTI, Franco: Conjectures on World Literature. In: *New Left Review*, 2000, č. 1, s. 54).

13 Diela W. Shakespeara sa týkal aj výskum Petra Plecháča z Ústavu českej literatúry AV ČR, ktorý zaznamenal široký medzinárodný ohlas (PLECHÁČ, Petr: Relative contributions of Shakespeare and Fletcher in Henry VIII: An analysis based on most frequent words and most frequent rhythmic patterns. In: *Digital Scholarship in Humanities*, fqaa032, <https://doi.org/10.1093/llc/fqaa032>).

ako všeobecné korpusy umožňujú skúmanie systému jazyka. Pre všeobecné textové korpusy platí, že pri ich tvorbe sa usilujeme o čo najväčšiu vyváženosť, preto okrem vyrovnaného zastúpenia jednotlivých štýlov (publicistický, umelecký, vedecký) sa snažíme do korpusu zaradiť čo najviac textov z rôznych tematických oblastí určitej činnosti alebo poznania, tzv. domén.¹⁴ Je zrejmé, že tieto kritériá nie je možné jednoducho aplikovať na literárne texty, preto je na mieste otázka, aké kritériá vyváženosti a reprezentatívnosti má teda ELTeC.

Hlavným kritériom je proporčné diachrónne zastúpenie diel vo fixne daných časových obdobiach podľa roku prvého vydania diela: 1840–1859, 1860–1879, 1880–1899, 1900–1920. Pod vyváženosťou korpusu (zbierky) máme na mysli, že nasledujúce parametre sa budeme snažiť dodržať pre každé časové obdobie, ako aj pre korpus ako celok, a to bez ohľadu na to, koľko textov máme k dispozícii. Môže sa stať, že za určité obdobie budeme mať nedostatok textov v niektorých kritériách, v inom zase bude potrebné urobiť užší výber, ktorý budú mať na starosti literárni historici orientovaní na dané obdobie. Zbierka v pilotnej verzii neobsahuje veľké množstvo textov, iba sto románov z každej literatúry, ktoré vyšli prvý raz v období rokov 1840–1920.¹⁵ Práve z týchto dôvodov sú pre každé obdobie kľúčové percentuálne podiely.

Základné kritériá vyváženosti zbierky ELTeC sú nasledovné:

Autorský podiel: ideálne 9 autorov – 1 autor s tromi textami/románmi v každom časovom úseku (ak to nie je možné, v záujme zachovania vyrovnanosti zaraďujeme do ELTeC len jeden text/román¹⁶ pre jedného autora).

Pohlavie autora: min. 10 % a max. 50 % by malo ísť o ženské autorky, pokiaľ je to vzhľadom na časové obdobie možné.

Dĺžka románov/textov: min. 20 % budú zastúpené krátke romány (rozsaň 10 000–50 000 slov), min. 20 % budú zastúpené dlhé romány (viac než 100 000 slov), zvyšných 60 % by mali tvoriť romány/texty štandardnej dĺžky, t. j. medzi 50 000 a 100 000 slovami.

Literárny kánon: min. 30 % vysoko kanonizovaných textov, ktoré sa vyznačujú opakovanými vydania, čiže ide o „povinné čítanie“ daného obdobia. Naopak, min. 30 % textov sú texty nezaraďované do kánonu, t. j. texty, ktoré vyšli v ďalších vydaniach len zriedkavo alebo neboli nikdy znova vytlačené. Orientačná báza pre posudzovanie ďalších vydání je, či text vyšiel opäť a koľkokrát v časovom období 1970–2009. Dôvodom pre toto kritérium je najmä to, aby sa do zbierky dostali takisto texty, ktorým nebola z hľadiska literárnych dejín určitej národnej literatúry

14 K tejto problematike viac v kapitole Bibliografická a štýlovo-žánrová anotácia dokumentov a textov v Slovenskom národnom korpusu. In: ŠIMKOVÁ, Mária – GAJDOŠOVÁ, Katarína – KMEŤOVÁ, Beáta – DEBNÁR, Marek: *Slovenský národný korpus: Texty, anotácie, vyhľadávania*. Bratislava : Jazykovedný ústav Ľ. Štúra SAV, Mikula, 2017.

15 Ďalšie romány, ktorých počet presahuje sto, sú zahrnuté v takzvaných „rozšírených zbierkach“. Aj keď vydanie takýchto rozšírených zbierok v rámci ELTeC nie je plánované v najbližšej budúcnosti (aby sa nevytvoril nepomer medzi jednotlivými zbierkami), sú ľahko dostupné a záujemcovia ich môžu použiť aj na iné účely (<https://github.com/COST-ELTeC>).

16 Pojem románu v rámci ELTeC je definovaný vágne, iba ako fiktívny prozaický text.

venovaná pozornosť, čo v určitom zmysle ospravedlňuje nedostatočnosť a zjednodušenie v prístupe ku kanonizácii textov.¹⁷

Domnievame sa, že naplnenie týchto kritérií na menšom počte textov umožní formulovať relevantné závery o širšej škále a obdobiach dejín národných literatúr. S určitým zjednodušením možno jeden román národnej literatúry považovať za „predstaviteľá“ niekoľkých ďalších v komparácii s ostatnými národnými „predstaviteľmi“. Silnou stránkou koncepcie projektu ELTeC v zmysle vyváženej je spojenie najvýznamnejších spisovateľov a druhoradých autorov, čím zbierka realizuje tézu Franca Morettiho o „veľkom nečítanom“ (*the great unread*).¹⁸ ELTeC navyše predstavuje nielen nekanonické texty v rámci konkrétnych európskych literatúr, ale aj „nekanonické literatúry“ ako celok, ktoré nie sú v západnom literárno-kritickom diskurze prakticky zastúpené.¹⁹ Na výstup projektu – získanie porovnateľných korpusov pre rôzne literatúry – sú potrebné univerzálne princípy výberu textu a rovnaký štandard anotácií. Texty v zbierke ELTeC sú preto jednotne anotované a kódované vo formáte XML štandardami TEI.²⁰ Výhodou je, že tieto anotácie robia konkrétni zástupcovia jednotlivých národných literatúr so svojimi tímami, ktorí sú odborníkmi na konkrétny literárny a kultúrny kontext. Môžu tak účinne a rýchlo vyriešiť zložité problémy, napríklad problémy spojené s historickými reáliami alebo so špeciálnymi jazykovými vlastnosťami.

Jednotné štandardy a nevyhnutný odstup od konkrétnych literatúr však okrem pozitív vedie nevyhnutne aj k určitým stratám: „Ak chceme porozumieť systému v jeho celistvosti, musíme sa zmieriť s tým, že niečo stratíme“,²¹ napísal F. Moretti v programovom článku o dištančnom čítaní ešte v roku 2000 a platí to aj dnes. To, čo strácame, sú charakteristické a jedinečné vlastnosti menšinových kontextov, napríklad pri malých národných literatúrach je to ich viacjazyčnosť. Pretože v zbierke ELTeC je možné prezentovať iba texty napísané v národných jazykoch, je zrejmé, že významné a často dôležité diela národnej literatúry napísané v inom jazyku sa vynechávajú. Na to dopláca najmä ukrajinská zbierka, a to z dôvodu, že ukrajinská literatúra obsahuje aj texty v ruštine, ale tiež v poľštine a nemčine, ktoré napísali nielen autori „druhého sledu“, ale aj autori považovaní za národných klasikov. Napríklad romány Tarasa Ševčenko napísané v ruskom jazyku by z ukrajinskej zbierky ELTeC boli vylúčené. Existuje tiež obmedzenie týkajúce sa miesta vydania publikácie, ktoré by malo byť v Európe. To znamená, že do zbierky sa nemôžu dostať romány vydané napríklad v Severnej Amerike, a to ani v príslušnom časovom období a v príslušných národných jazykoch.

Ďalší problém, ktorý sa ukazuje v našom geografickom priestore, je, že zbierka ELTeC uprednostňuje knižné verzie, t. j. samostatné dobové vydania, pred

17 Koncepcia kanonicity je spojená s počtom (opakovaných) publikácií diel za posledných niekoľko desaťročí. Je to zjavné zjednodušenie, pretože počet vydaní nemôže slúžiť ako spoľahlivý ukazovateľ kanonicity a dokonca ani popularity určitého diela.

18 F. Moretti si termín vypožičal od Margaret Cohenovej, ale pre ňu znamenal skôr „literárny archív“, neprečítané zdroje na štúdium literárnych diel (KHADEM, Amir: *Annexing the Unread: A Close Reading of „Distant Reading“*. In: *Neohelicon*, roč. 39, 2012, s. 410).

19 Viac k výkladom nekanonicity Khadem, c. d., s. 418.

20 Pokyny pre kódovanie: <https://distantreading.github.io/Schema/eltec-1.html>; <https://distantreading.github.io/Schema/eltec-2.html>

21 MORETTI, Franco: *Conjectures on World Literature*. In: *New Left Review*, 2000, č. 1, s. 54 – 68. Cit. podľa MORETTI, Franco: *Distant Reading*. London: Verso, 2013, s. 49.

seriálovými publikáciami v periodikách, kde často vychádzali dlhšie prózy na pokračovanie; vystupuje to ako problém v ukrajinskej i slovenskej zbierke. Takisto je žiaduce mať prvé vydanie románu, hoci táto podmienka patrí k tzv. mäkkým podmienkam – snaha držať sa prvého vydania značne zvyšuje pracovnú záťaž. Redakčná prax zostavovania zbierok zo slovanských literatúr obyčajne uprednostňovala texty neskorších vydání; spravidla sa vyberalo posledné vydanie za života autora, ktoré obsahuje jeho konečné zámery.²² Väčšina voľne dostupných ukrajinských a slovenských digitálnych textov preto predstavuje presne takúto verziu. To znamená, že zostavovatelia zbierok musia diela skontrolovať a porovnať ich s verziami prvých vydání, podľa ktorých je potrebné urobiť menšie, avšak niekedy významné zmeny. Nie je správne zazlievať projektu toto posledné obmedzenie, pretože ide o všeobecnejšiu otázku. Katherine Bode výstižne poznamenala, že dôsledné čítanie, a teda aj dištančné, odmieta skutočnosť, že literárne diela nie sú jediným a stabilným textom.²³ Výskumní pracovníci, rovnako ako čitatelia a vydavatelia, sa zvyčajne chcú zaoberať presne týmto jediným, klasickým, autentickým kanonickým textom, ktorý takmer nikdy neexistuje.

Aký je teda aktuálny stav zbierky ELTeC? Od novembra 2020 sa zverejňujú na platforme Zenodo jednotlivé národné zbierky, ktoré spĺňajú kritériá a obsahujú minimálne päťdesiat textov.²⁴ Pôvodne malo ísť o kolekciu prvých desiatich národných literatúr, existujú však dôvody domnievať sa, že ambiciózná úloha pripraviť viac ako desať reprezentatívnych čiastočných zbierok s polovičným počtom románov bude splnená do konca roka 2020 (projekt sa končí v novembri 2021). Aktuálny stav slovenskej a ukrajinskej zbierky ELTeC je taký, že slovenské aj ukrajinské zastúpenie v projekte COST sa usiluje napriek vyššie opísaným problémom s digitálnymi zdrojmi spracovať a pripraviť svoje národné verzie zbierok ELTeC. V prípade Slovenska prípravu výrazne sťažuje absencia kvalitných digitálnych zdrojov, teda profesionálne zostavenej fulltextovej zbierky slovenskej literatúry, ktorá by sa dala v tomto prípade využiť. Na jej vytvorenie vo verzii značne rozšírenej a prispôbenej slovenským špecifikám bol na jeseň 2020 v rámci výzvy Agentúry pre vedu a výskum podaný spoločný projekt troch pracovísk, dvoch ústavov SAV a jedného univerzitného pracoviska. Podpora projektu by výrazne napomohla nielen k skvalitneniu digitálneho výskumu slovenskej literatúry, ale aj k zapojeniu slovenských vedcov do európskych a širších medzinárodných výskumných infraštruktúr v rámci dištančného čítania a digitálnych humanitných vied.

Štúdia je výstupom grantového projektu MVTS SAV COST Action CA 16204: Distant Reading for European Literary History (1840–1920).
Zodpovedný riešiteľ: Mgr. Marek Debnár, PhD. Doba riešenia:
2019–2021

22 Princíp editio ultima (posledné autorské vydanie) dominoval nielen vo vedeckej redakcii ukrajinskej literatúry, ale vo väčšine slovanských literatúr (YESYPENKO, Dmytro: Printed and Digital Wor(l)ds: Retrospectives and Perspectives of Scholarly Editing in Slavic Countries. In: *Canadian Slavonic Papers*, roč. 62, 2020, č. 1, s. 50, 60).

23 BODE, Katherine: Abstraction, Singularity, Textuality: The Equivalence of „Close“ and „Distant“ Reading. In: BODE, Katherine: *A World of Fiction: Digital Collections and the Future of Literary History*. Ann Arbor: University of Michigan Press, 2018, s. 33.

24 <https://distantreading.github.io/ELTeC/>

638 Literatúra

- ARAC, Jonathan: Anglo-globalism? In: *New Left Review*, 2002, č. 16, s. 35 – 45.
- BODE, Katherine: Abstraction, Singularity, Textuality: The Equivalence of „Close“ and „Distant“ Reading. In: BODE, Katherine: *A World of Fiction: Digital Collections and the Future of Literary History*. Ann Arbor : University of Michigan Press, 2018, s. 16 – 35.
- COHEN, Margaret: Narratology in the Archive of Literature. In: *Representations*, zv. 108, 2009, č. 1, s. 51 – 75.
- DEBNÁR, Marek – GOGORA, Andrej: *Digitálne trendy v súčasných humanitných vedách*. Nitra : UKF, 2019.
- GOGORA, Andrej: The text mining of ethics and information technology. In: *Communications. Scientific Letters of the University of Žilina*, roč. 19, 2017, č. 1, s. 3 – 8.
- JUROVÁ, Jarmila – DEBNÁR, Marek: Interpretation of Values in the Epitaphs on the Tombstones of Old Jewish Cemetery in Bratislava, 2018. In: *Communications. Scientific Letters of the University of Žilina*, roč. 20, 2018, č. 1A, s. 71 – 79.
- KHADEM, Amir: Annexing the Unread: A Close Reading of „Distant Reading“. In: *Neohelicon*, roč. 39, 2012, s. 409 – 421.
- MORETTI, Franco: Conjectures on World Literature. In: *New Left Review*, 2000, č. 1, s. 54 – 68.
- MORETTI, Franco: *Distant Reading*. London : Verso, 2013.
- MORETTI, Franco: *Grafy, mapy, stromy – Abstraktní modely literární historie*. Prel. Olga Čaplyginová. Praha : Karolinum, 2014.
- MORETTI, Franco: *Graphs, Maps, Trees: Abstract Models for a Literary History*. London : Verso. 2005.
- NEWMAN, Mark: The Structure and Function of Complex Networks. In: *SIAM Review*, roč. 45, 2003, č. 2, s. 167 – 256. Dostupné na: <http://epubs.siam.org/doi/pdf/10.1137/S003614450342480>
- ŠIMKOVÁ, Mária – GAJDOŠOVÁ, Katarína – KMEŤOVÁ, Beáta – DEBNÁR, Marek: *Slovenský národný korpus: Texty, anotácie, vyhľadávania*. Bratislava : Jazykovedný ústav L. Štúra, Mikula, 2017.
- YESYPENKO, Dmytro: Printed and Digital Wor(l)ds: Retrospectives and Perspectives of Scholarly Editing in Slavic Countries. In: *Canadian Slavonic Papers*, roč. 62, 2020, č. 1, s. 45 – 66.
- ZUMRÍK, Miroslav: *Filozofické aspekty korpusovej lingvistiky*. Bratislava : Veda, 2020.

Elektronické zdroje

PLECHÁČ, Petr: Relative contributions of Shakespeare and Fletcher in Henry VIII: An analysis based on most frequent words and most frequent rhythmic patterns. In: *Digital Scholarship in Humanities*, fqaa032, DOI: <https://doi.org/10.1093/llc/fqaa032>

Mgr. Marek Debnár, PhD.
Ústav slovenskej literatúry SAV
Dúbravská cesta 9
841 04 Bratislava
Slovenská republika
E-mail: Marek.Debnar@savba.sk

Dr. Dmytro Yesypenko
Taras Shevchenko Institute of
Literature
National Academy of Sciences
of Ukraine
4 Mykhaila Hrushevskoho Street
Kyiv 01 001
Ukraine
E-mail: dm.yesypenko@gmail.com