

## ROMANIAN LITERATURE IN THE DIGITAL AGE

Luiza Marinescu

Assoc. Prof., PhD, Spiru Haret University of Bucharest

*Abstract:* There are blessed ages in which computer scientists and humanities scholars are supposed to cooperate in order to develop and share Knowledge and methodologies in a new common research area: the digital humanities. In Romania, among the academic personalities, Solomon Marcus (1st of March 1925, Bacău - 17th of March 2016, Bucharest) was perhaps the only one who happily combined linguistics and mathematics, being known in the Romanian philological environment especially for *Lingvistica matematică/ Mathematical Linguistic, Modele matematice în lingvistică/ Mathematical Models in Linguistics* 1963, 1966 și *Poetica matematică/ Mathematical Poetic* 1970 and in the international scientific field for numerous translation and also for numerous papers published in prestigious publishing houses in Paris, New York and Bologna.

Grounded in the Distant reading paradigm (i.e. using digital humanities in analyzing large collection of literary texts) the paper, *From close and to distant reading of Romanian novels from 1850 to 1920* provides a historical classification of texts in order to be included in a multilingual European Literary text Collection (ELTeC), permitting to test methods and compare results across national traditions and consider the consequences of such resources and methods for rethinking fundamental concepts in literary history. The paper is structured in the following chapters:

1. *Close and Distant Reading in Digital Humanities: Romanian literature case from 1850 to 1920.*
2. *Romanian Literary History and Ways of Transnational Knowledge through Distant Reading of Literary Facts: 100 Romanian Novels from 1850 to 1920*
3. *Transnational Redefinition of Periods in National History Through a Literary Species: the Novel*
4. *The Problem of "Translation / Export" of National Literary Histories*

*Keywords:* Romanian literary history, mathematical linguistics, digital humanities, Solomon Marcus, Franco Moretti

### 1. De la lectura clasică la cititul la distanță și domeniul de studii Digital Humanities: cazul literaturii române între 1850-1920

Se spune că o carte adevărată nu este cea pe care o citim, ci **cartea care își citește cititorul**. În lumea contemporană acestui obicei al lecturii i s-au asociat termeni internaționali precum *close reading*, *hyper reading*, *social reading*, *distant reading*, *historical contextualization*, *surface reading*. Și asta pentru că s-au diversificat mediile în care citim literatura în epoca digitală, fie ele cartea tipărită, e-reader, tabletă, telefon, computer sau smartphone, acest fapt generând puncte tari și puncte slabe ale diferitelor strategii de lectură. O metodă de studiu literar în lumea occidentală rămâne lectura supranumită *close reading*. În 1920 criticul literar I. A. Richards a făcut un experiment cu studenții săi, pe care i-a pus să interpreteze poezii, fără a le oferi nici un fel de informații despre autor, despre contextul publicării, despre curentul literar al perioadei, despre titlu poemului. Studenții au eșuat în acest experiment, iar I. A. Richards și-a dat seama că această abilitate de a interpreta un text literar, fără nici o informație contextuală, este foarte utilă și a numit-o *close reading*, încercând să o cerceteze și să o dezvolte izolat de toate celelalte abilități de lectură. Această abilitate sau metodă specifică presupune,

întocmai precum contextualizarea istorică, un loc de frunte în munca literaților, fiind interconectată cu metoda de contextualizare istorică. În cartea sa *Practical Criticism* I. A. Richards îi invită pe cititori să acorde atenția supremă cuvântului tipărit pe pagina respectivă și „să facă dreptate” în acest fel textului însuși.

Cleanth Brooks, Robert Penn Warren, Allan Tate și John Crowe Ransom, adică noua critică (New Critics) care a debutat prin anii 1930 și care a dominat viața literară cam până prin anii 1950-1960, au considerat *close reading* o formă particulară, foarte precisă de identificare a tuturor subtilităților semantice și sintactice ale textului literar, izolând-o și cultivând-o separat. Pentru ei, textele literare nu au nici un fel de funcție socială sau psihologică, și după părerea lor, ele nici nu educă și nici nu fac din lume un loc mai bun. Ele își poartă propria valoare literară care, la un moment dat, pierde întocmai precum o ființă umană. Comparația cu ființa umană e corectă în contextul în care ființa umană trăiește izolat, eventual într-un mediu experimental gen laborator. Pe cititori, în opinia acestor critici nu ar trebui să îi privească nici opinia autorului, nici biografia sa, nici contextul social, politic și istoric. Cei care citesc îndeaproape textul (Close readers) trebuie să ignore deliberat toate contextele istorice, sociale, politice și biografice pentru a se concentra doar asupra textului. Close readers lucrează doar pe textul literar, prin acordarea de maximă atenție a cuvintelor din pagină. Close reading este o formă foarte precisă de lectură, ce caută subtilitățile formelor literare, suporturile și structurile ce construiesc un roman, un text dramatic sau unul liric.

În epoca contemporană această Nouă critică/ New Critics e considerată formalistă, anistorică și apolitică de către Noii istorici / New Historicists, care discută în direct opoziție cu generația anterioară numită Noii critici/ New Critics și sunt de părere că această metodă nu mai este una crucială. Close reading are meritul de a lăsa cititorul să se concentreze asupra cuvintelor sau a frazei și să exploreze sensurile. Uneori însă această analiză microscopică a cuvintelor și a sensurilor e inutilă și complicată, forțată și fără legătură cu restul textului. Alteori, cel care realizează analiza, care e o disecție a textului, este cunoscător și al contextelor, și al detaliilor biografice, dar aceste aspecte sunt eludate de analiza finală, vitregindu-i înțelesul. Cititorul de tip close reader mărește detaliile textului, pe care îl supune unui zoom și „nu vede pădurea de copaci”. El vede copacul la microscop. Firește că textul contează pentru cititor și că în totdeauna cititorul trebuie să se întoarcă la text.

Literatura română a evoluat în alte condiții și, în loc de această etapă de teoretizare a lecturii, a avut parte de dezbateri legate de modernitatea literaturii în privința scrierii istoriei literare. Eugen Lovinescu și George Călinescu sunt cei care trasează coordonatele modernității istoriei literare românești, decalând cu mai multe decenii discuția integrării tipurilor de lectură.

În succesiunea close reading, când informatica a întâlnit literatura, în Europa occidentală s-au dezvoltat mai multe strategii de bază referitoare la lectura zilnică a textelor on-line. O formă de cooperare a lecturii on-line se numește *social reading*, un fel de lectură colectivă, diferită de cea făcută la colțul străduțelor medievale, când *Divina Commedia* își câștiga notorietatea. Formele de lectură istorică ale textelor literare și contextualizarea istorico-literară a constituit o etapă absolut necesară în evoluția procesului de lectură. Utilizările și limitările oferite de conceptul de distant reading formulat cu eleganță intelectuală de Franco Moretti a generat noi idei legate de utilizarea hărților literare pentru „*Atlasul romanului european 1800-1900*” și a oferit o soluție academică de abordare a textelor literare, care rezultă din analizarea de către computer a unor mari cantități de date. Aceste noi abordări academice ale lecturii nu se concentrează asupra interpretării datelor, ci asupra determinărilor de suprafață, referitoare la materialitate.

Această deghizare umanistică a informaticii și a matematicii a pornit de la articolul lui Noam Chomsky, *Three models for the description of language*, "I.R.E. Transactions on Information Theory", vol. IT-2, no. 3: 113-24 (1956) și de la cartea acestuia, *Syntactic Structures*. The Hague: Mouton, 1957; Berlin and New York: 1985, care detaliază cele trei modele matematice de descriere lingvistică. Ideile acestor lucrări s-au dezvoltat diferit în occident și în Europa răsăriteană.

O relatare edificatoare despre evoluția alianței dintre științele umaniste și matematică/informatică în România este oferită de academicianul Solomon Marcus (*Singurătatea matematicianului*, 2010), care menționează că și școala rusă propusese o alianță între lingvistică și matematică. „N. Kolmogorov propune un model algebric al cazului gramatical, V. A. Uspenski publica un model algebric al părții de vorbire, iar R. L. Dobrushin propune un model algebric al categoriei gramaticale. Primele experimente de traducere automată, începute încă în anii '40, au un caracter predominant ingineresc, dar în 1958 O. S. Kulagina extrage din acest tip de activitate o descriere a noțiunilor de bază ale gramaticii pe baza teoriei mulțimilor. Tatonarea posibilităților de traducere automată și de documentare automată în Europa occidentală, în cadrul Euratom, și în S.U.A., de exemplu, prin David Hays, conduce, spre sfârșitul anilor '50 și începutul anilor '60, la diverse idei de proiectivitate sintactică (Yves Lecerf și alții), o provocare interesantă pentru teoria grafurilor. În toate aceste activități sunt implicate esențial logica matematică (gramatica generativă a lui Chomsky este, în esență, un sistem formal în sensul lui Hilbert) și unele capitole de combinatorică (sistemele lui Post și probleme de tipul celor propuse la începutul secolului trecut de Axel Thue). Tot din direcție logico-matematică provin ideile lingvistice și logice ale lui Y. Bar-Hillel (1953) și J. Lambek (1958). F. Harary și N. Paper propun în 1957 un calcul al distribuției fonemelor, N. Chomsky prezintă în 1958 o analiza a relației dintre lingvistică, logică, psihologie și calculatoare; în același an, Y. Bar-Hillel analizează procedurile de decizie în limbile naturale. M. Masterman discută în 1957 relația dintre semantică și sintaxă în traducerea automată. La toate acestea trebuie să adăugăm articolul lui S. C. Kleene din 1956, privind reprezentarea evenimentelor în rețele nervoase și în automate finite, în ordinea de idei inaugurată de articolul din 1943 al lui W. S. McCulloch și E. Pitts, asupra unui calcul logic al ideilor implicate în activitatea nervoasă.” (Marcus, 2010:13, 14)

## 2. Istoria literară românească și modalitățile de cunoaștere transnațională prin intermediul lecturii la distanță: 100 de romane românești din perioada 1840-1920

### a. Rafinarea conceptului de lectură a romanelor cu ajutorul mijloacelor tehnologice diverse: tipuri de lectură și de media

Critica literară din spațiul anglo-saxon și american a rafinat teoretizarea conceptului de citire/lectură sub denumirea de *reading*. Romanele au rămas un gen de lectură iubită, căci și în zilele noastre cărțile tipărite, computerul, laptopul, tabletele, e-readerul, telefonul mobil, I pad sau Iphone-ul toate acestea medii tehnologice de lectură schimbă modalitatea de receptare, și implicit, de interpretare a textului romanelor.

În vechime, se spunea că zeul egiptean Thot al lunii, al aritmeticii, al vorbirii și inventatorul scrisului dăruise oamenilor această din urmă invenție ca un suport menmotehnic pentru lectură, pentru a-i face pe oameni să rețină ceea ce citesc. În ultimele decenii în spațiul european occidental s-a pus foarte mult în discuție:

1. În ce măsură aceste medii tehnologice, textele digitale influențează procesul lecturii și al înțelegerii textelor?

2. Dacă cititorul a realizat lectura unui text în medii diferite, înseamnă că el a efectuat lectura unor texte diferite?
3. Ce diferențe există între reacțiile/răspunsurile diferite ale creierului uman în cazul lecturii textului pe hârtie și în cazul lecturii cuvintelor pe ecran? Există reguli de prevenire a erorilor?

*Concluzia* este că ecranul afectează înțelegerea textului, diminuând capacitatea creierului de a-și aminti ceea ce a citit pe ecran și influențând claritatea redării informației parcurse. (Carr, 2010; Birkerts, 1994; Car, 2013; Hayles, 2012) Prin viteza de scriere utilizând funcțiile copy-paste CITOITORUL decalează capacitatea creierului omenesc de a înțelege ceea ce citește și de a scrie în cuvinte proprii ceea ce a înțeles. Funcțiile copy și paste măresc timpul de scriere și diminuează timpul de procesare în creier a informației deja parcurse. Lectorul on-line, care șterge barierele dintre texte, dintre cititor și autor este un căpcăun efemer de informație neprocesată, pe care nu are timp să o transforme în substanță nutritivă pentru evoluția lui spirituală. Informația din textul literar selectată rămâne ciuntită, ruptă, aleator combinată sau juxtapusă într-un alt text on-line ori printat și imposibil de povestit cronologic. Pentru literați și pentru oamenii de știință lectura în medii tehnologice variate, spre deosebire de lectura clasică a textului tipărit prezintă avantajul funcției de căutare, selectare și copiere. De asemenea este vorba și despre diseminarea materialului selectat, prelucrat. Ar mai fi o singură condiție: timpul de procesare umană a informației prezentate să nu se diminueze, iar corectarea rezultatului să se facă și clasic, pe hârtie.

Capacitatea de a se cufunda în lectură a cititorului este mărită în cazul textelor electronice dispuse pe dispozitive tip e-reader sau e-book pentru că acestea nu oferă posibilitatea de a naviga on-line, întrerupând lectura liniară, pentru a învăța mai multe despre autor, despre locurile descrise în roman, despre informații referitoare la personajele din roman, care sunt pasajele cele mai citite sau cele mai comentate.

În privința softurilor, foarte mulți cititori preferă textele care se pot rearanja (*reflowable texts*), adică textele fără un număr fix de pagini, care pot fi citite cu diverse mărimi de fonturi. Cititorii profesioniști, literații, pe de o parte, preferă textele cu număr fix de pagini, astfel ca acestea să poată fi citate în publicațiile academice.

#### **b. De la lectura de tip close reading spre lectura la distanță (distant reading) în domeniul științelor umaniste pe suport digital (digital humanities): cazul literaturii române din perioada 1850-1920**

În România, acest moment al joncțiunii științelor umaniste cu matematica și informatica este descris de către academicianul Solomon Marcus (*Singurătatea matematicianului*, 2010):

„Să rezumăm. Dezvoltări din direcții foarte diferite fac joncțiunea în a doua parte a anilor '50, aducând într-o albie comună discipline dintre cele mai diverse: lingvistica, psihologia (Chomsky considera lingvistica generativă un capitol al psihologiei cognitive), calculatoarele, matematica, logica și biologia; dar prin teoria informației, în plin elan atunci, s-a făcut legătura și cu fizica, în special cu termodinamica. Inutil să mai adaugăm că filozofia se afla în fața unor provocări fără precedent. Evenimentele enumerate aveau loc într-un moment în care se năștea informatica în România, sub bagheta extraordinarului dirijor de energii creatoare care a fost Grigore C. Moisil. Și pentru că, vorba poetului, toate aceste lucruri *trebuiau să poarte un nume*, s-au inventat diverse etichete, una dintre ele fiind lingvistica matematică. Sintactic, nu putem alătura decât doi termeni; dar era clar că noile preocupări nu combinau numai două domenii, ci

mai multe. Marea noutate consta în faptul că se aflau împreună cel puțin șase discipline, dintre care trei din domeniul socio-uman. Se mai aflau împreună știința și ingineria. Polaritatea pascaliană «spiritul de geometrie, spiritul de finețe» și contrastul dintre cele două culturi, la care se referea C. P. Snow, primeau o provocare fără precedent. Ne aflam în plină transdisciplinaritate. Cum puteam rămâne indiferent la aceste evoluții? Am intrat în joc. Am simțit că unor energii care așteptau de mult să se dezlănțuie le-a venit ceasul. Într-un timp record, m-am inițiat în lingvistica structurală, disciplina prin care te apropii de noile preocupări din direcția lingvisticii. Am fost ajutat în aceasta privință de discuțiile cu Emanuel Vasiliu, cel mai apropiat de logică și de matematică, dintre lingviștii români ai aceluși moment, și de Paula Diaconescu, entuziasta cercetătoare în analiza structurală a limbii române; amândoi, de la Catedra de limba română a Universității din București, catedra condusă de Profesorul Alexandru Rosetti. Lor, li s-au adăugat ulterior Edmond Nicolau și Sorin Stati. Între Rosetti și Moisil a existat o atracție magnetică, ei au încurajat și sprijinit o colaborare față de care cei mai mulți se arătau sceptici. Sprijinul lor, la Universitate și la Academie, a permis României să fie una dintre primele țări în care s-au ținut cursuri universitare de lingvistică matematică și computațională și în care s-a înființat o revistă de profil, în limbi internaționale.

### ***Un loz câștigător***

Drept rezultat, a urmat o dezvoltare vertiginoasă, oglindită parțial în recentul volum *Grigore C. Moisil and his followers in theoretical computer science* (Ed. Academiei Române, 2007). În această atmosferă, am redactat cursul de lingvistică matematică pe care Editura didactică și pedagogică mi l-a publicat în 1963, cu rezerva considerată normală față de o întreprindere aparent hazardată. Entuziasmul mă împiedica să sesizez caracterul aparent utopic al traseului pe care mă angajam. Cursul se baza în bună măsură pe cercetările mele personale, publicate în reviste. Pentru a mă testa, am trimis cartea la câteva adrese universitare potențial interesate într-o atare aventură. A fost un loz câștigător. A urmat publicarea ei la New York, la Paris, la Moscova și la Praga. Marile enciclopedii Brockhaus, Encyclopaedia Universalis, Enciclopedia Einaudi, Great soviet Encyclopedia, Encyclopedia of Mathematics și numeroase enciclopedii de lingvistică, de cibernetică, de informatică au menționat una sau alta dintre edițiile cărții. Trăiam astfel o experiență nouă, nu mai rămâneam cantonat într-un domeniu cu granițe destul de precise, ci mă aflam pe un traseu transdisciplinar, care mă obliga să învăț nu numai lingvistica, ci și biologia sistemului nervos, biologia eredității, logica, psihologie cognitivă, structura limbajelor de programare și anumite capitole de matematică discretă, care nu erau pe linia antrenamentului meu anterior, din studiul funcțiilor reale și al topologiei generale. Am simțit tot timpul, în această nouă etapă, sprijinul Profesorilor Rosetti, Moisil și Miron Nicolescu. Atunci am descoperit faptul că, prin interacțiune cu disciplinele socio-umane, matematica și calculatoarele dobândesc, pentru un public destul de larg, o valoare culturală.” (Marcus, 2010: 15-17)

În România adevărata digitizare a cataloagelor din biblioteci începe după 1990 (Biblioteca Academiei Române), iar scanarea primelor cărți începe să se producă în momentul în care computerele personale/PC-urile încep să facă parte din munca zilnică a literaților, cu aproximație la începutul primului mileniu. Primele biblioteci digitizate (colecții ale bibliotecilor municipale –digibuc, colecții ale bibliotecilor județene și universitare- BCU, BCU *Lucian Blaga*) încep să fie funcționale. Generalizarea folosirii internetului prin instalarea rețelelor corespunzătoare se face cu aproximație în 2010, când România e considerată una dintre țările în care viteza de navigare pe rețelele de internet este una maximă. Accesibilitatea dispozitivelor/

media pentru lectura on-line se produce cam tot din 2010, când, în fiecare an, generațiile de dispozitive se perfecționează și își fac apariția cititorii on-line de romane.

În Europa occidentală cele mai citite cărți în 2009 erau romanele. Dar strategiile de lectură clasică denumite fie *close reading*, fie *deep reading* ale romanelor tipărite pe hârtie s-au schimbat. A început să se vorbească despre *hiper lectură (hyper reading)*, adică lectura oricărui text dispus on-line (cu adresă electronică www, adică world wide web), descărcarea, accesarea hiperlinkurilor, folosirea operațiilor de tăiere și copiere (cut și paste) cu ajutorul computerelor, o tehnică de lectură care se petrece numai în fața ecranului, care are menirea să ne transforme capacitatea de a citi în general și care a început să joace un rol important. Deși hyper reading ne ajută să intrăm în legătură cu tot felul de oameni și de informații de pretutindeni și să schimbăm păreri, idei, impresii, fiind un mijloc accesibil de obținere a informațiilor și a cunoștințelor, totuși, acest tip de lectură distrage atenția și afectează negativ abilitatea noastră de a înțelege textele. Uneori induce stresul și alteori alterează rețeaua neuronală, în sensul că nu dispunem de o atenție continuă, care să ne permită înțelegerea instantanee a textelor parcurse. Hyper reading este o formă de lectură on-line, non-lineară de lectură, adică o formă nestructurată de lectură, în care hiperlinkurile îl pot conduce pe cititor oriunde în loc să aibă loc parcurgerea de la pagina 1 la pagina 2.

Solomon Marcus compara literatura cu matematica, în privința schimbării paradigmei lecturii de către hipertext discutând despre *dependența de contexte lungi* și subliniind diferențele, cu mult timp înaintea acestor schimbări ale media de lectură:

„Fenomenele de textualitate, de intertextualitate și de *hipertextualitate*, în linia de gândire a unor M Bakhtin, J. Kristeva și a celor care, prin hipertextualitate, au transgresat secvențialitatea textului tradițional, sunt la ele acasă în matematică. Într-adevăr, într-un text matematic se manifestă, mai mult decât în orice alt text, fenomenele de dependență la distanță. Suprimați dintr-o carte de matematică primele zece pagini și riscați să nu mai înțelegeți aproape nimic din rest. O operație similară într-o carte de geografie sau de istorie are un efect neglijabil. Faptul se explică prin structura textelor matematice; prin construcția în etape, în care fiecare etapă se bazează în mod riguros și explicit pe etapele anterioare. Desigur, în orice demers procedăm în etape care se folosesc de etapele precedente, dar de cele mai multe ori acest lucru se face prin reamintirea faptelor anterioare care urmează a fi utilizate. În matematică, preluarea noțiunilor, convențiilor și rezultatelor anterioare are o asemenea amploare, încât reluarea lor, de fiecare dată când ele sunt invocate, ar pune la grea încercare atenția și memoria și ar sabota funcția euristică a limbajului. Achizițiile etapelor anterioare trebuie ordonate cu grijă, așa cum se procedează într-o locuință, prin gruparea diferitelor obiecte în dulapuri, sertare, cutii diferite. În matematică, această ordonare impune folosirea unei anumite terminologii și a unui anumit simbolism, prin care desemnăm noile noțiuni și entități, în vederea folosirii lor cât mai comode în etapele următoare. Astfel emerge componenta artificială a limbajului matematic. Sub aspect istoric, acest fenomen s-a accentuat pe vremea lui Galilei și a lui Newton, accelerându-se apoi și atingând apogeul în secolul trecut.

### ***La fel în poezie, dar din cu totul alt motiv***

Să precizăm că dependența de contexte mari, practic, de întregul text, are loc în ambele direcții, deci atât la stânga cât și la dreapta. Așa cum un element al textului depinde strict, chiar dacă indirect, de întreaga desfășurare anterioară a textului respectiv, același element va fi invocat, direct sau indirect, în întreaga desfășurare ulterioară a textului. Limbajul matematic este deci, prin excelență, un teritoriu de desfășurare permanentă a relațiilor anaforice și cataforice. Este interesant faptul că, și în poezie, localul este solidar cu globalul, se vorbește chiar despre

modul în care o serie de metafore locale se acumulează, producând o metaforă globală. Dar această dependență nu are, în poezie, caracterul precis și explicit pe care îl are în matematică. Legătura dintre local și global este, în poezie, o operație ambiguă, interpretabilă într-o infinitate de feluri; ea ține deci de actul lecturii și al interpretării, aparține cititorului. Semnificațiile în matematică au un statut conceptual, iar conceptele sunt susceptibile de definiții. Acest fapt le distinge de semnificațiile poetice, care manifestă o tendință anticonceptuală. Poezia încearcă să recupereze cuajutorul contextului ceea ce pierde în materie de dicționar. De aceea ea are nevoie de contexte practic infinite, regăsind astfel, pe o cale complet diferită, o situație valabilă și în matematică.” (Marcus, 2010: 32-34)

În eseu *Hyper-Readers and Their Reading Engines (Hipercititorii și motoarele lor de citit)* James J. Sosnoski definește hyper reading în legătură cu „practicile de lectură asistate de computer, bazate pe necesitatea existenței unui ecran” (Sosnoski, 1999: 167) și identifică opt operații adiacente acestei metode: *filtrarea (filtering)*, *smântânirea/frunzărirea (skimming)*, *ciugulirea/scobirea/săparea cu târnăcopul (pecking)*, *filmarea (filming)*, *încălcarea legilor sau hotarelor textului (trespassing)*, *anularea autorității (de-authorizing)* și *fragmentarea (fragmenting)*. Hiper lectorul (hyper reader) e mai selectiv, supunând textul *filtrării (filtering)* cu textele citite pe ecran decât cu cele tipărite (Sosnoski, 1999: 163), pentru că are la îndemână motoarele de căutare (Sosnoski, 1999: 166). *Smântânirea sau frunzărirea (skimming)* presupune că din cuprinsul unui articol, se poate selecta doar ceea ce prezintă interes pentru a îl ajuta pe cititorul care are multe de citit (Sosnoski, 1999: 167). *Ciugulirea/scobirea/săpătura cu târnăcopul (pecking)* presupune că lectorul alege câte puțin din toate, fără a urmări logica textelor din care selectează diverse fragmente (Sosnoski, 1999: 168). *Impunerea (imposing)* este a patra strategie descrisă de Sosnoski care se referă la faptul că după lectura hyper textului (gen articol Wikipedia) cititorul se întreabă dacă ceea ce a citit e adevărat sau e o minciună, pentru că hyper textele au mai puțină coerență, unitate și autoritate decât are un roman sau un poem. Cititorul unui hyper text e mai interesat să îl mineze, pentru a își impune punctul de vedere în text și pentru a-l utiliza în propriul său interes, impunându-i propriul sens. (Sosnoski, 1999: 168) *Filmarea (filming)* este strategia de lectură care se referă la raportul elementele grafice și elementele verbale ale textului, raport care favorizează imaginea în hyper texte. (Sosnoski, 1999: 169) Cu alte cuvinte cititorul „filmează” imaginile, adică defavorizează lectura textului. Hipercititorii sunt considerați librocлаști, adică spărgătorii de cărți/texte, pentru că ei atribuie hipertextele cu nenumărați autori domeniului public. Orice fel de link pe care un site îl creează către un alt site este un act de anulare a autorității, pentru că acesta îl utilizează în scopuri proprii. (Sosnoski, 1999: 169) *Încălcarea legilor sau hotarelor textului (trespassing)* (Sosnoski, 1999: 163) este una dintre greșelile frecvente ale hiper lectorului /hyper reader, care trebuie să respecte rigorile restabilirii frontierelor textului pentru a evita plagiatul. (Sosnoski, 1999: 170) *Privarea de autoritate* a hyper textului are loc în momentul în care hiper cititorul îl concediază pe autor, asumându-și prin asamblarea și reasamblarea *fragmentelor* textului în structura/ forma proprie a textului pentru care semnează ca autor. (Sosnoski, 1999: 170, 171). În privința ultimelor trei strategii de lectură (*încălcarea hotarelor hyper-textului, privarea de autoritate și fragmentarea*), dispăre pur și simplu granița dintre cititor și autor, deoarece fragmentarea oferă libertatea logică, topologică și asociativă a fragmentelor. (Sosnoski, 1999: 172)

Celor opt strategii de hiperlectură, (hyper reading sau hyperreading), Catherine Hayles, autoarea studiului *Cum gândim: digital media și tehnogeneza contemporană (How We Think: Digital Media and Contemporary Technogenesis)* publicat în 2012 le adaugă strategia de

*juxtapunere* (adică operația de comparare a textelor, prin două sau mai multe ferestre deschise pe același ecran) și cea de *scanare* (adică operația de culegere a tuturor imaginilor care au legătură cu subiectul de interes).

Întrebarea fundamentală care rezultă din studiul lui Hayles era cum să transformi în abilități de lectură toate aceste teoretizări strategiilor lecturii în epoca digitală și cum să construiești o punte între lectura textului pe hârtie și lectura textului în format digitizat. (Hayles, 2012: 56)

Lectura socială (social reading) e o formă de colaborare în procesul de lectură on-line care implică discuții care augmentează timpul de lectură și înțelegerea și care se transformă într-o experiență de lectură colectivă. Proiectul citirii romanului *The Golden Notebook* de Doris Lessing bazat pe comunicare asincronă este un exemplu de social reading. Lectura socială este formală în opoziție cu mica discuție despre lectură, participarea putând fi sincronă sau asincronă (lectorii citesc când pot, simultan sau nu același text). Lectura socială este persistentă în opoziție cu o conversație oarecare. Ea creează un anumit mediu de lectură dar nu îl exclude pe celălalt, adică lectura adâncă, clasică.

*Lectura la distanță (distant reading)* este o formulare recent evaluată de către studiile italianului Franco Moretti, în opoziție cu lectura apropiată de intimitatea textului (close reading). În vreme ce cititorul clasic sau close reader se concentrează pe o singură particularitate a textului literar, cititorul de la distanță (distant reader) poate analiza sute sau mii de texte literare pentru a identifica tipare la scară istorică generală, matrice care nu țin seama de granițele istorice sau de limitările temporale. Cititorii de la distanță/Distant readers folosesc metode din științele sociale și naturale. Rezultatul acestor studii este un grafic, o hartă sau un arbore. Nu e pentru prima dată în istoria culturală când geometria apare în studiile umaniste. Cărțile lui Franco Moretti, profesorul italian de la Universitatea Standfort au generat o neliniște referitoare la renunțarea la cititorul clasic și la close reading și la studiile aferente, grație interesului pentru combinarea matematicii și a geografiei cu studiile literare, urmând ca în urma analizei computaționale, pe o bază de date, să se identifice tipare recurente, evoluții și involuții la scară istorică amplă a unor specii literare precum romanul. Fie că este vorba despre valul romanului epistolar, valul romanului gotic sau despre valul romanului istoric, hărțile sale ilustrează distribuția protagoniștilor în diverse medii sociale. Harta distribuției protagoniștilor în Paris se concentrează asupra axei de simetrie reprezentată de Sena, unde bărbații protagoniști ai romanului trăiesc în lumea artistică a malului stâng, în vreme ce femeile protagoniste ale romanelor, trăiesc în partea de nord a Senei, în zona comercială și în zona aristocratică Saint Germain. Prin intermediul arborilor, Moretti redă dezvoltarea stilului indirect. Studiile realizate de Franco Moretti reprezintă o analiză cantitativă a informațiilor din romanele europene din perioada 1800-2000. Principalele argumente ale autorului legate de utilizarea lecturii la distanță/distant reading sunt:

1. În loc să se concentreze asupra studierii semantice și sintactice a unui singur autor, critica are la dispoziție posibilitatea de a scana o mare cantitate de texte, de a le transforma în baze de date pentru studierea tiparelor evolutive și involutive recurente la nivelul diverselor specii literare preferate de cititori, precum romanul.
2. Lărgirea și redeschiderea canonului literar prin analizarea literaturii pierdute sau uitate (literatura scrisă de femei, literatura scrisă de persoanele de culoare) în canonul care includea doar nume de excepție precum Shakespeare, Goethe etc. toți autorii fiind „bărbați albi”.
3. O mai mare obiectivitate în analiza literaturii, dincolo de normele, valorile și prejudecățile fiecărui critic literar.

Noutatea acestei strategii profesionale de lectură (distant reading) propusă de Franco Moretti constă în asocierea programelor de computer în studiile literare și a bazelor de date, pentru analizarea unor anumite teme de cercetare.

Solomon Marcus explica interferența matematicii/informaticii în arhitectura spiritului uman oglindită în științele umaniste de-a lungul timpului:

„Mai întâi, urmărind firul dezvoltării matematicii la vechii greci, constatăm caracterul predominant spiritual al ei, vocația contemplării unor armonii de forme și arhetipuri. Inventarea teoremei este o achiziție spirituală care, numai ea singură, ar fi suficientă pentru a asigura prestigiul peste milenii al culturii vechilor greci. La Pitagora, matematica și muzica sunt inseparabile, amândouă raportate deopotrivă la cosmos și la arhitectura spiritului uman. Numerele, intervalele muzicale și mișcarea corpurilor cerești conduc la ceea ce s-a numit *muzica sferelor*. Cele cinci tipuri de poliedre regulate, puse în evidență de Platon, sunt entități la fel de fundamentale ca dreapta, cercul, pătratul și sfera la Euclid, și fac parte din viziunea lui Platon asupra matematicii ca reprezentare a universului. Le găsim în mituri, în diferitele religii, în simbolismul artelor și în rezultatele fundamentale ale științei. Numărul prim, șirul lui Fibonacci, proporția de aur, ideile de grup, de mulțime ordonată, de spațiu topologic, banda lui Möbius, sticla lui Klein, noțiunea de infinit mic, la Leibniz, și universul non-standard al lui Robinson rezumă structuri, prototipuri și procese sau comportamente cu valoare universală. De aceea pot apărea deopotrivă în natură și în cultură, în știință și în artă, în natura inertă și în cea vie. Pentru cultura vechilor greci, Platon reprezintă cea mai înaltă expresie a matematicii ca aspect fundamental al spiritului uman. Pentru Aristotel, discipolul lui Platon, matematica nu este o parte a științei și nu este subordonată acesteia; matematica se ocupă de obiecte al căror interes este de sine stătător și care admit o motivare estetică.

#### ***Spiritualitatea matematicii: secolele XIII-XVII***

Sf. Augustin (354-430) preluase de la Platon fascinația pentru numere, iar de la Euclid metoda de procedare axiomatic-deductivă. Această metodă avea să fie urmată de teologia catolică până spre secolul al XVII-lea. Dar nu numai teologia, ci și alte discipline au urmat aceeași cale; a se vedea *Etica* lui Spinoza (1632-1677) și mecanica newtoniană. Duns Scotus (secolul al XIII-lea) se ocupă de problema existenței și infinității lui Dumnezeu, folosind procedee care prefigurează noțiuni din ceea ce azi numim teoria mulțimilor ordonate. Nicolaus Cusanus (1401-1464) vede în matematică unicul mod de a ajunge la certitudine. Ca și Descartes, mai târziu, Cusanus adoptă ipoteza unui Univers indefinit (nu infinit). N. Copernic (1473-1543) propune, în lucrarea sa privind mișcările de revoluție ale sferelor cerești (1540), un model matematic al heliocentrismului. Opera lui Copernic a rămas în primul rând pentru valoarea ei științifică, dar și calitatea ei literară este remarcabilă; este un poem dedicat Soarelui și Cercului.

În perioada Renașterii (secolul al XV-lea) are loc o alianță fericită între artele vizuale și matematică, prin nume ca Leonardo da Vinci, Brunelleschi, Alberti, Albrecht Dürer, Piero della Francesca și Bombelli. Se realizează astfel un progres substanțial în înțelegerea perspectivei (reprezentarea spațiului cu trei dimensiuni în cel cu două dimensiuni).

Galileo Galilei (1564-1642), prin *Il Saggiatore*, *Sidereus Nuncius* și mai cu seamă prin opera sa *Dialog*, se înscrie în istorie drept unul dintre părinții științei moderne, prin recunoașterea rolului central al matematicii în înțelegerea lumii. Dar, după cum au atras atenția Leopardi și Italo Calvino, prin aceleași opere Galilei rămâne și ca unul dintre marii scriitori în proză ai Italiei. Un alt savant dublat de un scriitor este Johannes Kepler (1571-1630), care în *Astronomia Nova* (1609) face apel la cele cinci tipuri de poliedre ale lui Platon pentru a studia interacțiunea dintre om și cosmos. Kepler demonstrează că traiectoriile planetelor nu sunt circulare, cum se

credea, ci eliptice; sunt astfel aduse în atenție secțiunile conice ale lui Apollonius de Perga (262-180). Pasul următor: Isaac Newton (1642-1727) descoperă legea atracției universale (1687).

René Descartes (1586-1650) preconizează o știință unificată, având ca model matematica. Asemenea lui Galilei, Descartes crede că matematica este cheia care deschide drumul spre o imagine globală, unificată și coerentă a lumii. Plecând de la matematică, Descartes s-a simțit proiectat în fizică, filozofie, psihologie, fiziologie și cosmologie, în toate acestea devenind un pionier. În *Discurs asupra metodei*, Descartes strălucește nu numai prin deducție filozofică, ci și prin aspectul literar.” (Marcus, 2010: 54, 55, 56)

La fel ca și lectura la distanță, *lectura suprafeței (surface reading)* reprezintă o nouă strategie profesională de lectură, care se concentrează nu pe interpretarea textului literar, ci pe materialitatea cărților, pe materialul din care sunt confecționate. Pentru lectorii suprafețelor, faptul că citim o carte tipărită sau un e-book reprezintă o mare diferență. Tot o mare diferență este și în privința lecturii unei piese de Shakespeare într-o ediție bibliofilă, în ediția princeps învelită în piele, într-o ediție ieftină din secolul XXI, pe un hard sau tipărită. Materialitatea cărții contează foarte mult pentru acest tip de lectură. Lectorii suprafețelor pretind ca materialul din care este făcută cartea posedă așa numitele calități tactile (*haptic qualities*), care complinesc cunoașterea conținutului informațional. Cititul cu mâna examinează estetic calitatea cărții din mâna cititorului. Acesta din urmă apreciază „calitățile tactile și trăsăturile legate de mobilitate”. Studiul atingerii este privit ca un „mod de comunicare”. (*Reading by Hand-the haptic evaluation of artists' books*. <https://www.researchgate.net/publication/26410992> Reading by Hand the haptic evaluation of artists%27 books [accesat pe 8 mai 2018]).

### 100 de romane românești din perioada 1850-1920 în viziunea lecturii la distanță

Orice bază de date rămâne accesibilă sau nu, în funcție de tipurile de media și de programe utilizate pentru constituirea sa. Într-o lume în care punțile de comunicare în cazul programelor informatice se taie la fiecare 10 ani, problema durabilității și a accesibilității datelor de către generațiile următoare este una dintre cele mai importante. Acțiunea COST referitoare la lectura la distanță în istoria literară europeană, CA16204 DISTANTREADING FOR EUROPEAN LITERARY HISTORY privește proiectul selectării a 100 de romane din fiecare literatură europeană din perioada 1850-1920 și a propus o grupare de idei care să orienteze selecția pe [https://distantreading.github.io/sampling\\_proposal.html](https://distantreading.github.io/sampling_proposal.html)

#### a. Criteriile selecției

Cum fiecare literatură are particularitățile sale, propriile medii de publicare, propriile texte referitoare la romanul realist, romanul popular, principala provocare e legată de criteriile de comparare, adică elementele care se regăsesc sau nu în toate literaturile comparate.

De asemenea este important de subliniat că fiecare literatură a făcut cunoștință cu această specie literară (romanul) la o anumită dată când înțelesul noțiunii de autor era diferit. Proporția dintre autorii identificați în perioada respectivă între 1850 și 1920 (bărbați și femei, cel puțin 20%) ar putea duce la lărgirea canonului. Proiectul este centrat pe identificarea elementelor care sunt comune literaturilor din perioada amintită și a diferențelor dintre ele, dacă ceva diferit trebuie subliniat sau dacă ceva trebuie eliminat ori adăugat.

#### b. Digitizarea și sursele

Se preconizează selecția textelor din metadata, bibliografia bazată pe descrierile bibliografice standard, selecția autorilor, crearea unor liste cu fișiere disponibile pentru fiecare literatură pornind de la „inventarele” academice confirmate (*Dicționarul romanului românesc*), selectarea din aceste liste a elementelor legate de popularitate care să recomande romanele

respective pentru lista celor 100 de romane și digitizarea acestor materiale în decurs de doi ani, în format pdf.

Problema criteriilor comune de realizare a selecției a evidențiat faptul că limba de publicare și mai ales locul publicării operei literare schimbă apartenența lucrării respective. (Cazul literaturii Bucurei Dumbravă, scriitoare de limbă germană și română care a făcut cunoscută România prin subiectele românești ale romanelor sale în limba germană a scos în evidență faptul că romanele sale nu sunt incluse în nici un canon literar.)

Există perioade în care anumiți autori domină prin publicarea unui număr mai mare de romane, punându-și amprenta asupra epocii sau asupra operelor altor autori. Este importantă relevanța materialelor digitizate. Balanța dintre canonicitate și non-canonicitate trebuie echilibrată în vederea rescrierii canonului.

Edițiile traduse în limbi străine ale aceluiași roman sunt o probă de popularitate. În același sens, edițiile reeditate succesiv în aceeași limbă constituie o probă a popularității. În funcție de obiectivele cercetării se poate urmări:

1. Selectarea romanelor în care un anumit scriitor este menționat în opera respectivă, permite realizarea unei constelații de personaje, de autori.
2. Selectarea romanelor în care personajele mănâncă, citesc etc.
3. Selectarea romanelor în care complexitatea stilului generează cercetări morfo-sintactice și stilistice: numărul silabelor din cuvintele utilizate, numărul cuvintelor din frază; tipurile de metafore; tipurile naratologice (Gerard Genette)
4. Selectarea romanelor care vorbesc despre români: cum se văd românii, cum îi văd alții pe români.
5. Selectarea romanelor care privesc istoria cărții și a cititului: cititorii ca personaje; formulele de adresare referitoare la bărbatul scriitor sau femeia scriitoare; idiolectele cititorilor ca personaje sub influența lecturilor. (Vezi Dinu Păturică și biblioteca clasică din care selectează fragmentele referitoare la parvenire și parvenitism.)

**c. Exemplele de criterii pentru selecția ELTeC sunt:**

- *Sarcina de lucru:* se referă la liniile directe pentru schema de corpus, schemele de adnotări de bază și metadata și fluxul de lucru. Această preocupare aflată în lucru se concentrează pe designul corpusului și pe metadata, deoarece ambele sarcini se intersectează reciproc. Pentru crearea unui astfel de corpus de referință, avem nevoie de un design corporal care să permită comparabilitatea textelor și a sub-colecțiilor individuale în funcție de seturile diferite de metadata. Ar trebui să fie posibil ca fiecare membru al acțiunii COST să eșantioneze sub-colecții din ELTeC pentru sarcini specifice și întrebări de cercetare. Într-un prim pas, ne concentrăm asupra dezvoltării unor criterii de selecție clare, operaționalizate, transparente și motivate pentru corpus.

Este important să subliniem faptul că nu se intenționează definirea romanului ci a tipurilor de criterii de selecție folosite pentru ELTeC. Scopul acțiunii COST CA16204 este de a crea un corpus mare de literatură de referință din perioada (1840) 1850-1920 (prima perioadă) pentru diferite metode computaționale de citire la distanță pentru adnotarea și analiza corpusului.

- *Metodă:* În primul rând, discuția pornește de la întrebarea dacă sarcina de a construi corpusul necesită criterii de selecție normative (canon) sau criterii metodice de selecție (eșantionare corporală). Prima diferență importantă dintre abordări este motivația criteriilor. Un canon este portretul prestigiosului statut

social, cultural, economic și reflectă decizii normative de auto-promovare și de rating. Dimpotrivă, designul corporal urmează o întrebare sau un context de cercetare și, prin urmare, este orientat mai mult spre obiectivul cercetării. Un al doilea aspect important este modul de examinare a textelor reale. Moisil (2009: 876) afirmă că „datele sunt ontologic diferite de lume”. Deci există o diferență între textele din lume și datele pe care le creăm. Prin texte, putem lua în considerare manifestarea sau extinderea sau lucrarea unui text (conform IFLA 2009). Un canon poate conține o extensie a unui anumit text care este disponibil în diferite limbi și tipărite. Din punct de vedere ontologic, aceste nivele diferite de text diferă de ceea ce ar putea fi un text dintr-un corpus (cf. van Zundert și Andrews 2017). Aceasta înseamnă că digitizarea este un fel de adnotare, deci interpretare (Odebrecht et al 2017). O reprezentare a unui text într-un corpus (de exemplu, transcriere, OCR) este rezultatul interpretării. Un proiect de corpus trebuie să ia în considerare acest lucru pentru problemele de eșantionare și digitizare. La sfârșitul acțiunii, ELTeC ar trebui să conțină texte literare (romane) dintr-o perioadă distinctă și în mai multe limbi. Pentru fiecare limbă (și astfel pentru fiecare context cultural) există o diversitate de canoane, care reflectă diferite grupuri de prestigiu, cum ar fi editori, autori și cititori. Acești actori și posibila lor influență asupra construirii canonice sunt multi-relaționale. Fiecare canon este rezultatul textelor din diferite perspective. Evaluarea poate reflecta evaluarea intelectuală (un text este un reprezentant al unei anumite perioade de literatură, este influențat, este important), ratingul economic (un text este publicat în mai multe ediții tipărite) sau evaluarea cititorilor (un text este cel mai popular în cadrul un anumit grup de cititori) (Hermann 2011 sau Winko 1996). Toate aceste evaluări se pot schimba în timp și pot, de asemenea, să intervină între ele. Un canon poate, prin urmare, să reflecte diferite interpretări ale textelor „celebre”, „importante” sau „influențe”. Aceste criterii nu sunt în general comparabile. De exemplu, textele dintr-o comunitate de limbi mai mici, cum ar fi româna, sunt mai puțin probabil să fie reproduse frecvent decât textele în limba engleză din aceeași perioadă și gen. În plus, gradul de conștientizare a textelor depinde de influența țării și a editurilor lor.

Criteriile derivate dintr-un canon nu sunt complet comparabile și categorice. Ce canon prestigios al grupului ar trebui să fie luat în considerare, care ar trebui exclus și de ce? Există canoane comparabile pentru romane în toate țările limbii în cauză? Algee-Hewitt și McGurl (2015) arată o abordare a designului corporal bazat pe mai multe canoane și care sunt tipurile de probleme care apar. Fiecare analiză a corpusului arată doar efectele diferite ale deciziilor luate de grupul normativ. Considerarea canoanelor naționale este, de asemenea, foarte dificilă și oarecum problematică. Un exemplu este Canonul național românesc de literatură dezvoltat în secolul al XVIII-lea și promovat de sistemul educațional național până în 1989. Din 1989, sistemul educațional românesc nu a promovat un canon, ci a recomandat o listă formată din numele a 10 autori pentru examenul de bacalaureat, fără a recomanda o listă de cărți de citit în școală sau în universitate. România nu este singură în acest demers, Germania reunificată urmând același drum. Astfel, luarea unor astfel de canoane ca parte a unei baze de eșantionare a ELTeC ar însemna reflectarea unui trecut politic și social al educației și politicii din țările respective. Alegerea între canoane poate însemna alegerea între gusturile literaturii (actuale) (în trecut și

prezent) și gusturile din literatura anterioară, atunci când autorul canonului evaluează textele istorice. În cele din urmă, aceste canoane nu sunt construite pentru a fi linii directoare de eșantionare pentru un corpus pe care ne-ar plăcea să îl construim în acțiunea COST., „Scopul principal și obiectivul acțiunii COST CA 16204 este ca această acțiune să dezvolte resursele și metodele necesare pentru a schimba modul în care este scrisă istoria literară europeană”. Acest obiectiv necesită o nouă abordare a designului corporal, a modelelor de metadata și a modelelor de adnotări. După cum afirmă Fowler (2002, 214): „Actualul canon stabilește limitele înțelegerii noastre în literatură în mai multe moduri”. Bazarea pe canoane va împiedica obiectivul acțiunii într-un mod fundamental. Canoanele oferă acces tradițional și normativ la istoria literaturii. În schimb, acțiunea se concentrează pe noi abordări pentru a spune o altă poveste. În schimb, am putea decide dacă colecția noastră ar trebui să conțină un amestec de opere care nu au fost reproduse niciodată de la prima lor apariție, lucrări care au fost reproduse de câteva ori în decursul unuia sau două decenii de la prima lor apariție și lucrări reproduse în aproape fiecare deceniu de la prima lor apariție.

Prin urmare, se susține o abordare non-normativă, dar bazată pe metadata, a criteriilor de eșantionare, care vor urma o abordare de proiectare corporală. Criteriile de eșantionare ale Corpus-ului sunt orientate / dezvoltate în cea mai mare parte de întrebarea și / sau contextele de cercetare ale grupului de creatori de corpuri. În acțiunea CA16204, nu avem nici o întrebare de cercetare distinctă, nici un grup fix și anterior cunoscut de creatori de corpuri-uri. Contextul de cercetare al acțiunii este mai mult interesat de producerea de cunoștințe în sens metodologic și nu preferă o singură metodă, model sau teorie. În plus, grupul membru al acțiunii va fluctua și va consta din cercetări din diferite discipline cu diferite contexte teoretice și culturale. Astfel, intenția este să se construiască structura corporală pe o bază metodică. Prin această metodă se va putea, de asemenea, să se selecteze textele canonice, dar nu numai acestea.

*Reprezentativitatea:* În plus față de aspectul „prestigiului” (canonul), aspectul reprezentativității este problematic pentru proiectarea corpusului. Dezvoltarea unor criterii pentru proiectarea corpusului înseamnă a determina ce tip de eșantion din lume ce va fi inclus în baza de date. Evident, includerea întregii literaturi a populațiilor din secolul al XIX-lea în mai multe limbi este imposibilă. Așadar, trebuie să se facă un compromis între ceea ce dorim să avem în corpus (toată literatura) și ceea ce putem pune în corpus (eșantioane). Este adevărat că nu există un corpus „bun” sau „rău”, deoarece modul în care un corpus este proiectat depinde de tipul de corpus și de modul în care acesta va fi utilizat. (Hunston 2008, 155) Un proiect de corpus trebuie să urmeze obiectivul cercetării. Pentru acțiune, reprezentativitatea poate fi relația dintre corpus și corpul literaturii în cauză. „Reprezentativitatea se referă la măsura în care un eșantion cuprinde întreaga gamă de variabilitate într-o populație” (Biber 1993, 243). A spune ceva despre reprezentativitatea unui corpus necesită cunoștințe despre întreaga literatură a populație (în perioada în cauză). De fapt, nu se cunoaște fiecare carte a fiecărei limbi publicate / citite / discutate în Europa în perioada în cauză. În plus, este imposibil să se identifice o listă completă a „categoriilor” care ar explica exhaustiv toate textele produse într-o anumită limbă (Hunston 2008, 161) sau context. Astfel de categorii se referă la diferiți factori, cum ar fi caracteristicile autorilor, de exemplu, genul sau locul nașterii, editorii, subiectele textelor, cititorii etc. În contextul construcției canonice, nu există nici o măsură reală a „semnificației” a unui tip de discurs către o comunitate. Șansa ca un corpus să reprezinte întreaga populație dintr-un spațiu crește odată cu mărimea unui corpus. În acest fel, mărimea și reprezentativitatea sunt conectate. (Vezi Biber (1993) și Hunston (2008) pentru o discuție detaliată.)

Reprezentativitatea este, prin urmare, un fel de ideal, pe care ne-ar plăcea să-l urmăm, dar care nu poate fi atins ca întreg. În conformitate cu memorandumul de înțelegere, ELTeC poate fi proiectat ca un corpus model de monitorizare, în care textele (din diferite limbi și perioade) să poată fi adăugate în timp.

Un corpus complet echilibrat este un ideal pe care se încearcă să fie realizat alături de idealul de reprezentativitate (Hunston 2008, 163). Conform Memorandumului de înțelegere, corpusul va conține 2500 de romane cu texte complete, cel puțin în 10 limbi diferite:

a. Limbi: engleză, franceză, germană, greacă, italiană, poloneză, portugheză, rusă, spaniolă (core ELTeC)

b. prima tranșă: 6 sub colecții (100 de romane pe limbă) 1850-1920 începând cu limbile britanică, franceză, spaniolă, germană, greacă, poloneză

c. a doua tranșă: 4 sub colecții (100 de romane pe limbă) 1850-1920

d. a treia tranșă: 6 sub colecții în limbi suplimentare și sub colecții pentru toate cele 16 limbi 1780-1850,

e. Se va încerca de asemenea, includerea altor romane din alte literaturi în limbi, cum ar fi limba maghiara, sârba, elvețiana, româna, ceha, letona, norvegiana.

În acest fel, ELTeC este echilibrat în ceea ce privește limbile literaturilor pe care intenționează să le includă. În ceea ce privește genul, corpusul nu este echilibrat, ci omogen; toate textele din corpus vor fi romane complete. În ceea ce privește timpul, proiectul corpus-ului se va concentra pe perioada 1850-1920.

**Întrebarea este dacă se dorește să utilizăm fiecare criteriu, cu intenția de a reprezenta varietatea valorilor posibile sau ar trebui ca eșantionul să reprezinte distribuția acestor valori a eșantioanelor?**

Să presupunem că dorim să selectăm 100 de texte dintr-o populație de texte publicate pe o perioadă de (de exemplu) 20 de decenii. Am putea selecta cinci texte din prima decadă, cinci din cel de-a doua, și așa mai departe, formând cele 100 de titluri, împărțite uniform în decursul posibilelor decenii. Probabilitatea ca un text din corpusul nostru să apară într-o anumită decadă va fi întotdeauna aceeași: 1 din 5. Această selecție reprezintă varietatea valorilor posibile pentru criteriul respectiv. Să presupunem acum că ne uităm mai îndeaproape la numărul de titluri din fiecare deceniu disponibile efectiv în colecția pe care o prelevăm. Este mai mult decât probabil că acest număr va varia semnificativ: de exemplu, am putea observa că există 2000 de titluri publicate în deceniul x și numai 100 în decada y. Pentru a reprezenta statistic această populație, ar trebui să o facem de 20 de ori mai probabil ca un titlu ales aleatoriu să vină din decada x decât din decada y. Deoarece numărul total de titluri pe care le putem alege este destul de mic în raport cu numărul total disponibil în realitate, aplicarea strictă a acestui principiu ar putea însemna că nu putem alege deloc titluri din câteva decenii. Acesta este motivul pentru care preferăm să facem eșantionarea să reprezinte mai degrabă varietate decât frecvență; alt motiv este că nu putem alege un număr fracționat de titluri. Când începem să analizăm mai mult de un criteriu, sarcina de a ne asigura că numerele din eșantionul nostru reflectă cu acuratețe distribuția tuturor valorilor din întreaga colecție devine complexă.

Urmând abordarea reprezentării varietății unei colecții, trebuie să decidem care dintre criterii este echilibrat în ce fel și care interferează cu alte criterii. De exemplu, este posibil să dorim să alegem în mod echilibrat romane de la autori și bărbați. Acest lucru poate însemna că în totalul tuturor romanelor o jumătate va fi de la autori de sex feminin. Fără nicio altă

reglementare, am putea avea mai mulți autori de sex feminin într-un deceniu decât în alte decenii. Dacă dorim să avem un număr egal de autori de sex masculin și de sex feminin în fiecare deceniu, trebuie să conectăm criteriul genului autorului cu criteriul timpului. Făcând acest lucru, ar putea complica procesul de selecție (adică găsirea de romane pentru această proporție în fiecare deceniu al perioadei în cauză). Deci, trebuie să se decidem care categorii trebuie să fie prezente într-un mod echilibrat în corpus.

**d. Cerințe privind criteriile de eșantionare:**

Conform proiectului, designul corpusului ar trebui să fie echilibrat în ceea ce privește limba și data publicării textelor. Aceasta înseamnă că acest corpus nu trebuie să se bazeze exclusiv pe criterii cronologice, ceea ce înseamnă că e nevoie de un text din fiecare an al perioadei în cauză. Principalul criteriu de eșantionare „limba” nu va necesita includerea traducerilor. Se ia în considerare prima ediție a unui roman sau alte ediții ale acestor romane. Roman, înseamnă ediția cărții, excluzând astfel romane doar tipărite în reviste. Dacă luăm în considerare edițiile unui roman, aceste ediții ar trebui să fie disponibile în mod liber (licențe gratuite pentru reutilizarea acestora.) Prima ediție este cea mai interesantă din punct de vedere filologic, deoarece reprezintă textele autentice ale autorilor, îngrijite și scrise conform normei literare. Se vor combina toate formele de cuvinte, care sunt separate prin linii oblice. În momentul de față, trebuie să presupunem că nu există instrumente de normare (suficient de bune) pentru fiecare limbă. Edițiile ulterioare ale unui roman pot fi deja normate. Acest lucru ar putea duce la reprezentări diferite de text în ELTeC, care ar trebui indicate în metadata. Considerând, de asemenea, edițiile mai târziu disponibile ale unui roman au două avantaje: În primul rând, membrii acțiunii COST pot deja să ofere documente text (html, TEI etc.) editate ulterior și, în unele limbi, ar putea fi mai ușor să găsească mai târziu ediții care există deja într-un format care poate fi citit de mașină (în acest fel nu trebuie să depunem eforturi în digitizarea acestora).

Disponibilitatea electronică nu ar trebui să fie un criteriu de eșantionare principal. Un text nu trebuie exclus din ELTeC deoarece nu este digitizat, dar trebuie exclus, dacă textul nu poate fi disponibil în mod liber în ELTeC. Dacă utilizăm doar disponibilitatea ca criteriu de selecție, suntem expuși riscului de a copia proiecte precum "Gutenberg", de exemplu. Apoi avem nevoie de criterii suplimentare, care pot fi aplicate fără să fie nevoie să cunoașteți (citiți) textele în cauză. Criteriile trebuie verificate fără o cunoaștere aprofundată a textelor. În caz contrar, acest lucru se va opune obiectivului întregii acțiuni și abordării metodice a lecturii îndepărtate. Criteriile trebuie să fie operaționabile, ceea ce înseamnă că pot fi decise din metadatale de text. Aici, se definesc metadatale de text într-un domeniu mai larg decât doar metadatale bibliografice clasice. În acest fel, designul corpus interacționează cu metadatale. Unele metadata ale textului pot fi folosite drept criterii de eșantionare. Aceste criterii sunt criterii text-externe și -interne (cf. Hunston 2008) pe care trebuie să ne bazăm pe atunci.

Sugestia este să se utilizeze un tabel online ca mijloc de colectare a nominalizărilor pentru includerea în ELTeC, dar alte metode sunt fezabile.

**Criterii de eșantionare**

**Nivelul 1 Criteriul eligibilitate:**

Pentru a fi inclus, un text trebuie:

- Să fi fost publicat pentru prima oară ca o carte între 1850 și 1920, sub-grupat și împărțit în mod egal într-o perioadă de zeci de ani.
  - Să fi fost publicat pentru prima dată într-o țară europeană. [fără să fie chiar „primul” din acea decadă]

- Să fie un roman, adică o narațiune de proză fictivă cu o lungime minimă de 10.000 de cuvinte.

- Să fi fost inițial scrise în limba subcolecției date

### **Criteriul compoziției sau nivel 2.**

Printre romanele din fiecare limbă, subcolecția trebuie să conțină:

- cel puțin 10% -50% au fost scrise de autori de sex feminin pentru subcolecția limbii.

- cel puțin 10 autori sunt reprezentați cu trei romane.

- cel puțin 30% sunt romane foarte canonizate, cel puțin 30% ar trebui să fie romane non-canonizate, bazate pe retipărirea a două grupe:

1 Grup: canon. retipărite de mai multe ori

2. Grupa: non-canon. nu sau o dată retipărită, perioada: 1980-2000

cel puțin 20% sunt romane scurte ( 10-50k cuvinte-token), cel puțin 20% sunt romane lungi (> 200k cuvânt-token).

### **3. Redefinirea transnațională a periodizărilor din istoria literară națională prin intermediul unei specii literare: romanul**

- Data: 1850 - 1920  
(prima etapă)

- Se va împărți în decenii

- grupa A (1850-1860)

- grupa B (1861-1870)

- grupa C (1871-1880)

- grupa D (1881-1890)

- grupa E (1891-1900)

- grupa F (1901-1910)

- grupa G (1911-1920)

- Întrebare deschisă: Ar trebui să extindem perioada până în 1840?

#### **Metadate pentru textele din ELTeC**

Limba textului

MoA definește limbile care urmează să fie eșantionate. Nu propune diferențierea variației regionale (de exemplu, în limba germană), nici a variațiilor geografice (de exemplu, francezii din Belgia sau Franța sau Elveția). Aceasta presupune numai tipuri de limbi europene, astfel încât engleza exclude engleza americană; franceza exclude Quebecois.

Întrebare deschisă: Poate câteva excepții: - Locul publicării este uneori o problemă: de ex. pentru romanele portugheze publicate în publicații braziliene sau autori români, croați publicați în Germania.

Limbi posibile pentru prima iterație: olandeză, engleză, franceză, germană, greacă, italiană, poloneză, portugheză, rusă, spaniolă, maghiară

Numărătoarea edițiilor

Propunem să utilizăm de câte ori este reprodusă o lucrare, ca o măsură obiectivă a recepției, folosind categorii cum ar fi:

scăzut: retipărită de mai puțin de 10 ori

mare: reprodus de mai mult de 100 de ori

mediu: retipărită de 10 până la 100 de ori

Sexul autorului

- Utilizăm următoarele trei categorii pentru genul de autor actual (nu este revendicat)

Masculin

Feminin

mixt (nedefinit sau mai mult de un autor)

- Lungime:

Ar trebui să încercăm să includem o varietate de lungimi

scurt (mai puțin de 5000 de cuvinte)

mediu (5000 - 20000 de cuvinte)

lung (mai mult de 20000 de cuvinte)

- Felul romanului: Este o problemă deschisă cum se clasifică romanele tematic. Nu e clar cum să poate folosi acest criteriu de eșantionare.

Tema principală

Tabelul următor sugerează numărul minim și maxim de titluri care trebuie selectate pentru fiecare criteriu. De exemplu, pentru fiecare limbă, s-a propus selectarea între 80 și 120 de titluri diferite. Dintre acestea, nu vor reveni mai mult de 20 de la fiecare perioadă dată definită mai sus. Dintre acestea, se va încerca să existe între 5 și 10 exemple din fiecare dintre categoriile de reprint definite (mare, mediu, scăzut). Dintre acestea, numărul produs de autori va fi cuprins între 5 și 8, precum și numărul de autori de sex feminin și numărul emis de autori de sex necunoscut sau nedeterminat.

Limba Data Reprint Autor 80 - 120 20 5-10 5-8

Metadata pentru textele din ELTeC

Afișăm aici elementele de metadata care trebuie colectate pentru fiecare text. Acestea vor fi furnizate de componente specifice ale structurii TEI Header, ca definind în hârtie WG privind Encoding.

Vezi propunerea de Encoding pentru mai multe detalii:

titlu	a treia persoana
subtitlu	povestitorul autorului
data publicării	canonicitate / recepție
loc de publicare	evaluat de un canon: y / n
editor	canon
serie	Cuvinte cheie
editor	de la OPAC? alții? Proprii?
autor	limba
Nume	posibil: zona de limbă, tipul de limbă (de
sex	ex. germană, bavariană)
limba maternă	Referința sursei (dacă este disponibilă)
locul nasterii	de exemplu. DTA, textgrid etc.
introducerea unei baze de date a unei	
persoane, cum ar fi GND, dacă este	
disponibilă	
mărimea	
în cuvinte / în jetoane	
pagini	
subiecte (vezi Falkentheorie)	
politică, criminalitate, spionaj, provincial,	
școală, aventură, război, credință, intern,	
natură, fabrica	
subgen	
lucrare în curs de dezvoltare	
Narator	
prima persoana	

### Cazuri de primă utilizare a corpusului:

1. intertextualitatea: cum sunt menționate lucrările literare
2. constelații de caracter: măsuri de rețea
3. MoU afirmă „atribuirea de autor, modelarea temelor, analiza rețelei de caractere, analiza stilistică”
4. Ce tipuri de locuri / nume de locuri sunt descrise în diferite romane
5. proprietăți structurale, organizare formală
6. descrierea caracterelor; tipuri de caractere; cititorii ca personaje
7. metodele de atribuire a stilului de autor stilometric funcționează în toate limbile în același mod
8. Ce mănâncă oamenii
9. locație, locuri unde participă povești. morfo-sintaxă, lizibilitate; de exemplu. cuvinte / propoziții
10. despre metafore
11. trăsăturile naratologice vorbesc direct
12. prezentarea altor culturi europene
13. analiza sentimentului
14. aspecte bibliometrice: istoricul cărților
15. dezvoltare diacronică; locul publicării; gen
16. adresarea formularelor
17. Evoluția și validarea mișcărilor literare; definiție empirică
18. expresii temporale
19. toate cele de mai sus într-o perspectivă comparativă variind de ex. bărbat / femeie, data, locul
20. subgenuri?
21. rata și măsurabilitatea schimbării; modul în care se răspândește inovația; atât în subiect, cât și în modul de exprimare
22. Idiolecte de caractere; locuri
23. evoluția cuvintelor în timp
24. coerența locală; coeziunea; evoluția de la realism la modernism.

#### 4. Concluzii: problema translării & exportării romanelor din istoriile literare naționale

În cazul unui proiect de tip distant reading, posibilitatea ca literatura română să se afle pe harta literaturii europene într-o anumită perioadă este condiționată de efortul conjugat al mai multor actori în timp: criticul/ istoricul literar, bibliograful, bibliologul și echipa de informaticieni din cadrul bibliotecii (naționale, județene, academice, metropolitane), finanțatorul (Primăria, universitatea, Academia, consiile județene), echipa de cercetare a proiectului.

În cazul învățământului universitar actual din România lipsa a unui domeniu de studiu denumit generic *Digital humanities* este un element, care prin utilizarea sa ar putea schimba coordonatele de percepție și difuzare a literaturii române în lume.

Bibliografia primară pentru acest domeniu formată din inventare de corpusuri pe specii literare și perioade (*Dicționarul cronologic al romanului românesc* proiect realizat de Institutul de lingvistică și istorie literară „Sextil Pușcariu” al Academiei Române), sau inventare de lucrări traduse în limbile de circulație europeană, pe epoci istorice (Bibliografia titlurilor traduse, Bibliografia traducerilor din literatura română, *Bibliografia relațiilor literaturii române cu literaturile străine în periodice* coordonată de profesorul Dan Grigorescu și realizată la Institutul de critic și Istorie Literară George Călinescu din București) reprezintă primul pas în acest demers.

Împachetarea informatică a conținuturilor literare sub formă de corpus reprezintă provocarea referitoare la accesarea acestora, iar aceste riscuri trebuie evaluate și rezolvate de comun acord cu învățații umaniști interesați de dezvoltarea domeniului digital humanities în România, de producătorii de softuri și harduri, de specialiștii bibliotecilor digitale din țara noastră.

## BIBLIOGRAPHY

- Birkerts, 1994: Birkerts, Sven. *The Gutenberg Elegies: The Fate of Reading in an Electronic Age*. New York: Fawcett Columbine.
- Car, 2013: Carr, Nicholas. 'Is Google Making Us Stupid? What the Internet is Doing to Our Brains'. *The Atlantic Magazine*. 1.7.2008. Web. 20.5.2013. Web.
- Carr, 2010: Carr, Nicholas. *The Shallows: What the Internet is Doing to Our Brains*. New York: W.W. Norton.
- Hayles, 2012: Hayles, N. Katherine. *How We Think: Digital Media and Contemporary Technogenesis*. Chicago: University of Chicago Press.
- Marcus, 1963: Solomon Marcus, *Lingvistica matematică. Modele matematice în lingvistică*. Ed. Didactică și Pedagogică. București, 220 p.
- Marcus, 1964: Solomon Marcus, *Gramatici și automate finite*. Ed. Academiei, București, 256 p.
- Marcus, 1966: Solomon Marcus, Edmond Nicolau, Sorin Stati *Introducere în lingvistica matematică* București, Editura științifică și enciclopedică, 344p.
- Marcus, 1966: Solomon Marcus, *Introducere în lingvistica matematică*. Ed Științifică, București, 336 p (în colaborare cu Edmond Nicolau și Sorin Stati)
- Marcus, 1966: Solomon Marcus, *Lingvistica matematică* (a doua ediție, revizuită și completată cu patru capitole noi). Ed Didactică și Pedagogică, București, 254p
- Marcus, 1967: Solomon Marcus, *Algebraic Linguistics; Analytical Models*. Academic Press, New York, XIV + 254 p.
- Marcus, 1967: Solomon Marcus, *Introduction mathématique a la linguistique structurelle*. Dunod, Paris, XII + 282 p.
- Marcus, 1969: Solomon Marcus, *Algebraicke modely jazyka*. Ed. Academia, Prague, 289 p.
- Marcus, 1970: Solomon Marcus, *Introduzione alla linguistica matematica*. Casa editrice Riccardo Patron, Bologna, 448 p. (in collab. with E. Nicolau and S. Stati).
- Marcus, 1970: Solomon Marcus, *Proceedings of the Symposium on Algebraic Linguistics held 10-12 February 1970, Smolenice*. Publishing House, Slovak Academy of Science, Bratislava, (in colaborare cu Jan Horecky si Laszlo Kalmar).
- Marcus, 1970: Solomon Marcus, *Teoretiko-mnozestvennye modeli jazykov*. Ed. Nauka, Moscova, 1970, 332 p. (translation of the first five chapters of the book 10 and of the last chapter of the book 9).
- Marcus, 1973: Solomon Marcus, *Mathematische Poetik*. Ed. Academiei, București-Athenaum Verlag, Frankfurt am Main. 437 p.
- Marcus, 1974: Solomon Marcus, *Matematicka Poetika*. Ed. Nolit, Belgrad. 337 p.
- Marcus, 1974: Solomon Marcus, *Poetics and Mathematics*. Special issue of the journal POETICS, Mouton, The Hague, no.10. (editor și coordonator).
- Marcus, 1975: Solomon Marcus, *Semiotica folclorului. Abordare lingvistico-matematică*. Ed. Academiei, București. 268 p.(co-autor)
- Marcus, 1975: Solomon Marcus, *Semiotica folclorului. Abordare lingvistico-matematică*. Editura Academiei, București. (editor și studiu introductiv).
- Marcus, 1976: Solomon Marcus, *Matematicka analiza ctenu podruhe*. Ed. Academia, Prague, 234 p.

- Marcus, 1977: Solomon Marcus, *A nyelvi szépség matematikája*. Ed. Gondolat, Budapest, 400 p.
- Marcus, 1977: Solomon Marcus, *Metode distribuționale algebrice în lingvistică*. Ed. Academiei, București, 256 p. (co-autor).
- Marcus, 1977: Solomon Marcus, *Metode distribuționale-algebrice în lingvistică*. Editura Academiei, București, 1977 (editor și studiu introductiv).
- Marcus, 1977: Solomon Marcus, The formal study of drama. Special issue of the journal POETICS, North Holland, Amsterdam, vol.6, no.3/4, (editor).
- Marcus, 1978: Solomon Marcus, *Introduccion en la linguistica matematica*. Ed. Teide. Barcelona, 386 p. (revised and completed Spanish translation of the book 5).
- Marcus, 1978: Solomon Marcus, *La semiotique formelle du folklore. Approche linguistico-mathematique*. Ed.Klincksieck, Paris - Editura Academiei, București, (editor și studiu introductiv).
- Marcus, 1978: Solomon Marcus, *La semiotique formelle du folklore. Approche linguistico-mathematique*. Ed. Klincksieck, Paris - Ed. Academiei, București, 309 p.
- Marcus, 1981: Solomon Marcus, *Contextual ambiguities in natural & artificial languages*. Vol. 1, Ed. Communication and Cognition, Ghent, Belgium, 138 p.
- Marcus, 1981: Solomon Marcus, *Contextual Ambiguities in Natural and in Artificial Languages*, vol.1, Communication and Cognition, Ghent, Belgium, (prefațator și coordonator).
- Marcus, 1981: Solomon Marcus, *Semiotique roumaine*. Tipografia Universității din București, (în colaborare cu P.Miclău).
- Marcus, 1981: Solomon Marcus, *Snmeia gia ta snmeia*. Ed. Pneumatikos, Atena, 1981, 119 p.
- Marcus, 1981: Solomon Marcus, *Theorie et pratique de la reception*. Special issue of the journal "Degres", Brussels, (în colaborare cu I.Coteanu, P.Miclău și R.Munteanu).
- Marcus, 1982: Solomon Marcus, *Semiotica matematică a artelor vizuale*. Editura științifică și enciclopedică, București, (editor și studiu introductiv).
- Marcus, 1983: Solomon Marcus, *Contextual Ambiguities in Natural and in Artificial Languages*, vol.II, "Communication and Cognition", Ghent, Belgium, (prefațator și coordonator).
- Marcus, 1984: Solomon Marcus, *The Formal Study of Drama, II*. Special issue of the journal POETICS, Amsterdam, vol.13, no.1/2, (editor și coordonator).
- Marcus, 1986: Solomon Marcus, *To Paradoxo*. Ed. Pneumatikos, Atena, 126 p.
- Marcus, 1996: Solomon Marcus, *Language, Logic, Cognition and Communication; A Semiotic, Computational and Historical Approach*. Report 9/96. Grup de Recerca en Linguistica Matematica i Enginyeria del Llenguatge. Reports Universitat Rovira i Virgili, Tarragona, Spain, 184 p.
- Marcus, 2004: Solomon Marcus, *Mathematics in Romania*, Baia Mare, CUB Press 22, 84 p.
- Marcus, 2007: Solomon Marcus, *Words and languages everywhere*, Polimetrica, International Scientific Publisher, 542 p.
- Marcus, 2010: Solomon Marcus, *Singurătatea matematicianului*, Editura Liternet, colecția Palimpsest.
- Moretti, 2013: Moretti, Franco *Distant Reading*, Verso, London, New York.
- Moretti, 2015: Moretti, Franco *Atlas of the European Novel 1800-1900* Verso, London, New York.
- Moretti, 2014 a: Moretti, Franco *The Bourgeois Between History and Literature* Verso, London, New York.

- Moretti, 2014 b: Moretti, Franco *The Way of the World The Bildungsroman in European Culture New Edition*, Verso, London, New York.
- Sosnoski, 1999: Sosnoski, James J. 'Hyper-Readers and Their Reading-Engines.' *Passions, Politics, and 21st Century Technologies*. Ed. Gail E. Hawisher und Cynthia L. Selfe. Urbana: Utah State University Press, 161–77.
- Stein, 2015: Stein, Bob. 'A Taxonomy of Social Reading: A Proposal'. *The Institute for the Future of the Book*. June 10, 2015.