



Tehnici digitale pentru analiza romanului românesc

**Vlad POJOGA, Ștefan BAGHIU,
Emanuel MODOC, Daiana GÂRDAN,
Andreea COROIAN GOLDIȘ**

Universitatea „Lucian Blaga” din Sibiu, Facultatea de Litere și Arte;
Universitatea „Babeș-Bolyai” din Cluj-Napoca, Facultatea de Litere
Lucian Blaga University of Sibiu, Faculty of Letters and Arts;
Babeș-Bolyai University of Cluj-Napoca, Faculty of Letters

Personal e-mail: vlad.pojoga@ulbsibiu.ro, ermocod@gmail.com, alexandra.gardan@gmail.com,
andreeacoroian@gmail.com, stefan.baghiu@ulbsibiu.ro

Digital Tools for the Analysis of the Romanian Novel

Drawing from the most recent developments in the field of Digital Humanities, the article showcases some of the most promising methods as well as some of the most useful tools for computational textual analysis. Textometry, stylometry and network analysis (alongside topic modelling or distributional semantics) are nowadays methods that, through the recent developments in digital formalism proposed by Matthew L. Jockers or the researchers from Stanford Literary Lab, are able to reveal literary and cultural patterns at a macroanalytical method. The article will delve into computational stylistics by testing the Stylo package, developed by Maciej Eder, Jan Rybicki and Mike Kestemont, on a sample corpus provided by the *ASTRA Data Mining Project*, attempting to explore the merits of such methods in the field of Romanian literary studies.

Keywords: digital humanities, Romanian literature, 19th century novel, stylometry, textometry, network analysis, macroanalysis



Începând cu anii '90, odată cu dezvoltarea la scară largă a capacităților digitale și cu creșterea gradului de accesibilitate al acestora – după emergența internetului și în directă proporționalitate cu evoluția rapidă a tehnologiei informatice (atât *hardware* cât și *software*) –, literatura și studiile literare își extind aria de concentrare, profitând de câteva din modificările fundamentale ale mediului de transmitere și stocare a informației. Astfel, apar noi trăsături ale literaturii *per se*, precum interacțiunea ca posibilitate concretă de a influența *calea* pe care o parcurge acțiunea¹, concomitent cu creșterea capacităților studiilor literare atât în privința analizei unui volum mult mai

ridicat de texte, cât și în utilizarea unor instrumente digitale de analiză, care facilitează prezentarea unor imagini *macro* și reprezentări grafice complexe. În spațiul cultural românesc, cu toate că a pătruns cu relativă întârziere, paradigma *digital humanities*² (în care se încadrează prezenta cercetare) începe să fie tot mai discutată în ultimii ani³, iar proiecte de cercetare și culturale de tipul *INTELLIT* sau *ASTRA Data Mining* cresc tot mai mult profilul public al acesteia. Articolul de față, structurat în trei părți, se concentrează asupra posibilităților digitale de analiză cantitativă, inventariind câteva unelte de prelucrare a datelor textuale și a metadatelor extrase din arhive de mari

dimensiuni, și propune câteva previzualizări aplicative ale unor astfel de demersuri⁴.

De la analiză cantitativă la analiză computațională

Cu toate că își are punctul de pornire în analiza cantitativă, cea din urmă noțiune se desparte de cea dintâi în câteva puncte nodale. Însă înainte de a-i trasa contururile, merită să acoperim conceptul de bază, întrucât el se află în continuă dezvoltare chiar și autonom de analiza computațională. Termenul-cheie în cazul acestui tip de analiză este „cantitativ”, termen care se opune celui de „calitativ”. În contemporaneitate, studiile literare cantitative au devenit cvasi-sinonime cu conceptul de *distant reading* propus de Franco Moretti în 2013, în opoziție cu tradiționalul *close reading* anglo-saxon. În timp ce analiza cantitativă presupune discutarea anumitor date centralizate pentru deoalarea relațiilor, frecvențelor, contextului etc., fiind utilizată mai ales pentru instrumentalizarea conceptelor teoriilor formaliste și structuraliste, analiza calitativă se poate rezuma adesea la interpretări subiective/impresioniste asupra fenomenelor literare. De aceea, majoritatea studiilor cantitative și a statisticilor literare au fost asociate încă de la origini unor științe auxiliare ale studiului literaturii (bibliologia și bibliografia) și unor științe socio-umane cu o puternică bază statistică în metodologia proprie (economia, sociologia) sau lingvisticii. În cercetarea literară, analiza cantitativă este importantă în studierea producției literare, în stabilirea paternității textelor literare, în analiza genurilor literare, în geografia literară sau în prin utilizarea statisticii sociologice și a stilometriei. Odată cu era digitală, graficele, hărțile sau arborii⁵ au devenit instrumente fundamentale în analiza cantitativă. Dacă statistica literară este cel mai des legată de studiile de sociologie a literaturii ale lui Robert Escarpit (1966), analiza cantitativă a depășit statisticile producției literare, consolidându-și o tot mai puternică poziție în interiorul teoriei literare sui generis.

Cu toate că majoritatea cercetătorilor plasează începuturile analizei cantitative în jurul anului 1850, câteva precedente merită menționate. Foarte puțin cunoscut în studiile cantitative, pionierul acestei analize în domeniul literaturii este istoricul și bibliograful francez Charles Louandre (1812-1882), care a deplâns încă din 1842 lipsa unor cercetări a tuturor cărților publicate, descriind canonul literar ca o „constelație” în care anumite „stele” sunt mai vizibile decât altele și comparând o listă bibliografică a unei biblioteci cu un „cimitir” din cauza masei impresionante de texte necitite rămase în arhive. Această perspectivă asupra producției a fost reluată în diferite forme, cel mai celebru fiind conceptul de Great Unread al lui Margaret Cohen⁶, care descrie totalitatea

textelor „uite” sau „necitite”. Acest concept a fost extrem de util în cercetarea cantitativa a literaturii, mai ales în studierea plot-urilor narative pentru distant reading („citirea de la distanță”) teoretizată de Franco Moretti în 2000 în articolul „Conjectures on World Literature”⁷. Teoreticianul italian a devenit astfel una dintre figurile centrale ale cercetărilor cantitative mai ales prin pledoaria pentru un tip de analiză a literaturii care exclude lectura efectivă a cărților și care cere permanent descoperirea unor metode prin care pot fi „citate” cantități uriașe de literatură prin analiza anumitor tipare narative. Pentru acest tip de analiză cantitativă sunt extrem de importante conceptele și ideile formalismului rus, de la care teoreticienii distant reading se revendică explicit. În *Canon/Archive*⁸, un loc privilegiat este deținut de formalismul cantitativ (*quantitative formalism*) formulat de Stanford Literary Lab pentru analiza canonului literar și a prestigiului autorilor, a genurilor literare, a relațiilor dintre personaje sau a „sonorității” (*loudness*) stilului literar al unui autor. Alături de conceptul de *distant reading* propus de Franco Moretti, unul la fel de important este cel de *macroanalysis* preluat din economie în studiile literare de către Matthew L. Jockers⁹, care declară *close reading*-ul „nepractic” în contextul big data (al posibilităților digitale de analiză a unei cantități uriașe de texte).

Din acest punct, diferențele dintre analiza cantitativă și cea computațională încep să devină mult mai clar definite. Cu toate că, așa cum am spus, analiza computațională este, în fond, un tip de analiză cantitativă, se poate face analiză cantitativă și fără a apela strict la instrumente de analiză computațională. De altfel, cele mai recente studii cantitative în spațiul românesc – mă refer strict la cele elaborate în sialul tendințelor *World Literature*, a se vedea studiile semnate de Ștefan Baghiu¹⁰, Daiana Gârdan¹¹ sau Andrei Terian¹² – utilizează instrumente de lucru terțiare și, detaliu important, analogice, în format tipărit. Ne referim la *Dicționarul Cronologic al Romanului Românesc* și la *Dicționarul Cronologic al Romanului Tradus*, editate de Academia Română, ambele conținând date utile pentru demersuri cantitative, dar finalmente limitate atât sub aspect tehnic (inexistența formatului digital al acestor dicționare), cât și din punctul de vedere al metadatelor care pot fi coroborate prin intermediul lor. Ceea ce suplimentează, în plan metodologic, analiza computațională, prin diversele *soft*-uri elaborate în limbaje de programare precum R și Python, este o analiză cantitativă care lucrează direct cu un corpus literar dat (fie că e vorba de poezie, proză, sau publicistică) și care generează date care urmează a fi analizate și interpretate de către cercetători.¹³



Instrumente de analiză textuală digitală și de prelucrare a datelor

Un instrument de bază pentru analiza textuală computațională este TXM (acronim pentru *Textométrie*). Elaborat în anii '80 în Franța, TXM este larg utilizat pentru analize lexicologice elementare: frecvența relativă a cuvintelor într-un corpus dat, lungimea medie a cuvintelor într-o operă, concordanțe și co-ocurențe. Pe lângă aceste analize, instrumentul mai oferă și o varietate de operațiuni pre-programate: progresii, lematizări, prelucrări grafice ale configurației lexicale a unei cărți etc. Chiar dacă TXM poate procesa și texte în format simplu (de tip .txt), cele mai elaborate analize depind de cantitatea (și existența, în primul rând) a metadatelor dintr-un corpus. De pildă, un roman digitalizat în format .txt nu va conține altceva decât suma vocabulelor care îl compun, pe când același text în format .xml poate conține, pentru fiecare cuvânt din conținutul romanului, o serie de metadata (categoria gramaticală din care face parte, paradigme, declinări etc.) folosite pentru analize mult mai nuanțate (de tipul „frecvența adverbilor de mod într-un roman” sau „reprezentarea grafică a câmpului lexical al călătoriei”). Dincolo de asta, TXM poate oferi, prin funcțiile lui cele mai simple, date utile care depășesc cu mult o elementară funcție de căutare dintr-o carte în format electronic.

În ceea ce privește uneltele de analiză textuală online, gratuite (al căror avantaj evident este reprezentat de gradul lor ridicat de accesibilitate și relativa ușurință a utilizării), de menționat sunt Textalyser (textalyser.net), cu capacități de procesare reduse – poate analiza un număr maxim de 1000 de cuvinte –, dar cu utilitate imediată în numărarea cu acuratețe a unor cuvinte – a căror dimensiune poate fi selectată în funcție de numărul de caractere –, generând un tabel de ocurențe și procentaje ale cuvintelor în textul dat și Voyant Tools (voyant-tools.org), care produce analize mult mai complexe de text, precum densitatea vocabularului, numărul mediu de cuvinte per propoziție, graficele apariției cuvintelor în funcție de secțiunile textului încărcat, legăturile imediate ale cuvintelor în text (ce se regăsește imediat înaintea sau imediat după un cuvânt) sau corelări ale unor termeni între care există legături directe.

Un alt instrument, cel mai ofertant din punct de vedere al potențialului computațional, dar și al multiplelor roluri pe care le poate acoperi, este pachetul *Stylo*¹⁴, elaborat de Maciej Eder (cercetător la Institutul Limbii Poloneze din Cracovia), Jan Rybicki (de la Universitatea Jagielonă din Cracovia) și Mike Kestemont (de la Universitatea Antwerp din Belgia). Stilometria în general și stilistica computațională în particular vizează studiul cantitativ al stilului unei scriituri și este utilizat cu preponderență în studiile de atribuire a autoratului

(în cazul cărților semnate sub pseudonim). În ceea ce privește stilometria, o ramură a cercetărilor cantitative care se bazează pe analiza formală a limbajului literar și nonliterar, originile pot fi plasate în 1851, când Augustus de Morgan a intuit că autorul unui text poate fi stabilit după analiza cantitativă a lungimii cuvintelor unui text¹⁵. Deși nu s-au concretizat într-o metodă, ideile sale au fost continuate de George Zipf sau de G. Udny Yule. Dincolo de stabilirea autorului unui text, stilometria a fost utilizată și în cercetări ale cronologiei operelor literare, conform ideii că modificarea stilului unui autor poate fi observată prin analiza cantitativă a cuvintelor, după cum au arătat David R. Cox și Leonard Brandwood studiind operele lui Platon. Conform lui David I. Holmes, adevăratul potențial al stilometriei a fost atins la începutul anilor 1960 prin Charles Frederick Mosteller și David Wallace și mai ales odată cu „metoda Morton”, care analizează cantitativ locul cuvintelor în propoziții pentru stabilirea autorului unui text și care au fost utilizate mai ales în analiza textelor atribuite lui William Shakespeare. În prezent, studiile de stilometrie urmează calea trasată de John Borrows, care a utilizat clustere de cuvinte secundare (cele mai puțin importante într-un text, propoziții și conjuncții) pentru a stabili stilul propriu al unor autori ca Walter Scott, Byron sau Jane Austen. Cele mai utile instrumente astăzi pentru stilometria literară computațională au fost propuse de Matthew L. Jockers (2014). Aspectul computațional al stilometriei contemporane nu face decât să continue această tradiție, dar la un cu totul alt nivel de reprezentare și la un cu totul alt potențial de analiză. La cel mai elementar nivel, pachetul *Stylo* (realizat în limbajul de programare R) analizează semnătura auctorială a unui corp literar dat prin cuantificarea unor aspecte ce țin de lexic și de frecvența relativă a unor cuvinte dintr-un text. Deoarece este popularizat în studiul identificării mărcilor auctoriale, *Stylo* lucrează cu colecții mari de texte, de la câteva zeci la câteva sute. Un caz, notoriu în domeniu, de utilizare de succes a analizei stilometrice îl reprezintă descoperirea lui Patrick Juola (în 2013), care a identificat în romanul *The Cuckoo's Calling*, semnat de un Robert Galbraith, marca auctorială a lui J.K. Rowling.

Unul dintre cele mai utilizate softuri de vizualizare a datelor și metadatelor este Tableau, cu versiunile sale Tableau Desktop (contra cost) și Tableau Public (gratuit, dar cu capacități reduse de salvare a rezultatelor, acestea fiind publicate online, pe platforma proprie). În esență, Tableau preia o serie de date indexate în prealabil într-un document care poate fi de mai multe tipuri, de la foi de calcul Excel și fișiere .txt și PDF până la fișiere cu extensii specifice, spațiale sau statistice, și generează grafice complexe, diagrame GANTT, arbori, combinații duale, cercuri suprapuse, buline proporționale, hărți¹⁶ și nu numai.

Oferind control utilizatorului asupra multor detalii ale reprezentării vizuale rezultate, de la dimensiune, culoare, text suprapus și posibilitatea de a procesa și combina mai multe seturi de date complexe, Tableau necesită un grad mai mare de atenție și cunoaștere a softului din partea utilizatorului, însă produce, deseori, rezultate spectaculoase.

Un ultim instrument, folosit deseori alături de analiza stilometrică, este *Gephi*, un soft care poate prelua datele oferite de *Stylo* și pune la dispoziție o suită extrem de complexă de operațiuni statistice. *Gephi* este folosit cu preponderență în analiza de rețea și, în general, în reprezentări rețelare ale unor date obținute extern. Departe de fi un model extravagant, de dată recentă, în câmpul umanioarelor, rețeaua și analizele rețelare ale fenomenelor sociale, culturale și literare cunosc, odată cu emergența formalismului digital, deschideri spectaculoase. Aceste noi posibilități se leagă în special de amploarea recentelor analize, de capacitatea – încă în proces de testare, la nivel global – a noilor instrumente de metabolizare a unor date pe care tradiția comparatistă le-ar fi considerat pe de-o parte irelevante (ocupându-se doar de vârfuri), dar, pe de-altă parte, și cu atât mai important, le-ar fi considerat imposibil de realizat. Suntem, fără îndoială, martori ai unei schimbări de paradigmă în ceea ce privește noțiunea de rețea și demersurile analitice aferente ei: se petrece, în noua vârstă a lecturii (distant reading) o mutație, de la înțelegerea și aplicarea, din punct de vedere metodologic, a rețelei ca metaforă, la instrumentarea ei ca auxiliar grafic, vizual, generator al unei viziuni de tip macro, ce expune cu succes vase comunicante care, de la mai puțină distanță, ar fi imposibil de văzut.

Despre câștigurile unui asemenea instrument s-a discutat, cu efervescență, în numeroase proiecte colective, în presă și în volume de autor. Autorii volumului *Networks. A very Short Introduction*, semnaleză poate cel mai școlăresc marele atu al acestei metode: „În timp ce abordările rețelare elimină foarte multe dintre aspectele și caracteristicile individuale ale fenomenelor pe care le tratează, o păstrează poate pe cea mai importantă, anume, rețeaua nu alterează dimensiunea sistemului, i.e. numărul de elemente, modelele de interacțiune, setul specific de relații și conexiuni dintre elemente. O metodă atât de simplificată este suficientă să reprezinte proprietățile unui sistem”.

Implementarea generatoarelor digitale de rețele în câmpul studiilor literare naționale combină două tipuri de câștiguri, 1. ele recuperează și fac vizibile sensuri de circulație, importuri, relații intra-naționale și internaționale, modele și moduri de emergență a unor curente și fenomene literare/culturale, facerea și desfacerea unor grupări, măsura capitalului simbolic al altora ș.a. și 2. la nivel individual, inserează un

fenomen particular, tratat în general la firul ierbii de către tradiția critică, într-un tablou mai mare, îl pune pe o hartă (geografică, politică, literară), reșapând astfel conglomeratul de verdicte și clișee premergătoare.

S-a discutat, mai ales în cazul culturilor periferice – spațiul românesc fiind unul dintre culturile vizate –, despre un așa-zis impact dăunător al noilor metode digitale, în sensul în care acestea, depărtându-se de text, ar conduce la crearea unui deficit de înțelegere a fenomenelor particulare. Or, o conversație despre statutul noului tip de lectură (ne referim, evident, la lectura distantă), care să răspundă, poate, la întrebări legate de statutul paradigmatic al acesteia (o paradigmă științifică o anulează și o înlocuiește întotdeauna pe cea anterioară, este acesta cazul formalismului digital? Și dacă nu, mai putem vorbi despre o nouă paradigmă?), este fără îndoială un pas necesar. Deocamdată, și în economia proiectului de față, aderăm la un punct de vedere mai degrabă complementar, unul pe care îl exprimau și Matthew Jockers și Franco Moretti în primele volume care teoretizau și puneau la lucru în același timp noile metode. Aplicat la analizele rețelare, îl redau în termenii lui Jeffrey Drouin, care scrie în *Close and Distant-Reading Modernism*: „Metodele digitale și lectura distantă nu au în intenție înlocuirea metodelor tradiționale, ci mai degrabă conducerea spre locurile precise în care ar trebui aplicate, sugerarea răspunsurilor acolo unde datele lor sunt neconcludente (..) Mașinăriile nu înlocuiesc agentul uman în facerea sensului, dar îl ajută să privească mai departe și să sape mai adânc”.

Previzualizări aplicative

Dintre cele mai spectaculoase rezultate pe care analizele de tip rețea le pot genera cu instrumente specializate (de pildă *Gephi*, probabil cel mai accesibil și accesat instrument de generare a rețelelor în câmpul studiilor literare, dar nu numai), dar și cele care prezintă cel mai mare interes pentru corpusuri de roman, se numără: atribuirea auctoratului și devoalarea pseudonimului, apropierea și depărtările stilistice dintre un autor și altul, dar și între romane semnate de același autor, particularități de gen (scriitura feminină vs. masculină), toate semnaleză importante pentru înregistrarea evoluției și a modernizării, mai ales în cazul unei culturi tineri.

Înainte de a trece la câteva ilustrări, o scurtă precizare cu privire la aceste reprezentări vizual-abstracte de rețea se impune: cu toate că nodurile de rețea își păstrează o anumită centralitate în interiorul „câmpului” stilistic, natura stocastică a instrumentului de analiză statistică face ca poziția „în spațiu” a fiecărui roman să fie supusă hazardului. De câte ori datele sunt încărcate și grupate algoritmic, plasarea în spațiu

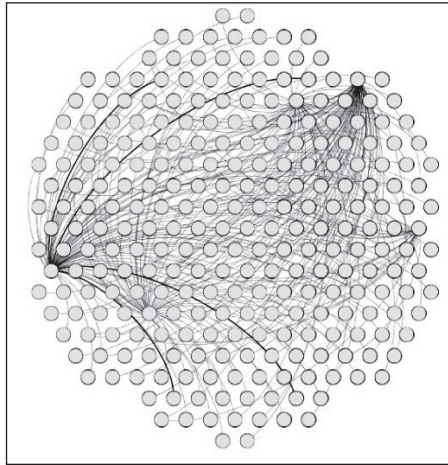


Fig. 1: O rețea înainte de procesarea datelor prin clustering

nu e niciodată identică cu cea anterioară, cu toate că dispunerea centru-periferie rămâne aceeași. În același timp, o anumită doză de precauție se cuvine cerută în cazul interpretării propriu-zise a rețelelor generate prin programul folosit (*Gephi*)¹⁷, întrucât rezultatele pot diferi de la un algoritm de grupare utilizat la altul. Pentru a încerca, pe cât posibil, să reducem aceste riscuri, am utilizat un singur algoritm, ale cărui rezultate au fost supuse constant unor testări multiple, pentru a putea respecta principiul reproductibilității datelor.

Cu toate că demersul nostru a presupus o analiză cantitativă incompletă (fiind vorba doar de un eșantion de 23 de romane din totalul de 158, reprezentând 15% din corpusul total), rețelele produse cu ajutorul

instrumentelor de analiză computațională utilizate au presupus un efort semnificativ de procesare „subiectivă” a datelor. Ceea ce programul de realizare a rețelelor încarcă, într-o primă etapă, e o masă nediferențiată de noduri și ramuri (vezi Fig. 1). Pentru a putea obține o reprezentare inteligibilă, care să redea informațiile necesare unui demers analitic-interpretativ al datelor, trebuie selectată și reglată o schemă de dispunere (*layout*), bazată pe algoritmi de grupare (*clustering*) utilizați în analize statistice în alte domenii. Pentru crearea rețelelor finale, am folosit un algoritm de grupare (numit „Force Atlas”) care tratează nodurile de rețea ca pe niște particule dinamice, care se resping, în timp ce ramurile de rețea, a căror „greutate” (calculată după datele cantitative pe care le pun la dispoziție) determină întreaga dinamică a nodurilor și acționează ca niște corzi elastice, care își apropie nodurile unul de celălalt. Din acest punct de vedere, un nod care angajează un număr mai mare de „conexiuni” cu alte noduri devine, astfel, central, în timp ce un alt nod care e mai puțin conectat la rețea rămâne „expulzat” la margine.

Să luăm, așadar, mostră de 23 de romane publicate între anii 1860 și 1900. Selecția, cu toate că e una inegală (8 romane de mistere și de senzație la 15 romane haiducești), a luat în calcul și testarea autorilor care scriau serii întregi de romane (N.D. Popescu și Panait Macri sunt exemplari în acest sens). O analiză stilometrică, realizată prin pachetul *Stylo*, redă următoarea analiză de *cluster*:

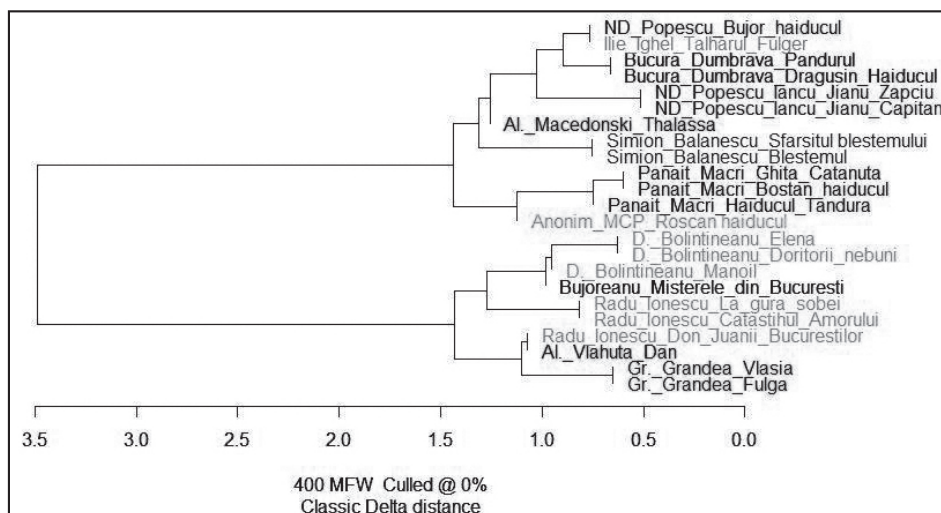


Fig. 2. Analiză stilometrică pe 23 de romane. Cu cât două sau mai multe romane sunt mai apropiate în stil, cu atât mai mult poziția lor în diagramă va fi mai aproape de 0.

Ceea ce face, propriu-zis, analiza stilometrică realizată de *Stylo*, e să calculeze, utilizând o varietate impresionantă de metode statistice, anumite „afinități” de construcție a enunțurilor pe baza unui criteriu elementar (cele mai frecvente 400 de cuvinte). Așa cum se poate vedea în **Fig. 2**, autorii cu două sau mai multe romane sunt grupați împreună și împărtășesc „brațe” ale rețelei doar cu autori apropiați stilistic. În câteva cazuri excepționale (romanul *Bujor haiducul* al lui N.D. Popescu și *Tâlharul Fulger* al lui Ilie Ighel), apropierea dintre autori sunt atât de mari încât unele romane scrise de același autor se „îndepărtează” stilistic în rețea. În alte cazuri, rezultatele sunt și mai surprinzătoare: potrivit analizei stilometrice, nuvela *Roșcan haiducul*, semnată MCP (dar având, la final, mențiunea: „Această nuvelă este întocmită după o narațiune germană”), pare cea mai apropiată stilistic de scriitura lui Panait Macri, un cunoscut autor de colportaj literar și al doilea cel mai prolific autor de romane după mai celebrul N.D. Popescu.

O reprezentare rețelară mult mai nuanțată a acestor rezultate o poate oferi *Gephi*, datorită căruia pot fi decelate nu doar apropieri de rang 1 între doi autori, ci și de rang 2 sau 3. În cazul romanului haiducec românesc, rezultatele sunt frapante prin caracterul lor aproape nediferențiat între multipli autori:

Din punct de vedere stilometric, autorii care formează ochiuri de rețea la „periferie” (Simeon Bălănescu, Panait Macri, Radu Ionescu, N.D. Popescu, D. Bolintineanu) sunt cei mai banali stilistic, în vreme ce autori precum Al. Vlahuța și Al. Macedonski apar doar conjunctural în „vecinătatea” acestor autori, relațiile lor stilistice ținând strict de faptul că eșantionul este unul izolat. Ceea ce ne relevă chiar și o incipientă analiză computațională stilometrică este, credem noi, grăitor: departe de a fi opere valoroase estetic, apropierea stilistică dintre autori poate fi dezvoltată ca argument pentru strategiile de publicare a diferiților autori. Cu totul anodini estetic, acești autori au fost considerați printre cei mai populari scriitori ai epocii, în cazul lui N.D. Popescu fiind vorba, după toate datele, de numeroase ediții succesive ale romanelor sale haiducești sau istorice¹⁸. Romanele acestor autori nu reprezintă nimic altceva decât un barometru a ceea ce se putea numi, la secolul al XIX-lea, literatură de duzină, iar afinitățile stilistice dintre ei nu pot decât să confirme o asemenea presupuziție.

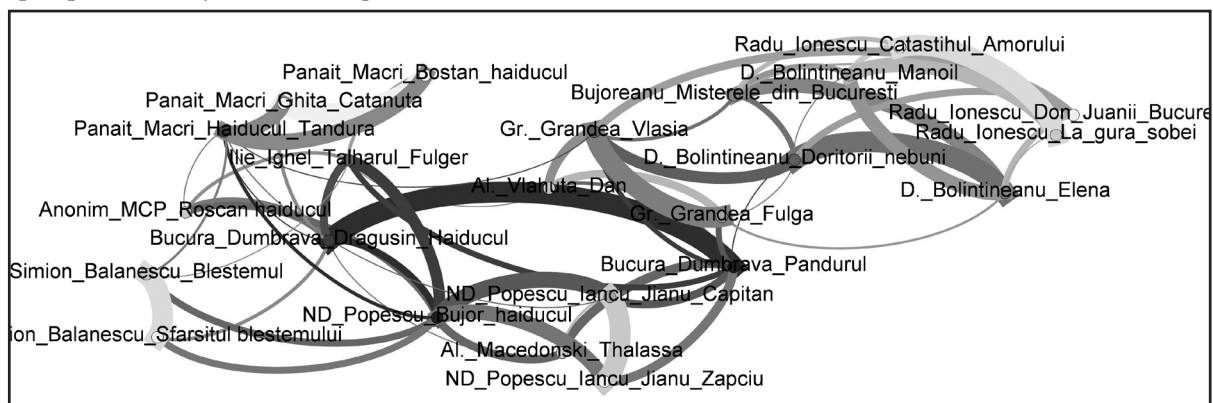


Fig. 3. Reprezentare rețelară a 23 de romane publicate în secolul al XIX-lea. Cu cât este mai accentuată o ramură de rețea, cu atât mai evidentă relația de afinitate stilistică între doi autori.

Note:

1. pentru o analiză amănunțită a felului în care schimbarea mediului influențează producția și lectura literaturii și o taxonomie a literaturii digitale narative, vezi Vlad Pojoga, „The Change of Medium and the Medium that Changes: Narrative Literature, Networks, and the Digital”, *Transylvanian Review*, vol. XXVIII, Supplement 1 (2019): 203-215.

2. una dintre primele discuții in extenso asupra acesteia în România poate fi văzută în Alex Goldiș, „Digital Humanities – o nouă paradigmă teoretică?”, *Revista Transilvania*, 12 (2014).

3. vezi Mihaela Ursa, „Is Romanian Culture Ready for the Digital Turn?”, *Metacritic Journal for Comparative Studies and Theory*, vol. 1, nr. 1 (2015): 80-97; Ovio Olaru, „What is Digital Humanities and What’s It Doing in Romanian Departments?”, *Revista Transilvania*, 5-6 (2019): 30-37.

4. Pentru o panoramă a studiilor cantitative în spațiul românesc vezi Mihnea Bălci, „Studii cantitative recente în spațiul românesc. Între analiză instituțională și problema traducerilor” în *Revista Transilvania*, 2 (2019): 11-18.

5. Franco Moretti, *Graphs, maps, trees. Abstract models for a Literary History* (London, New York: Verso, 2005).

6. Margaret Cohen, *The Sentimental Education of the Novel* (Princeton: Princeton University Press, 1999).



7. Franco Moretti, „Conjectures on World Literature”, *New Left Review*, 1 (2000).
8. Franco Moretti, ed., *Canon/Archive* (New York: n+1 Foundation, 2017).
9. Matthew L. Jockers, *Macroanalysis. Digital Methods & Literary History* (Chicago: University of Illinois Press, 2013).
10. Ștefan Baghiu, „Translating Novels in Romania: The Age of Socialist Realism. From an Ideological Center to Geographical Margins”, în *Studia UBB Philologia*, vol. LXI, nr. 1 (2016): 5-18;
11. Daiana Gârdan, „Evoluția romanului erotic din prima jumătate a secolului al XX-lea. Între exercițiu și canonizare”, *Revista Transilvania*, nr. 7 (2018): 1-7; Daiana Gârdan, “The Great (Female) Unread. Romanian Women Novelists in the First Half of the Twentieth Century: A Quantitative Approach”, *Metacritic Journal for Comparative Studies and Theory*, vol. 4, nr. 1 (2018): 109-124.
12. Andrei Terian, „Big Numbers: A Quantitative Analysis of the Development of the Novel in Romania”, *Transylvanian Review*, vol. XXVIII, Supplement 1 (2019): 55-71.
13. Cercetări similare au fost efectuate în Emanuel Modoc, „Traveling Avant-Gardes. The Case of Futurism in Romania”, în *The Culture of Translation in Romania / Übersetzungskultur und Literaturübersetzen in Rumänien*, ed. Maria Sass, Ștefan Baghiu și Vlad Pojoga (Berlin: Peter Lang, 2018), 45-62; Ștefan Baghiu, „Strong Domination and Subtle Dispersion: A Distant Reading of Novel Translation in Communist Romania (1944–1989)”, în *The Culture of Translation in Romania / Übersetzungskultur und Literaturübersetzen in Rumänien*, ed. Maria Sass, Ștefan Baghiu și Vlad Pojoga (Berlin: Peter Lang, 2018), 63-84; Vlad Pojoga, „A Survey of Poetry Translations in Romanian Periodicals (1990-2015)”, în *The Culture of Translation in Romania / Übersetzungskultur und Literaturübersetzen in Rumänien*, ed. Maria Sass, Ștefan Baghiu și Vlad Pojoga (Berlin: Peter Lang, 2018), 99-121; Ovio Olaru, “Translating Nordic Noir Bestsellers. Towards a Comparative View on German and Romanian markets”, în *The Culture of Translation in Romania / Übersetzungskultur und Literaturübersetzen in Rumänien*, ed. Maria Sass, Ștefan Baghiu și Vlad Pojoga (Berlin: Peter Lang, 2018), 231-241.
14. Vezi Eder, M., Rybicki, J. and Kestemont, M., „Stylometry with R: a package for computational text analysis”, în *R Journal*, vol. 8, no. 1 (2016): 107-121, <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>.
15. Vezi, pentru un istoric cuprinzător, Anthony Kenny, *The Computation of Style. An Introduction to Statistics for Students of Literature and Humanities* (Oxford: Pergamon Press, 1982). Vezi și David I. Holmes, „The Evolution of Stylometry in Humanities Scholarship”, în *A Companion to Digital Literary Studies*, ed. Ray Siemens, Susan Schreibman (Blackwell Publishing: New Jersey, 2013).
16. De exemplu, atât hărțile din Pojoga, „A Survey”, cât

- și cele dintr-un alt articol al acestui dosar, Ștefan Baghiu, Vlad Pojoga, Teodora Susarenco, Radu Vancu, Emanuel Modoc, „Geografia internă a romanului românesc în secolul al XIX-lea”, *Revista Transilvania*, 10 (2019): 29-43, au fost generate în Tableau. Vezi recent utilizat softul pentru hărți ale traducerilor romanelor străine în România între 1944-1989 în Ștefan Baghiu, „Translating Hemispheres: Eastern Europe and the Global South Connection through Translationscapes of Poverty” în *Comparative Literature Studies*, 56.3 (2019): 487-503.
17. Vezi, pentru o expunere mai detaliată a mecanismelor din spatele acestui *soft* și a riscurilor de utilizare în domeniul studiilor literare, Dennis Tenen, *Visual-Quantitative Approaches to the Intellectual History of the Field. A Close Reading*, în *Futures of Comparative Literature. ACLA State of the Discipline Report*, ed. Ursula K. Heise, (London and New York: Routledge, 2017), 258-266.
 18. Ioana Drăgan notează că „N.D. Popescu (...) înregistrează 25 de ediții între 1881 și 1912 pentru cele 42 de titluri”, vezi Ioana Drăgan, *Romanul popular în România – literar și paraliterar* (Cluj-Napoca, Casa Cărții de Știință, 2006), 122.

Bibliography:

- Bălăci, Mihnea. „Studii cantitative recente în spațiul românesc. Între analiză instituțională și problema traducerilor”. *Revista Transilvania*, 2 (2019): 11-18.
- Baghiu, Ștefan, Vlad Pojoga, Teodora Susarenco, Radu Vancu, Emanuel Modoc. „Geografia internă a romanului românesc în secolul al XIX-lea”. *Revista Transilvania*, 10 (2019): 29-43
- Baghiu, Ștefan. „Translating Hemispheres: Eastern Europe and the Global South Connection through Translationscapes of Poverty”. *Comparative Literature Studies*, 56.3 (2019): 487-503.
- Baghiu, Ștefan. “Translating Novels in Romania: The Age of Socialist Realism. From an Ideological Center to Geographical Margins”. În *Studia UBB Philologia*, vol. LXI, nr. 1, (2016): 5-18.
- Baghiu, Ștefan. “Strong Domination and Subtle Dispersion: A Distant Reading of Novel Translation in Communist Romania (1944–1989)”. În *The Culture of Translation in Romania / Übersetzungskultur und Literaturübersetzen in Rumänien*, 63-84, editat de Maria Sass, Ștefan Baghiu și Vlad Pojoga. Berlin: Peter Lang, 2018.
- Cohen, Margaret. *The Sentimental Education of the Novel*. Princeton: Princeton University Press, 1999.
- Drăgan, Ioana. *Romanul popular în România – literar și paraliterar*. Cluj-Napoca, Casa Cărții de Știință, 2006.
- Eder, M., Rybicki, J. and Kestemont, M. „Stylometry with R: a package for computational text analysis”. În *R Journal*, vol. 8, no. 1 (2016): 107-121, <https://journal.r-project.org/archive/2016/RJ-2016-007/>

- index.html.
- Gârdan, Daiana. "The Great (Female) Unread. Romanian Women Novelists in the First Half of the Twentieth Century: A Quantitative Approach". *Metacritic Journal for Comparative Studies and Theory*, vol. 4, nr. 1 (2018): 109-124.
- Gârdan, Daiana. „Evoluția romanului erotic din prima jumătate a secolului al XX-lea. Între exercițiu și canonizare”, *Revista Transilvania*, nr. 7 (2018): 1-7.
- Goldiș, Alex. „Digital Humanities – o nouă paradigmă teoretică?”. *Revista Transilvania*, 12 (2014).
- Holmes, David I. "The Evolution of Stylometry in Humanities Scholarship". În *A Companion to Digital Literary Studies*, editat de Ray Siemens și Susan Schreibman. New Jersey: Blackwell Publishing, 2013.
- Jockers, Matthew L. *Macroanalysis. Digital Methods & Literary History*. Chicago: University of Illinois Press, 2013.
- Kenny, Anthony. *The Computation of Style. An Introduction to Statistics for Students of Literature and Humanities*. Oxford: Pergamon Press, 1982.
- Modoc, Emanuel. "Traveling Avant-Gardes. The Case of Futurism in Romania". În *The Culture of Translation in Romania / Übersetzungskultur und Literaturübersetzen in Rumänien*, editat de Maria Sass, Ștefan Baghiu și Vlad Pojoga, 45-62. Berlin: Peter Lang, 2018.
- Moretti, Franco ed. *Canon/Archive*. New York: n+1 Foundation, (2017).
- Moretti, Franco. „Conjectures on World Literature”. *New Left Review*, 1 (2000).
- Moretti, Franco. *Graphs, maps, trees. Abstract models for a Literary History*. London, New York: Verso, 2005.
- Olaru, Ovio. "Translating Nordic Noir Bestsellers. Towards a Comparative View on German and Romanian markets". În *The Culture of Translation in Romania / Übersetzungskultur und Literaturübersetzen in Rumänien*, 231-241, editat de Maria Sass, Ștefan Baghiu și Vlad Pojoga. Berlin: Peter Lang, 2018.
- Olaru, Ovio. "What is Digital Humanities and What's It Doing in Romanian Departments?". *Revista Transilvania*, 5-6 (2019): 30-37.
- Pojoga, Vlad. "A Survey of Poetry Translations in Romanian Periodicals (1990-2015)". În *The Culture of Translation in Romania / Übersetzungskultur und Literaturübersetzen in Rumänien*, 99-121, editat de Maria Sass, Ștefan Baghiu și Vlad Pojoga. Berlin: Peter Lang, 2018.
- Pojoga, Vlad. "The Change of Medium and the Medium that Changes: Narrative Literature, Networks, and the Digital". *Transylvanian Review*, vol. XXVIII, Supplement 1 (2019): 203-215.
- Tenen, Dennis. "Visual-Quantitative Approaches to the Intellectual History of the Field. A Close Reading". În *Futures of Comparative Literature. ACLA State of the Discipline Report*, 258-266, editat de Ursula K. Heise. London and New York: Routledge, 2017.
- Terian, Andrei. "Big Numbers: A Quantitative Analysis of the Development of the Novel in Romania". *Transylvanian Review*, vol. XXVIII, Supplement 1 (2019): 55-71.
- Ursa, Mihaela. "Is Romanian Culture Ready for the Digital Turn?". *Metacritic Journal for Comparative Studies and Theory*, vol. 1, nr. 1 (2015): 80-97.



PROIECT CO-FINANȚAT DE:

Prezentul articol a fost realizat în cadrul proiectului *ASTRA Data Mining. Muzeul Digital al Romanului Românesc din Secolul al XIX-lea*, organizat de Complexul Național Muzeal ASTRA și co-finanțat de Administrația Fondului Cultural Național. Proiectul nu reprezintă în mod necesar poziția Administrației Fondului Cultural Național. AFCN nu este responsabilă de conținutul proiectului sau de modul în care rezultatele proiectului pot fi folosite. Acestea sunt în întregime responsabilitatea beneficiarului finanțării.

