

MIHAI EMINESCU

**G. COȘBUC, V. ALECSANDRI,
T. ARGHEZI, I. BARBU,
A. PĂUNESCU, G. BACOVIA, R. GYR.**

Comparații statistico-informaționale

TORSAN ILIE

MIHAI EMINESCU

**G. COȘBUC, V. ALECSANDRI,
T. ARGHEZI, I. BARBU,
A. PĂUNESCU, G. BACOVIA, R. GYR.**

Comparații statistico-informaționale



**Editura Universitară
2009**

Tehnoredactare computerizată: Angelica Mălăescu (Badea)
Coperta: Angelica Mălăescu (Badea)

Copyright © 2009
Editura Universitară
Director: Vasile Muscalu
B-dul. N. Bălcescu nr. 27-33,
Sector 1, București
Tel. / Fax: 021 – 315.32.47 / 319.67.27
www.editurauniversitara.ro
e-mail: redactia@editurauniversitara.ro

EDITURĂ RECUNOSCUTĂ DE CONSILIUL NAȚIONAL AL
CERCETĂRII ȘTIINȚIFICE DIN ÎNVĂȚĂMÂNTUL SUPERIOR
(C.N.C.S.I.S.)

Descrierea CIP a Bibliotecii Naționale a României
TORSAN, ILIE

**Mihai Eminescu, G. Coșbuc, V. Alecsandri, T. Arghezi, I.
Barbu, A. Păunescu, G. Bacovia, R. Gyr : comparații
statistico-informaționale / Torsan Ilie. - București : Editura
Universitară, 2009**

Bibliogr.

ISBN 978-973-749-787-1

821.135.1.09

© Toate drepturile asupra acestei lucrări sunt rezervate Editurii Universitare.

Distribuție: tel./fax: 021-315.32.47
021-319.67.27
comenzi@editurauniversitara.ro

ISBN 978-973-749-787-1

O nouă carte despre vechi Poeți

Pendulând cu delicatețe între matematicieni și lingviști, autorul acestei cărți îi aduce și pe unii și pe alții la Poarta Sărutului, după lehamitea ce-i unise pe sceptici la Masa Tăcerii.

Bătând cu curaj la ușile ferecate ale absconselor noțiuni de lingvistică matematică, profesorul Ilie Torsan împacă sobrietatea Teoriei Informațiilor cu ritmurile secrete ale unor poeme și catrene deșirate în elevații și ecuații.

El se ia la trântă cu capcanele imaginației scormonitoare, găsind soluții de grad superior unor teoreme ruginite între foile unui manual academic, pe care unii abia se încumetă să-l răsfoiască, necum să-l înțeleagă, nefiind la îndemîntea multora. Altminteri, cu răbdare monotonă și cu siguranța celui ce nu duce lipsă de idei geniale, autorul acestor calcule trebuincioase doar minților neleneșe, se încumetă să arate cărări și să ghideze cititorii spre miezul unor adevăruri uitate. Șiruri fibonaciene, țesute cu fir roșu, în nuanțe de laitmotiv, urzeli de energii informaționale și entropii calculate cu abacul, ne dezvăluie în persoana autorului un dresor al cifrurilor și un inițiat al cheilor.

Magistru al criptografiei române, autorul a pornit cu mulți ani în urmă un război împotriva ignoranței și arhisuficienței celor ce credeau că Eminescu este un geniu plecticos, epuizat și antimodern. În cartea de față, el propune o ipoteză, pe care, cu migală și înțelepciune, o demonstrează cu sârg și aplecare, meșterind formule pe

ogorul câtorva poeți clasici. Și, ca prin minune, teoria se adevărește: adică este - mai bine zis, devine- o Teoremă, care îi va purta numele: Este „Teorema Torsan cu privire la aplicațiile șirurilor Fibonacci în domeniul Poeziei”.

Aflându-mă în pragul vârstei de 70 de ani, aș îndrăzni o profeție. Cei ce se vor apleca în viitor spre înțelegerea integralității geniului eminescian vor avea nevoie, printre altele, și de studiile deschizătoare de drum ale acestui autor. Pașii profesorului Torsan nu sunt în pustiu. Cărțile sale de analiză criptografică, dedicate lui Eminescu, adică, vreau să zic, cel mai cunoscut „încă necunoscut Poet”, clare, ca orice operă menită să dureze, își adaugă prin prezenta încă o surată, mezină, dar altele, după cum ne încredințează autorul, stau gata să se nască.

Bucuros că i-am încurajat în ultimii 15 ani nobila-i străduință întru Eminescu, nu îmi rămâne decât să îi urez domnului Profesor Ilie Torsan să dezlege tainicele mistere ce mai învăluie, în unele privințe, opera nemuritorului astru al Poeziei române.

I. C. Rogoianu
03.11.2009

Cuvânt înainte

Cercetarea matematică a poeziilor lui Eminescu are în țara noastră o anumită tradiție.

Amintim în acest sens cercetările lui Dimitrie Macrea asupra frecvenței fonemelor în unele poezii ale lui Eminescu, din 1942, ca și cercetările întreprinse în același domeniu de către D. Caracostea, tot în 1942, ambele investigații au loc cu cel puțin doi ani înainte de a apare în lume o carte dedicată metodelor statistice în stilistică.

În perioada următoare au apărut numeroase lucrări interesante din acest domeniu.

Cu titlu de exemplu, reținem în acest sens interesanta cercetare întreprinsă de Luiza Seche, „**Lexicul artistic eminescian în perspective statistice**” în *Studii de limbă literară și filologie*, Ed. Acad. 1969, Tudor Vianu „**Statistica lexicală și o problemă a vocabularului eminescian**”, în *Limba română*, anul VIII, 1959, Solomon Marcus „**Entropie et énergie poétique – avec application à trois poésies de Eminescu**” din *Cahiers de linguistique théorique et appliquée*, vol. 4, 1967, Nicolau E., Sala C., Rocerit A., „**Observații asupra entropiei limbii române**”, în *Studii și cercetări lingvistice*, 1959, 10,1. etc.

În lucrarea, „**Mihai Eminescu. Caracteristici statistice și informaționale**”, Ed. Universitară, 2008, am calculat valorile unor astfel de statistici (frecvențele globale ale literelor; frecvențele literelor ca inițiale și finale de cuvinte, numărul mediu de litere distincte pentru versuri, valorile unui indicator pentru compararea redundanțelor, entropia și energia informațională etc.),

pentru un număr de peste 25 de poezii eminesciene, în vederea comparării lor din aceste puncte de vedere.

În această lucrare vom compara valorile unor caracteristici statistice și informaționale, calculate pentru eșantioane de poezii semnate de **M. Eminescu, V. Alecsandri, G. Coșbuc, T. Arghezi, I. Barbu (Dan Barbilian), G. Bacovia, Radu Gyr și A. Păunescu.**

Diversitatea de stiluri și tematici ale poeziilor analizate, precum și volumul eșantionului folosit, ne va permite să estimăm valorile acestor caracteristici pentru textele poetice din limba română.

Ne vom rezuma la următoarele caracteristici; entropia poeziilor, indicatorul R pentru compararea redundanței versurilor, raportul „consoană – vocală”, distribuția cuvintelor după numărul silabelor componente, distribuția numărului de litere distincte din textele de o anumită lungime.

Referitor la acestea facem următoarele precizări:

În cazul în care literele alfabetului se consideră independente, informația medie care revine unei litere dintr-un text este dată de expresia:

$$H = - \sum_{k=1}^n p(k) \log p(k)$$

unde $p(K)$ este probabilitatea apariției literei cu numărul k , din alfabetul cu n litere, expresie care poartă numele de entropie și a fost introdusă în teoria informației de matematicianul C.E. Shannon.

Pe parcursul lucrării se va observa că, cu cât – în textele analizate – sunt mai numeroase elementele de

organizare, cuvinte rimate numeroase, versurile scurte, numeroase repetiții, entropia are valoarea mai mică.

Indicatorul R pentru compararea redundanțelor este definit astfel:

Să considerăm un text S de lungimea L, având deci L litere, pe care îl vom partiționa după următoarele reguli:

- începând cu poziția „i” vom considera secvențele, S (1, i-1) și S (i, i + k) unde $k = 1, 2, 3, \dots, (L-i)$;
- găsim cel mai mic număr k, pentru care S (i; i+k) nu este un subșir al lui S (1; i-1) și atunci secvența S (i; i+k) este următorul segment în partiționarea lui S, urmând să trecem la poziția (i+k+1) și se reia procesul de mai sus.

O astfel de partiție se numește istorie exhaustivă.

Dacă notăm cu N numărul secvențelor dintr-o istorie exhaustivă a unui text de lungime L, atunci raportul:

$$R = N/L$$

poate fi folosit ca indicator al redundanței. Cea mai mică redundanță o au textele în care nici o literă nu se repetă, și în acest caz indicatorul R are valoarea maximă egală cu 1 (unu). Cu cât redundanța textului este mai mare, R are o valoare mai mică.

De exemplu textul: „Codrul este o mare bogăție”, are următoarea istorie exhaustivă:

C.o.d.r.u.l.e.s.t.e.o.m.a.re.b.o.g.ă.ț.i.e

care conține 19 secvențe, indicatorul R are valoarea,

$$R = 19/22 = 0,863$$

Textul, „Codrule codruțule te ador” are următoarea istorie exhaustivă:

C.o.d.r.u.l.e.codruț.ulet.ea.do.r

având 12 secvențe, indicatorul R are valoarea,

$$R = 12/22 = 0,545$$

valoare mult mai mică decât aceea obținută pentru primul text, lucru firesc, deoarece este evident că al doilea text este mai redundant.

Pentru a compara diferitele valori ale indicatorului R, funcție de lungimea textului, facem următoarele precizări:

Considerăm un alfabet cu 30 de litere, atunci numărul maxim de segmente dintr-o istorie exhaustivă a unui text de 100 de litere poate fi calculat, rezultând în mod evident valoarea maximă corespunzătoare a indicatorului R. Într-adevăr, dacă alfabetul are 30 de litere, atunci primele 30 segmente sunt chiar cele 30 litere. Următoarele 70 de poziții pot fi completate cu 35 de bigrame alese astfel încât toate perechile de litere adiacente să fie distincte. Acest lucru este posibil deoarece cu cele 30 de litere putem forma 900 de perechi, de unde putem alege perechile care îndeplinesc condiția de mai sus. Obținem o istorie exhaustivă cu $30 + 35 = 65$ perechi și deci indicatorul R are valoarea,

$$R = 65/100 = 0,65.$$

Repetând raționamentul de mai sus, pentru diferite lungimi de text, pentru valorile maxime ale lui R obținem valorile:

lungime text	40	50	60	64	70	80	100
R, maxim	0,875	0,8	0,75	0,734	0,714	0,687	0,65

O caracteristică interesantă este aceea referitoare la numărul literelor distincte.

Dacă notăm cu N lungimea în litere a versurilor și cu d (N) numărul literelor distincte ale versurilor de lungime N, pentru toate poeziile analizate există relația,

$$d(N) = A(i) [1 - \exp(-\frac{N}{a})]$$

unde N este lungimea textului, A reprezintă numărul literelor din alfabetul folosit, în cazul nostru A=30, exponențiala reprezintă numărul „e” cu valoarea aproximativă, e=2,718 ... iar A(i) este o constantă multiplicativă specifică fiecărui poet. Dacă eșantioanele de poezii sunt semnificative, putem considera că media numerelor A(i) notată cu A(r), reprezintă numărul specific al textelor poetice din limba română.

În același timp vom analiza distribuția cuvintelor poeziilor după numărul silabelor componente, ajungând la concluzia că aceasta este o distribuție de tip Poisson (sau distribuția evenimentelor rare), adică:

$$P(n) = a^{n-1} e^{-a} / (n-1)!$$

unde P (n) este probabilitatea apariției unui cuvânt format din „n” silabe, m = a + 1 este lungimea medie, în silabe, a unui cuvânt, iar (n-1)! = 1.2.3... (n-1).

Se va observa că numărul „a” este specific fiecăruia dintre poeții analizați.

Pentru fiecare poezie s-a calculat valoarea raportului „consoană-vocală” și lungimea medie a cuvintelor, precum și distribuția substantivelor și a vocalelor.

În încheiere, pentru divertisment, s-au prezentat câteva „întâmplări” interesante.

Ilie Torsan