
Statistici descriptive, entropie și internet

Profesor univ. dr. habil. Gheorghe Săvoiu

Rezumat

Investigarea evoluției limbajului în general și a celui arhivat pe Internet nu constituie o problemă unidisciplinară izolatoare de tip antropologic, etimologic, lingvistic, istoric, fizic, statistic, cibernetic sau informatic ci, cu siguranță, necesită o abordare inter-, trans- și multidisciplinară, incluzând mult mai multe științe în raport chiar și cu cele menționate anterior. Articolul acesta prin dimensiunile restrânse caracteristice genului și prin tematica strict statistică a revistei în care este publicat, rămâne unul interdisciplinar cel mult de ordin II, pornind în demersul introductiv al cercetării de la câteva repere definitorii ale limbajului, Internetului și entropiei, dezvoltând conceptul termodinamic al entropiei plasându-l în universal cuvintelor și rețelelor www, pentru a evolua structural cu ajutorul teoriei sondajului în direcția unei investigații ce inferențiază estimări și oferă parametri aproximativi și cu un anumit nivel de eroare limită ai comunicării pe Inteneti și, mai ales, ai evoluției entropiei în rețea. Cum de altfel este și natural, investigația naște statistici descriptive, matrici și modelări, alături de comentarii și remaci finale al căror grad real de utilitate poate fi evaluat de cititorii articolului în integralitatea acestuia.

Cuvinte cheie: limbaj, Internet, sondaj, eșantion aleator, eroare medie de reprezentativitate, eroare limită, statistici descriptive, entropie, entropie termodinamică, entropie statistică, entropie informațională, entropie economică, matrice de corelație, model econometric.

Clasificare JEL : C10, C13

Introducere

Limbajul, înțeles inițial ca transmitere exclusiv orală a informațiilor are ecouri ancestrale în evoluția condiției umane și a comunicării interumane, definit drept condiția formatoare definitorie a omului și societății sale extinse sub denumirea de populație. Protolimbajul ca strămoș al limbajului în viziunea antropologiei moderne pare legat de Homo Ergaster, hominidul care a trăit în Africa acum aproape două milioane de ani (1,4 – 1,9 milioane de ani), a recurs pentru prima oară la gândirea simbolică și a dispus de artă figurativă sau de imaginație pentru a-și proiecta gândurile pe termene lung (Carrizo, 2018). Argumentul științific al unei astfel de îndelungate evoluții de la protolimbaj la limbaj modern este dat de dimensiunea craniului lui Homo Ergaster, care atingând deja 850 cm³ a permis apariția centrului vorbirii

odată cu zonele Broca și Wernicke capabile a sprijini o formă de comunicare reală lingvistică sau simbolică. O teorie moderată axată pe dovezi genetice, respectiv pe descoperirea relativ recentă a unei gene legate de limbaj și a cărei răspândire a durat între 10 și 20 de mii de ani, plasează apariția limbajului ca motor al evoluției speciei umane acum aproape 200 de mii de ani. Teoria nu poate preciza însă aspecte calitative legate de faptul dacă limbajul era unul primitiv de tip *mimesis* (imitativ) sau unul mai evoluat (personalizat). Varianta maximal reduționistă leagă limbajul și comunicarea nu de apariția lui Homo Sapiens Sapiens (acum circa 50 de mii de ani) ci de maturizarea lui prin dezvoltarea unei structuri sociale complexe a comunităților umane, implicit comunicării intracomunitare.

Limbajul modern a devenit mai mult decât „o metodă de comunicare umană, vorbită ori scrisă, constând în utilizarea cuvintelor într-un mod structurat și convențional” (Oxford English Dictionary, 2020), multiplicându-se permanent în conceptualizări științifice tot mai variate de la cele psihologice ca *metodă non-verbală de exprimare* la cele informatice ca *sistem de simboluri și reguli pentru scrierea de programe sau algoritmi*, detaliindu-se simultan lingvistic și geografic drept *sistem de comunicare utilizat de o anumită țară sau comunitate*, ori structural în calitate de *frazologie și vocabulare specifice anumitor profesii, domenii sau grupuri sociale* etc.

Internet este un termen relativ recent intrat în limbajul general și științific. Acest concept se poate substitui în limbajul modern într-un mod limitat, cu acronimul WorldWideWeb (www sau W3), iar semnificația sa este una legată de accesare directă a informației, prin motoare de căutare, dar în special de păstrare sau arhivare a informațiilor (Săvoiu, Cudanov, 2020). Internetul înseamnă simultan atât “o inițiativă de accesare și regăsire a informațiilor extinse sau hypermedia” (Greene, 2000) cât și o ofertă care “urmărește să asigure informații generale printr-un acces rapid la un univers mare de documente” (Mowbray, 2017). Cu toate că Nikola Tesla, la începutul secolului XX, a fost în realitate primul om de știință care a crezut că „sistemul wireless mondial” este realizabil, imaginând-și cum va arăta această lume virtuală, iar după alte șase decenii, Licklider a devenit acela care a extins idee de rețea de calculatoare la nivel intergalactic, Internetul modern s-a născut oficial abia în 1990, când a fost practic inventat de designerii săi creativi Tim Berners Lee și Robert Cailliau. Acest fapt nu ar fi fost însă posibil, dacă în 1972, ARPAnet (Advanced Research Projects Agency Network - creată și finanțată de Departamentul Apărării al SUA) nu ar fi inventat poșta electronică, iar informaticienii Robert Kahn și Vint Cerf nu ar fi pus la punct planuri pentru primul protocol Internet. Internetul, sugestiv descris ca „*haos organizat*”, a fost folosit mai întâi de către fizicieni iar ulterior și de către

profioniștii din asistența socială, în special cei din mediile academice și asociațiile profesionale, de la universități la companii economice, de la asociații științifice, la organizații dedicate artei și culturii etc. Istoria reală a Internetului leagă într-un mod inedit domenii de comunicare cu domenii lingvistice, domenii statistice cu domenii informatice, sectoare academice, cu sectoare militare, acțiuni de inginerie cu interacțiuni sociale online, teoria educațională cu practică în cercetare. Utilizarea actuală a Internetului a depășit demult intențiile inventatorilor săi, sensul conceptului extinzându-se de la sursa tehnică „*disponibilă printr-un sistem de telecomunicații ce include cel puțin un computer*” (Manovich, 2001) la semnificații mai largi de conectare, accesare, servire cu informații sau chiar la extensia maximă de disponibilitate de comunicare și informare. În final, toate aceste semnificații multiple se reduc la *schimbare și adaptare* cu accent pe limbaj, alături de *comunicare și informație*, deoarece Internetul facilitează și redimensionează mai bine volumul și importanța informației, dar și a sistemului informational, a limbajului și chiar a comunicării (Săvoiu, Cudanov, 2020).

Asocierea între Internet și orice alt sistem termodinamic induce firesc nevoia de a identifica dimensiuni statistice legate de valori maxime, minime, medii, modale sau mediane, de dispersie și abatere standard, de asimetrie și boltire, de normalitate sau anormalitate evolutivă a limbajului etc. Propagarea și schimbarea limbajului Internetului deține o dinamică investigativă cel puțin interesantă, care are șanse ipotetice să identifice altceva decât eterogenitate și anormalitate în rețeaua potențial detectabilă a unor miliarde de cuvinte, aflată într-o continuă dilatare cauzată de creșterea numărului de pagini web, site-uri etc. Variabila numărului de cuvinte are o dimensiune și o dinamică doar aparent prestabilă și o semnificație dominant eclectică fiind dependentă de mulți factori ca limbă, dimensiuni cuantificate ale cuvintelor, sinonimie și omonimie, spațializare prin regionalizare împreună cu temporalizare dată de de profunzimea arhivelor, precum și din alte cauze deloc semantice sau gramaticale, ci mai curând legate de abreviere și acronimizare etc.

Volumul în cuvinte al Internetului ar putea fi analizat și interpretat atât ca fluxuri cât și ca stocuri statistice urmărind cel puțin patru tipuri de evaluări distincte, axate pe: i) valori maxie și minime, mediane și medii ale aproximărilor; ii) valori ale împrăștierii sau dispersiei statistice a aceluiași aproximărilor; iii) asocieri sau corelații majore ale valorilor specifice statisticilor descriptive ale aproximărilor; iv) estimări și modelări cu instrumente de investigație statistică, recunoscute și adecvatecum sunt sondajul statistic și modelul econometric validat parțial sau integral, asociind metode statistice investigații tiparului variabil al unei pagini web. Dimensiunea Internetului pornește astfel de la apariția și evoluția unui sens prototipic

sau standardizabil al formei și conținutului paginii web și ajunge în final la rețeaua acestora. Ca urmare, cercetarea din acest articol investighează statistic dimensiunea Internetului ca sursă, rezervor, corpus, sistem sau ansamblu statistic valorificând un eșantion de cuvinte prelevat aleator aleator în limba română și, la final, încearcă să precizeze relativ sau să estimeze parametrii rețelei www, cu un nivel de eroare acceptabil. Articolul urmărește în paralel și evoluția menționărilor unui eșantion de cuvinte prelevat aleator, apelând la motorul de căutare Google Search considerat cel mai important ca dimensiuni de accesare. Cronologia investigației pornește de la ipoteza că o cercetare poate fi considerată irelevantă pentru intervale mici de timp, iar din acest motiv se încearcă în paralel o determinarea estimativă a entropiei eșantionului cuvintelor selectate, pornind de la ipoteza simplă a creșterii acesteia în orice sistem expus trecerii timpului. Câteva remarci finale închid articolul iar această succintă prezentare introductivă (reminder) remodelează raționamente lingvistice cu ajutorul gândirii statistice, care, împreună evoluează în direcțional interdisciplinarității statisticii lingvistice.

Metodologie, date, rezultate și discuții

• Despre entropia statistică sau entropia sistemelor descrise statistic și ansamblurilor izolate

În orice tentativă mai mult sau mai puțin reușită de definire sau reconceptualizare, „*subiectul sau definitul este egal „prin definiție” cu predicatul sau definatorul*” (Săvoiu, 2009). Univocitatea ca semnificație neschimbată în timp și spațiu a termenului devine astfel dezideratul unei definiții precise și în cazul entropiei, iar ambiguitatea și imprecizia constituie expresii ale dezechilibrului și inegalității de conținut materializate în propoziția sau afirmația oricărei definiții. „*În esență, criteriile necesare și suficiente definiției sunt legate de prezența sinonimiei sau echivalenței sensurilor, dar și de absența ambiguității. Subiectul sau termenul bine definit devine în final sinonim lingvistic, echivalent logic și necontradictoriu sau lipsit de ambiguitate, în raport cu predicatul sau definatorul*” (Săvoiu, 2009). Cea mai simplă dintre definițiile entropiei, păstrând totuși precizia specifică fizicii și o localizare în rețeaua www sau un plasament în Internet, este aceea de „*măsură a gradului de indisponibilitate al unui sistem www de a produce ordine*” sau gândind mai aproape de limbaj de „*măsura a gradului de dezordine sau haos a limbajului și Internetului ca sisteme izolate*”.

Cu cât un sistem este mai apt să producă ordine, cu atât este mai mică entropia acestuia, limita fiind dată în timp de faptul că sistemul atinge pragul său final atunci când nu mai poate produce ordine și generează exclusiv haos, moment ce consemnează entropia maximă a acestuia. Problematika majoră

unei definiții de acest fel este aceea că definitorul (lat. definiens) este menit să faciliteze cunoașterea și să precizeze înțelesul entropiei, deși termenul entropie oferă natural accepțiuni consistente diferite (Săvoiu, 2009). Definiția științifică privită ca un ansamblu identifică în final o *convenție* a folosirii limbajului și configurează conceptul entropiei ca măsură a lipsei ordinii și acest lucru face ca semnificația acesteia să fie mai greu de înțeles. Clasicii încercau să definească esența fenomenului, prin similitudini (genul proxim) sau prin diferență, dar nu prin absență. Teoria științifică modernă și limbajul acesteia au renunțat practic la orice restricționare a definițiilor. Din acest motiv, acest articol propune ca *entropia asimilată limbajului și sistemului Internetului* să fie definită lingvistic și informațional prin *dezordonarea sau haotizarea sistemului* limbajului plasat pe www, cu ajutorul determinărilor statistice (probabilistice).

În calitate de autori sau cititori, emitenți și respondenți ai unor mesaje, oamenii anunță, inițiază și răspund unui dialog scris sau comunică pur și simplu pe Internet, distribuie informații cu ajutorul cuvintelor, dilatând permanent numărul cuvintelor folosite sau volumul limbajului comun. Informațiile curg în mod firesc *de la cei care știu mai mult, către aceia care știu mai puțin*, așa cum în termodinamică și căldura circulă de la corpul fierbinte care o pierde astfel prin contact cu corpul mai rece care se încălzește. Cu cât oamenii sunt mai intens plasați în procese de comunicare informațională, structurați în emitenți și receptori de mesaje, apreciază și comentează, modifică și redistribuie mesaje (texte) pe Internet, cu atât fiecare deține la final mai multe date decriptate pentru discuții ulterioare, amplificând informațiile primite în memoria personală și a Internetului, formând noi grupuri de comunicare. Prin comunicare permanent sau intermitentă, informațiile disponibile în Internetul astfel redefinit ca univers multidisciplinar al cuvintelor sau limbajului științific sau ca multivers al unor universuri de limbaje unidisciplinare, se dezvoltă în manieră ascendentă și exclusiv continuă și dețin astfel o șansă tot mai mare de a deveni mai încurcate, mai dezordonate, mai haotice etc., confirmând cu trecerea timpului creșterea entropiei, atât în cazul limbajului, cât și al rețelei (www). **Entropia** a fost concepută și poate fi redefinită ca o *reflexie cantitativă a evoluției gradului de asimetrie sau ireversibilitate al oricărui proces informațional*. Entropia dobândește o identitate tot mai precisă, reflectând o săgeată a timpului și caracterizând comportamentul oricărui sistem de comunicare inclusiv a Internetului. Această dinamică multiplicativă și disipativă a cuvintelor și a limbajului, numai într-un singur sens, respectiv în sensul unei dilatări existente oriunde în rețea și în mod continuu, descrie o inexorabilitate termodinamică de tip evolutiv: chiar și Internetul trece printr-un proces natural de maturizare și în final de îmbătrânire. Acest lucru impune a adapta logic și Internetului toate principiile termodinamicii (Săvoiu, 2021):

i) principiul “*echivalenței dintre formele schimbului de energie între sistemul Internetului, definit de energia internă informațională a limbajului propriu ca o funcție de stare și limbajul existent în lumea înconjurătoare*” (principiul inexistenței unui perpetuum mobile de speța I);

ii) principiul conform căruia “*transformările spontane de energie se realizează de la potențialul mai înalt spre potențialul mai scăzut*” (principiul inexistenței unui perpetuum mobile de speța II), caz în care entropia Internetului ca mărime de stare este aditivă în subsisteme sau rețele substructurate și mai mare ca zero pentru transformări ireversibile (entropia Internetului neputând să tindă spre zero sau chiar neputând să scadă în afara unui transfer de energie din afara sistemului Internetului iar alternativa Arnold Sommerfeld nefiind practic posibilă);

iii) principiul după care “*atunci când temperatura (informația) tinde către zero absolut, entropia unui sistem (Internet) devine nulă*” sau al ipotezei sistemului cu entropia nemodificabilă cunoscut ca alternativa Walther Nernst și Max Plank.

În cazul Internetului, principiul al doilea al termodinamicii conține cea mai mare aplicabilitate dintre toate cele trei principii de bază, constituind prin simetrie logică cel mai important principiu pentru acest articol. Astfel, afirmația după care în orice sistem izolat (închis), **entropia termodinamică a sistemului crește**, se transformă în rețeaua www în *entropia Internetului crește cu certitudine conform dinamicii volumului limbajului* așa cum sublinia Josiah Willard Gibbs (1902): “*Într-o transformare oarecare a unui sistem fizic izolat, entropia acestuia crește întotdeauna către valoarea maximă ce corespunde unei stări de echilibru termodinamic. Entropia rămâne constantă numai în cazul limită al transformărilor reversibile*”. Măsurarea efectivă a entropiei sistemelor a început cu Rudolf Julius Emanuel Clausius, care a extins legile termodinamicii la un univers infinit, ajungând la concluzia că “*energia lumii rămâne constantă, entropia lumii tinde spre maximum*”, spre un moment de echilibru în care întregul univers ar intra în “*moarte termică*”. Conceptul de **entropie absolută** (S_{abs}) se particularizează prin valoarea zero, conform celei de-a treia legi a termodinamicii, definind de fapt entropia la zero absolut, în timp ce variația entropiei absolute se notează cu ΔS și descrie distanța dintre oricare S și S_{abs} . Ludwig Eduard Boltzmann a relativizat moartea termică de tip clausian și a emis “*ipoteza fluctuațiilor*”, stabilind natura statistică a entropiei (1780).

Entropia statistică este o noțiune esențială pentru macrostarea unui sistem macroscopic, în raport cu care se notează microstările ce definesc un ansamblu statistic format dintr-un număr mare de sisteme identice, fiecare având aceleași probabilități w_i de a se afla în microstarea I , precum și condiția

lor de normare. Entropia statistică a lui Ludwig Eduard Boltzmann constituie astfel o măsură a numărului de stări microscopice posibile ale unui sistem în echilibru, în concordanță cu proprietățile sale macroscopice. Într-un sistem descris de variabile multiple, unde se identifică un anumit număr de configurații (Ω), ipotetic la fel de probabile statistic, entropia (S) este dată de logaritmul natural al numărului de configurații, multiplicat cu constanta Boltzmann ($1,3806504 \times 10^{-23} \text{J/K}$ sau $8,617343 \times 10^{-5} \text{eV K}^{-1}$):

$$S = -k_B \times \ln \Omega \quad (1)$$

Pentru ca entropia unui sistem (limbaj sau Internet) să scadă ar trebui ca acestuia să i se transfere energie din afara sistemului. Primul care a tratat similar și entropia și informațiile a fost Léon Brillouin urmat de Josiah Willard Gibbs, la care entropia devine cantitatea maximă de informații din sistem:

$$S = -k_B \sum_{i=1}^n p_i \times \ln(p_i) \quad (2)$$

unde k_B este aceeași constantă Boltzmann, p_i este probabilitatea ca o energie E_i să apară în timpul fluctuațiilor sistemului (operatorul Σ este cuantificat pentru toate stările posibile ale sistemului).

Entropia informațională reprezintă măsura gradului de uniformitate sau de diversitate a distribuției de probabilitate și a fost expusă de Claude Elwood Shannon și Warren Weaver în *The Mathematical Theory of Communication* (1949), transformată într-o entropie condiționată a două evenimente X și Y luând valori x_i și y_j , când își pierde nu numai semnificația termodinamică inițială, dar și povara istorică a unităților de măsură (J/K), fiind redefinită de relația:

$$S(X|Y) = - \sum_{i=1}^n p(x_i|y_j) \times \log \frac{p(x_i|y_j)}{p(y_j)} \quad (3)$$

unde $p(x_i | y_j)$ este probabilitatea condiționată ca $X = x_i$ și $Y = y_j$. Conform relației Shannon, cu notața entropiei modificată ușor față de formula originală din $H(X)$ în S , aceasta se poate determina și în cazul unei variabile discrete X cu valorile $\{x_1, \dots, x_n\}$ și cu funcția de probabilitate $p(x_i)$:

$$S = H(X) = - \sum_{i=1}^n p(x_i) \times \log_b p(x_i) \quad (4)$$

în care b este o bază pentru logaritmi, reală și supraunitară de tipul e , 2, 10 etc. (pentru $b = 2$ unitatea de măsură a informației devenind celebrul bit). Entropia informațională indică practic volumul de informație existent într-un cuvânt/mesaj în biți (lungimea minimă) sau în biți per simbol, constituind limita optimă a unei compresii fără pierderi aplicabilă unui/mesaj comunicat

(entropia Shannon în biți per simbol × număr de simboluri din cuvânt/mesaj). **Entropia economică** este descrisă ca definitorie pentru orice proces economic. Acest tip de entropie “domină toate transformările materiale într-o economie, dezvoltându-se într-un mod ireversibil în direcția epuizării resurselor” (Georgescu-Roegen, 1971), constituind varianta teoretică (bio)economică a entropiei aptă să descrie măsura de degradare continuă și iremediabilă a unui potențial de resurse existente (apariția dezorganizării sau haosului).

$$S = \sum_{i=1}^n w_i \times \ln(f_i), \text{ cu } f_i = N_i/N \quad (5)$$

unde w_i = coeficient de ponderare (agregare).

O variantă de calcul simplificată în logica lui Octav Onicescu devine energie informațională:

$$E_i = \sum_{i=1}^n (N_i/N)^2 = \sum_{i=1}^n (f_i)^2, \text{ with } f_i = N_i/N \quad (6)$$

unde un receptor așteaptă un cuvânt/mesaj N_i din N cuvinte/mesaje posibile, cu probabilitate evident diferită de apariție (Săvoiu, 2013).

În final, entropia informațională Shannon ca măsură a incertitudinii asociată unei variabile aleatoare determină așadar cantitatea de informație necesară pentru a reprezenta toate stările variabilei a fost aceea selectată în acest articol ca fiind cea mai adecvată soluție în investigarea limbajului și Internetului.

• **Un eșantion reprezentativ pentru dimensionarea limbajului activ al Internetului**

Autorul articolului a luat în calcul o valoare aproximativă a erorii limită probabile (Δ_{ma}), extrasă din cercetări anterioare și a selectat la final un nivel de eroare corespunzător unui prag de semnificație $\alpha = 0,01$ și unei abateri standard de 2091593 mențiuni (Săvoiu, 2021). Eroarea limită probabilă (Δ_{ma}) a căutărilor Google pe Internet ($\alpha = 0,01$ și $z = 2,58$) este determinată după relația:

$$\Delta_{ma} = z \times \sigma_{ma} = z \times \sqrt{\frac{\sigma^2}{n-1}} = 525000 \text{ mențiuni} \quad (7)$$

și numărul de cuvinte sau volumul unui eșantion extras aleator rețut rezultat la final este $n = 106$:

$$n = [z \times \sigma]^2 : [\Delta_{ma}]^2 = \{[z \times \sigma] : [\Delta_{ma}]\}^2 = \{[2.58 \times 2091593] : [525000]\}^2 = (10.278686)^2 = 105,65 \text{ cuvinte} \quad (8)$$

Determinările pentru un sondaj simplu aleator nerepetat (SSANR) nu aduc modificări semnificative, deoarece corecția finită de populație este din punct de vedere dimensional foarte mică în condiții concrete ($n = 106$ cuvinte în eșantion și $N = 65920$ cuvinte în dicționarul explicativ - DEX, 2012), impunând schimbări abia la nivelul zecimalei a patra:

$$\sqrt{\left[1 - \frac{n}{N}\right]} = 0,9991957. \text{ Acest fapt nu influențează practic volumul noului}$$

eșantion de tip SSANR (pragul de semnificație $\alpha = 0,01$ și $z = 2,58$):

$$n = \left[\frac{z^2 \sigma^2}{(\Delta_{ma})^2 + \frac{z^2 \sigma^2}{N}} \right] = \{(2,58^2 \times 2091593^2) : [(525000)^2 + (2,58^2 \times 2091593^2) : (65290)]\} = 105,5 = 106 \text{ cuvinte} \quad (9)$$

Entropia statistică a lui Ludwig Eduard Boltzmann conduce la un raționament în care echilibrul corespunde unei condiții interne *dinamice* din punct de vedere fizic, dar stabilizate statistic în modul în care particulele (cuvintele în cazul acestui articol) își ocupă pozițiile în volumul disponibil (limbajul și Internetul), fără ca acest lucru să însemne configurații ale particulelor rămase neschimbate în timp (menționarea cuvintelor cu rezultate asemănătoare). Ocuparea spațiului în timp cea mai probabilă a fost aceea pentru care configurațiile particulelor (cuvintelor) au fost mai distribuite iar energia internă a sistemului izolat (limbajul sau Internetul) a devenit practic suma energiilor cinetice ale constituenților în condiții de mișcare haotică (comunicare reală informațională, de la cei ce știu către cei ce nu știu).

Pentru sistemele reale, în mecanica de tip cuantic și ulterior în fizica cuantică, numărul de particule este unul colosal (10^{23}). Dacă într-o singură picătură de apă există aproximativ 10211021 molecule, iar în dicționarul limbii române 65920 de cuvinte și în cel al limbii engleze 600 000 de cuvinte, extins de investigațiile universității Harvard și ale companiei Google deja din 2010 la 1022000 de cuvinte, atunci estimările cuvintelor plasate pe Internet în limba română de aproximativ 1000 de miliarde, în limba engleză de aproape 15000 de miliarde și în limbajul total al Internetului de circa 30 000 de miliarde de cuvinte (Săvoiu, 2021), se confirmă similitudinea investigațională din articol conform căreia limbajul într-o comunicare modernă se va apropia în curând de dimesiunile unui ansamblu statistic specific fizicii statistice (10^{23}).

Analogia este justificată și pornind de la raționamentul lui Ludwig Eduard Boltzmann, în care echilibrul nu este altceva decât staționaritatea probabilității distribuției microstărilor (menționilor cuvintelor investigate cu Google search), unde haosul crește firesc (entropia), iar particulele

agitate (cuvintele) ocupă în timp toate pozițiile (probabilitățile calculate ale cuvintelor) care le sunt permise. În ipoteza ergodică boltzmanniană, distribuția statistică a microstărilor sau structura probabilistică internă depinde doar de numărul de microstări pe care sistemul le poate include, fapt care justifică prelevarea în acest articol a unui eșantion ($n=106$) și cuantificarea entropiei parțiale. După cum ordinea unui ansamblu fizic evoluează întotdeauna către dezordine, tot așa comunicarea axată pe cuvinte și limbaj, beneficiază de ipoteza asumată a existenței entropiei, estimată cantitativ sau inferențiată (Tabel nr. 1) cu valorificarea logaritmului natural ($\ln(x_i)$ sau $\log_e(x_i)$), unde x_i reprezintă frecvența sau numărul de mențiuni în urma căutărilor pe Internet ale cuvântului prelevat în eșantion în baza e . Valoarea matematică a numărului transcendent și irațional e poate fi aproximată cu 20 de zecimale la $e = 2,71828182845904523536$, deși aceasta deține în realitate un număr infinit de zecimale și cuantifică o constantă a matematicianului John Napier (1550-1617), care descrie funcția exponențială cu proprietatea de a avea valoarea derivatei $F(x) = e^x$ întotdeauna egală cu 1, în punctul $x = 0$. Inversa acestei funcții a fost transpusă de același John Napier în logaritmi naturali (logaritmi în baza e). În opinia autorului, caracterul maxim abstractizat al acestor logaritmi le conferă o adecvare firească la calculul entropiei (tabel nr.1).

Eșantionul extras aleator de 106 cuvinte în română de pe Internet și calculul entropiei

Tabel no. 1

Pag.	Coloana 1	Coloana 2	Cuvant extras aleator	Numar de litere	Mențiuni Google pe Internet		Valori determinate* per cuvânt în	
					21.01.2021	15.03.2021	21.01.2021 $P(x_i) \times \ln P(x_i)$	15.03.2021 $P(x_i) \times \ln P(x_i)$
50	4		antispumant	11	13400	15800	-2.05465E-05	-2.37418E-05
372		39	exponent	8	23700000	24900000	-0.016040637	-0.016594925
488	50		ieपुरoi	7	33700	33500	-4.81131E-05	-4.74794E-05
1109		60	tâmpană	7	3	3	-6.57817E-09	-7.42759E-09
251		44	criticism	9	237000000	225000000	-0.097901496	-0.093704606
92		49	bavură	6	36800	41300	-5.21685E-05	-5.75525E-05
1038		43	spectacol	9	5560000	5390000	-0.004686447	-0.004529025
18	73		afet	4	12200000	13100000	-0.00918511	-0.009686205
672	20		monosepal	9	3890	4200	-6.51502E-06	-6.94311E-06
759	11		ovreică	7	4490	4670	-7.4473E-06	-7.66382E-06
277	12		decăt	5	53700000	55200000	-0.031314354	-0.031797899
118	46		bregmă	6	481	497	-9.20948E-07	-9.4207E-07
487		8	idiotism	8	141000	120000	-0.00017819	-0.000152686
160		35	căruțaș	7	58700	44200	-8.0074E-05	-6.1253E-05
861		50	prăvălie	8	142000	153000	-0.00017934	-0.000190453
559		1	învălui	7	94500	207000	-0.000123757	-0.000250565
1195	63		veșteji	7	36400	37600	-5.16472E-05	-5.27972E-05
503	70		individualizare	15	119000	123000	-0.0001527	-0.000156158
676		37	moșneguț	8	10300	13600	-1.61032E-05	-2.06676E-05
172	53		cerdac	6	129000	120000	-0.00016434	-0.000152686

870		14	prestigiu	9	826000	1180000	-0.000876619	-0.001195084
38	55		amplixiune	10	526	692	-1.00086E-06	-1.28568E-06
541	3		încleștare	10	55800	55600	-7.64425E-05	-7.56024E-05
964	13		rotocol	7	525000	528000	-0.000584425	-0.000582973
29	64		alicuantă	9	700	760	-1.31037E-06	-1.40393E-06
433		59	geobotanic	10	1850000	497000	-0.001792512	-0.00055216
98	18		băzai	5	769000	784000	-0.000822425	-0.000830427
928		10	receptioner	11	526000	693000	-0.000585424	-0.000743749
489	40		iezuitic	8	2170	2130	-3.77914E-06	-3.6854E-06
263		63	curpen	6	293000	263000	-0.000345738	-0.000311199
976	37		sarcosporidioză	15	577	580	-1.09277E-06	-1.08922E-06
1159	27		țăcănit	7	68600	65900	-9.23552E-05	-8.83358E-05
1052		42	stigmă	6	56900000	56200000	-0.032803155	-0.032259353
684		47	muslim	6	54000000	57500000	-0.17213221	-0.178194773
329		57	drag	4	784000000	817000000	-0.216430252	-0.220597924
399	21		fixism	6	25800	24700	-3.76231E-05	-3.58621E-05
497	7		imunotransfuzie	15	281	303	-5.55492E-07	-5.9137E-07
193		61	clasicism	9	76800	78700	-0.000102401	-0.000103907
1005	69		senar	5	14700000	13000000	-0.010753442	-0.009623578
262		76	curea	5	8940000	10700000	-0.007049081	-0.008157555
608		72	machtetist	9	19400	18200	-2.89264E-05	-2.70559E-05
1065	23		subîntelege	11	238000	195000	-0.000286506	-0.000237362
1043	39		sprânceană	10	3460000	2490000	-0.003104364	-0.002310649
378	74		faclă	5	382000	774000	-0.000439151	-0.000820963
779		25	pas	3	2990000000	2920000000	-0.36698091	-0.366031729
40	2		analitic	8	1240000	1100000	-0.001258289	-0.001122832
935		59	regal	5	230000000	235000000	-0.095799688	-0.096708665
897	29		pumnișor	8	2020	1890	-3.53559E-06	-3.2958E-06
1095		20	știre	5	5760000	5556000	-0.00483171	-0.031964258
1082		9	sfadă	5	114000	129000	-0.000146845	-0.000163078
714	16		neurolimfă	10	387	171	-7.501E-07	-3.44853E-07
579	66		laterit	7	638000	677000	-0.000695971	-0.000728373
798		41	penetrație	10	2970000	3640000	-0.002716676	-0.003220848
735		28	odolean	7	129000	139000	-0.00016434	-0.000174541
23		15	ahtiat	6	32300	36800	-4.6272E-05	-5.17638E-05
442	75		glorie	6	8700000	8920000	-0.006886961	-0.006984828
1068		38	succesor	8	499000	564000	-0.000558385	-0.000618496
983	45		sângerare	9	495000	583000	-0.000554365	-0.000637138
225	36		constituție	11	1290000	1310000	-0.001303185	-0.001311197
353		67	endometrită	11	34600	52900	-4.92938E-05	-7.22301E-05
145		56	cardanic	8	407000	458000	-0.000464937	-0.000513083
663	53		moale	5	7850000	9050000	-0.006024946	-0.007071753
155		33	căciulă	7	2650000	2800000	-0.002458572	-0.002561007
298		49	desfăta	7	120000	131000	-0.000153868	-0.000165377
479	63		homeoterm	9	13000	13300	-1.99785E-05	-2.02454E-05
829	32		platelaj	8	3530	3200	-5.95174E-06	-5.38882E-06
845		40	politicesc	10	4020	3940	-6.71777E-06	-6.5419E-06
216	6		concentra	9	79700000	39900000	-0.042871346	-0.02445522
598	12		loc	3	258000000	313000000	-0.104067462	-0.118619035
485	76		iaz	3	3620000	3300000	-0.003229175	-0.002956751
252		61	cromocistoscopie	16	4500	4690	-7.46213E-06	-7.69437E-06
25		28	albie	5	3570000	3300000	-0.00319026	-0.002956751
636		72	meditație	9	5690000	5580000	-0.00478096	-0.00466672
923		57	răuri	5	1730000	1420000	-0.001689529	-0.001408294
702	17		nefresc	8	190000	147000	-0.000233625	-0.000183652
273		9	dârmoz	6	9480000	9620000	-0.007411182	-0.007450423
1000		4	seimon	6	648000	393000	-0.000705724	-0.000447097
637	75		megafon	7	15300000	17000000	-0.011122251	-0.012066736
962	7		rontgen	7	8240000	7210000	-0.006574092	-0.005820082
1193		73	vermis	6	52500000	48200000	-0.030750491	-0.028507874
237	67		corijent	8	7710	6610	-1.23103E-05	-1.05867E-05

836		21	pluti	5	960000	842000	-0.0010023	-0.000885037
639	74		melodramă	9	19400000	21300000	-0.013575167	-0.014573424
1099	60		tachina	7	560000	572000	-0.000619247	-0.000626354
907		1	radioastronomie	15	690000	646000	-0.000752304	-0.00069846
21		34	agoniza	7	3370000	4490000	-0.003033788	-0.003865949
10	3		act	3	2820000000	2880000000	-0.365022909	-0.365529208
916	10		răgușit	7	93500	134000	-0.000122561	-0.00016882
325		2	documentație	12	8280000	8680000	-0.006601412	-0.006823782
1024	41		sistem	6	372000000	324000000	-0.134458975	-0.121516765
1066	5		subordonat	10	360000	422000	-0.000416306	-0.000476677
1113	35		teleancheta	11	133	87	-2.7363E-07	-1.82129E-07
871		56	preursi	7	996	1940	-1.82402E-06	-3.37724E-06
979	37		sălbatic	8	3030000	3080000	-0.002764617	-0.002783768
373		70	extenuant	9	108000	144000	-0.000139785	-0.000180241
949	66		retrimiteri	11	153000	128000	-0.000191925	-0.000161927
346	36		electricitate	13	4350000	4770000	-0.003788835	-0.004074261
99	11		belcanto	8	5540000	5780000	-0.004671876	-0.00481087
1218		55	zărghală	9	959	1040	-1.7595E-06	-1.88412E-06
641		30	mercur	6	4330000	4500000	-0.0037737	-0.003873422
864		27	predecedat	10	9850	10400	-1.12909E-05	-1.61215E-05
653		23	mijloc	6	8280000	9080000	-0.006601412	-0.007091783
772	51		parafrazare	11	27200	11600	-3.95006E-05	-1.78378E-05
572	44		kilometru	9	2680000	2780000	-0.00248295	-0.002544978
1107		18	tălgerei	8	16900	15900	-2.54642E-05	-2.38807E-05
14	43		adjunct	7	38200000	40300000	-0.023757872	-0.024654729
Total = 106 cuvinte		Total = 837 (7,9 litere/cuvânt)			8730721923	8805003403	-1.890196788	-1.909348493
					Media = 82365301	Media = 83066070	Entropie = 1.890196788	Entropie = 1.909348493

Sursa: Realizat de autor.

*Notă: Determinarea valorii probabilității $p(x_i)$ este de natură frecvențială $[p(x_i) = f(x_i) : \Sigma f(x_i)]$

Entropia mențiunilor cuvintelor eșantionului, evaluată în timp, la 21.01.2021 respectiv 15.03.2021 evoluează firesc ascendant, de la **1.890196788** la **1.909348493** prin determinările de tip informational ($S = H(X) = - \sum_{i=1}^n p(x_i) \times \log_b p(x_i)$, unde $b = e$ și neluând în calcul și constanta Boltzmann în fenomenele statisticii lingvistice. Accentul acestei cercetări este plasat pe evoluția (dinamica) proceselor de cuantificare a entropiei și nu pe nivelul efectiv al entropiei. Procesul permanent al creșterii entropiei atestă astfel scurgerea timpului în sistemul aparent izolat al Internetului, în ceea ce privește menționarea cuvintelor în limbajul comunicărilor în rețea.

Entropia se redefiniște astfel și constituie în statistica lingvistică gradul de dezordine al comunicării în univers, care crește continuu și în mod natural, așa cum arată de altfel și determinările efective. Comunicarea modernă multiplicată prin limbajul sau rețeaua Internetului, descrie un sistem care evoluează ireversibil și urmează săgeata timpului, adică valoarea entropiei crește întotdeauna, printr-o curgere informațională continuă de la cei care știu către cei care nu știu (uneori cei din urmă nu știu în modul atât de plastic exprimat de Socrate, conform căruia cu cât aceștia știu mai mult cu atât își dau

seama cât de multe nu știu sau că de fapt ceea nu știu crește mult mai rapid decât ceea ce știu).

Analiza statistică descriptivă și econometrică identifică în paralel alte câteva aspecte importante, legate de asocierea mențiunilor cu dimensiunile cuvintelor (tabel nr. 2) și mai ales câteva modele cu intensități corelative performante între variabilele utilizate (tabel nr. 3 și 5) precum și modele cu dinamici și intensități diferențiate între cele două prelevări sau extrageri (tabel nr. 4 și 6-8).

Seriile de date inițiale (SER02 și SER03) descrise în tabel nr. 2 sunt caracterizate de o anormalitate distribuțională de tip accentuat, însoțită de o eterogenitate ridicată și o asimetrie foarte pronunțată, în timp ce prin logaritizarea mențiunilor se normalizează, omogenizează și simetrizează întreaga bază de date (SER04 și SER05):

Statistici descriptive pentru primele trei serii de bază și cele derivate prin logarimare

Tabel nr. 2

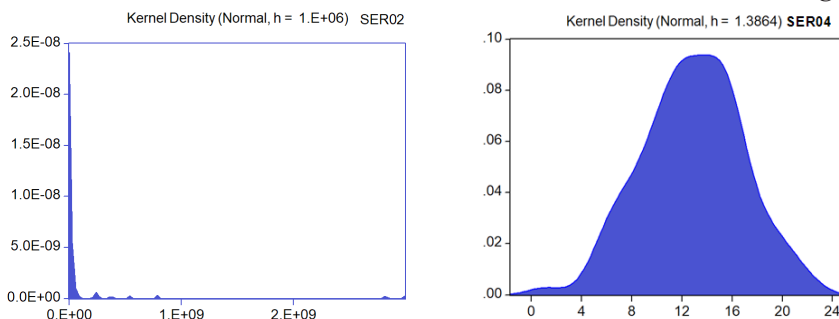
Sample: 1 106	Dimensiunea cuvintelor	Numărul de mențiuni			
Indicatori	SER01	SER02 21.03.21	SER03 15.03.21	SER04 log(SER02)	SER05 log(SER03)
Mean	7.896226	82365301	83066070	12.85743	12.88265
Median	8.000000	497000.0	512500.0	13.11634	13.14660
Maximum	16.00000	2.99E+09	2.92E+09	21.81854	21.79485
Minimum	3.000000	3.000000	3.000000	1.098612	1.098612
Std. Dev.	2.658067	4.07E+08	4.07E+08	3.959881	3.983315
Skewness	0.772488	6.451298	6.405496	-0.138605	-0.156136
Kurtosis	3.892969	44.72987	44.16034	2.907912	2.908925
Jarque-Bera	14.06419	8426.377	8207.468	0.376857	0.467321
Probability	0.000883	0.000000	0.000000	0.828260	0.791630
Sum	837.0000	8.73E+09	8.81E+09	1362.888	1365.560
Sum Sq. Dev.	741.8585	1.74E+19	1.74E+19	1646.469	1666.014

Sursa: Realizat de autor cu ajutorul pachetului de programe EViews

Distribuțiile de tip Kernel relevă clar în fig. nr. 1 procesul de normalizare statistică prin apelul la funcția exponențială și logaritizarea specifică derivată, așa cum se poate constata din confruntarea a două grafice care expun seria inițială (SER02) și aceeași serie logaritmată ulterior (SER04):

Confruntarea grafică a distribuției mențiunilor și logaritmilor acestor valori

Fig nr. 1



SER02 = Number of mentions (21.03.21)

SER04 = log (SER02)

Sursa: Realizat de autor cu ajutorului pachetului de programe EViews

Asocierile potențiale sau anticipate sunt identificate cu ajutorul unei matrice de corelație (tabel nr. 3) și descriu o legătură statistică inversă ce trebuie testate statistic, între dimensiunile și mențiunile cuvintelor în comunicarea și limbajul Internetului. Se remarcă ascendența intensității corelației, care crește odată cu entropia sau cu săgeata timpului.

Matrice de corelație

Tabel nr. 3

	SER01	SER02	SER03	SER04	SER05
SER01	1.000000	-0.311137	-0.315902	0.364028	0.373932
SER02	-0.311137	1.000000	0.999415	-0.938801	-0.936403
SER03	-0.315902	0.999415	1.000000	-0.940829	-0.940017
SER04	0.364028	-0.938801	-0.940829	1.000000	0.997897
SER05	0.373932	-0.936403	-0.940017	0.997897	1.000000

Sursa: Realizat de autor cu ajutorului pachetului de programe EViews

Testul F validează corelația în ambele momente de timp cu accente de intensitate mărită simultan cu creșterea entropiei, apelând la două modele econometrice similar descriind variația mențiunilor în seria logaritmă în raport cu dimensiunile în litere ale cuvintelor, modele realizate în tabel nr. 4, unde, în stânga sunt modelate datele din 21.03.21 și în dreapta cele din 15.03.21.

Confruntarea în timp a modelelor mențiunilor în funcție de dimensiunile cuvintelor în Internet

Tabel nr. 4

Dependent Variable: SER04 Method: Least Squares				
Sample: 1 106 Included observations: 106				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.081595	0.016871	-4.836329	0.0000
SER01	0.008075	0.002026	3.985851	0.0001
R-squared	0.132517	Mean dependent var		-0.017832
Adjusted R-squared	0.124175	S.D. dependent var		0.058963
S.E. of regression	0.055181	Akaike info criterion		-2.937723
Sum squared resid	0.316669	Schwarz criterion		-2.887469
Log likelihood	157.6993	F-statistic		15.88701
Durbin-Watson stat	2.094178	Prob(F-statistic)		0.000125

Dependent Variable: SER05 Method: Least Squares				
Sample: 1 106 Included observations: 106				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.083688	0.016846	-4.967945	0.0000
SER01	0.008317	0.002023	4.111649	0.0001
R-squared	0.139825	Mean dependent var		-0.018013
Adjusted R-squared	0.131554	S.D. dependent var		0.059123
S.E. of regression	0.055097	Akaike info criterion		-2.940764
Sum squared resid	0.315708	Schwarz criterion		-2.890510
Log likelihood	157.8605	F-statistic		16.90566
Durbin-Watson stat	2.109717	Prob(F-statistic)		0.000002

Sursa: Realizat de autor cu ajutorul pachetului de programe EViews

În evaluarea entropiei în etapa a doua, în 15.03.21, trei serii de date suplimentare sunt generate prin calcul parțial (Anexa nr.1) reprezentând probabilitatea determinată frecvențial [$SER06 = p(x_i)$], logaritmul acestei probabilități [$SER07 = \log_b p(x_i)$, unde $b = e$] și entropia (SER08) detaliată per cuvânt. Aceste serii corelează mult mai intens numărul literelor din cuvinte cu seria logaritmată a mențiunilor ($R = -0.447$) și entropia ($R = -0,374$) așa cum se poate constata din tabel nr. 5:

Matrice de corelație a datelor din Anexa 1

Tabel nr. 5

	SER01	SER03	SER06	SER07	SER08
SER01	1.000000	-0.315902	-0.315902	-0.447054	0.373932
SER02	-0.315902	1.000000	1.000000	0.419067	-0.940017
SER06	-0.315902	1.000000	1.000000	0.419067	-0.940017
SER07	-0.447054	0.419067	0.419067	1.000000	-0.552637
SER08	0.373932	-0.940017	-0.940017	-0.552637	1.000000

Sursa: Realizat de autor cu ajutorul pachetului de programe EViews

Testul F validează în seria de date din 15.03.21 corelația de intensitate maximă descrisă în modelul econometric al mențiunilor (SER03) în raport cu

datele din seria logaritmată a acestora (SER07) în raport cu dimensiunile în litere ale cuvintelor (SER01), model expus în tabel nr. 6:

Model econometric al mențiunilor în raport cu seria logaritmată a probabilităților acestor mențiuni și cu dimensiunile cuvintelor

Tabel nr. 6

Dependent Variable:SER03 Method: Least Squares Sample:1 106 Included observations: 106				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.33E+08	1.22E+08	5.172132	0.0000
SER07	35500626	10096203	3.516235	0.0007
SER01	-24615515	15129929	-1.626942	0.1068
R-squared	0.196272	Mean dependent var		83066070
Adjusted R-squared	0.180665	S.D. dependent var		4.07E+08
S.E. of regression	3.69E+08	Akaike info criterion		42.31633
Sum squared resid	1.40E+19	Schwarz criterion		42.39171
Log likelihood	-2239.766	F-statistic		12.57638
Durbin-Watson stat	2.177465	Prob(F-statistic)		0.000013

Sursa: Realizat de autor cu ajutorului pachetului de programe EViews

În final, articolul oferă în tabel nr.7 și o soluție de calcul a entropiei specifice statisticii lingvistice, pornind de la modelul econometric al acesteia (SER08) în raport de numărul mențiunilor pe Internet a cuvintelor selectate (SER03) și de seria logaritmilor naturali ai probabilităților frecvențiale de apariție a cuvintelor în limbajul specific al Internetului (SER07):

Model econometric al entropiei în funcție de numărul mențiunilor cuvintelor selectate și de logaritmi naturali ai probabilităților frecvențiale ale mențiunilor

Tabel nr. 7

Dependent Variable:SER08 Method: Least Squares Sample:1 106 Included observations: 106				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.036269	0.005185	-6.994869	0.0000
SER03	-1.25E-10	4.62E-12	-27.03143	0.0000
SER07	-0.002857	0.000472	-6.055791	0.0000
R-squared	0.914186	Mean dependent var		-0.018013
Adjusted R-squared	0.912520	S.D. dependent var		0.059123
S.E. of regression	0.017487	Akaike info criterion		-5.226849
Sum squared resid	0.031496	Schwarz criterion		-5.151469
Log likelihood	280.0230	F-statistic		548.6349
Durbin-Watson stat	1.450300	Prob(F-statistic)		0.000000

Sursa: Realizat de autor cu ajutorului pachetului de programe EViews

O extensie a modelului cu includerea numărului de litere din cuvânt (inclusiv în model a SER01) este susceptibilă de multicolaritate, așa cum se poate constata din tabel nr. 8, conform valorii confruntate a coeficientului de determinație sau a lui R^2 (0.914531 vs 0.914186):

Model econometric al entropiei în funcție de cu numărul literelor din cuvinte, de numărul mențiunilor și de logaritmi naturali ai probabilităților frecvențiale ale mențiunilor

Tabel nr. 8

Dependent Variable: SER08 Method: Least Squares Sample: 1 106 Included observations: 106				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.038808	0.006535	-5.938500	0.0000
SER01	0.000468	0.000729	0.641458	0.5227
SER03	-1.24E-10	4.69E-12	-26.51275	0.0000
SER07	-0.002738	0.000508	-5.386895	0.0000
R-squared	0.914531	Mean dependent var		-0.018013
Adjusted R-squared	0.912017	S.D. dependent var		0.059123
S.E. of regression	0.017537	Akaike info criterion		-5.212007
Sum squared resid	0.031370	Schwarz criterion		-5.111500
Log likelihood	280.2364	F-statistic		363.8039
Durbin-Watson stat	1.470832	Prob(F-statistic)		0.000000

Sursa: Realizat de autor cu ajutorul pachetului de programe EViews

Concluzii

Există și excepții notabile referitoare la sensul lipsit de orice ambiguitate terminologică al limbajului prezent pe Internet, dar confruntat temporal, iar exemplul cel mai elocvent este acela al cuvântului *recesiune* (recession) care a dat naștere celebrului Rindex cu rol de semnal în apariția crizelor sau recesiunilor. Statistici descriptive, matrice de corelație și modelări econometrice constituie împreună un prilej de estimare a entropiei și de verificare a creșterii ei permanente în sistemul de comunicare și în limbajul Internet.

În acest articol, entropia marchează și în spațiul lingvisticii statistice trecerea inexorabilă a timpului și generalizează îmbătrânirea tuturor sistemelor filologice generate de limbile naționale (izolate teoretic), inclusiv a comunicării cu ajutorul limbajului și Internetului. Entropia este și aici tot o săgeată a timpului conform unor repere statistice de cuantificare a parcursului sau traseului săgeții... Pentru a aspira la asigurarea unei continuități evolutive a limbajului și Internetului, se va respecta cu siguranță și cunoscutul principiu “zero al termodinamicii”, aplicat și în cazurile particulare ale economiei, sociologiei și demografiei. Conform acestuia “*orice ansamblu statistic sau sistem particular - descris de științele cu care se asigură interdisciplinaritatea de ordin II a fizicii statistice, respectiv cazul distinct al econofizii, sociofizicii și demografizicii - situat în condiții externe invariabile în timp va atinge, după un timp suficient de lung, o stare de echilibru necesară și așteptată, și simultan una de tranzitivitate a acestui echilibru*”, de la o generație la alta (Săvoiu, 2021). Acest adevăr se validează practic așa cum se poate vedea în acest articol și în lingvistica statistică, în comunicarea pe Internet destructurată la nivel de limbi naționale și limbaje științifice etc.

Referințe bibliografice

1. A short history of the Web | CERN. 2020. *Where the Web was born. 2020*. Available online at: <https://home.cern/science/computing/shot-history-web>. Accessed on the 27th of January 2021.
2. Anderson, R.H., Bikosn, T. K., Law, S. A. & Mitchell, B.M. 1995. *Universal Access to E-Mail: Feasibility and Societal Implications*. Santa Monica: Rand Corporation.
3. Bellamy, D. 1987. *Innovative applications of computer technology in social work*. Paper presented at the Conference of the Canadian Association of Schools of Social Work, Learned Societies Meeting, Hamilton, Ontario.
4. Carrizo, P. G. 2018. *Cultura generală în 365 de zile*, București: Editura Lider Internațional.
5. Georgescu-Roegen, N. 1971. *The Entropy Law and the Economic Process*, Cambridge: Harvard University Press.
6. Gibbs, J.W. 1902. *Elementary Principles in Statistical Mechanics*, Dover: Yale Univ. Press.
7. Greene, R., 2000. *Web work: a history of internet art*, Art Forum International, pp.162–190.
8. Manovich, L. 2001. *The Language of New Media*. Cambridge MIT: Cambridge, Press.
9. Marson, S. M. 1997. A Selective History of Internet Technology and Social Work. *Computers in Human Services*, vol. 14(2), pp. 35–49.
10. Mowbray, S. 2017. *Did Switzerland Invent the Internet?* Available online at: <https://theculturetrip.com/europe/switzerland/articles/did-switzerland-invent-theinternet/>, Accessed on the 26th of January, 2021.
11. Săvoiu, G. 2009. *Statistica: mod de gândire și metode*, București: Editura Universitară.
12. Săvoiu, G., Cudanov, M. 2020. New Typology of Online Scientific Research and Multi-disciplinarity, *ESMSJ*, vol 9(2), pp. 4-8. Available online at: https://www.esmsj.upit.ro/No20_2020.html Accessed on the 23th of January 2021.
13. Săvoiu, G. 2013. *Situații statistice financiar-contabile și sisteme de indicatori statistici derivați*, București: Editura universitară [*Financial and accountancy statistical situations and systems of derived statistical indicators*, Bucharest: Universitara Publishers].
14. Săvoiu, G. 2021. *Metode statistice și interdisciplinaritate*, București: Editura Universitară.
15. Shannon, C.E. și Weaver, W. 1949. *The Mathematical Theory of Communication*. Urbana, Illinois: University of Illinois Press.
16. The birth of the Web | CERN 2020. *The World Wide Web was invented by British scientist Tim Berners Lee in 1989 while working at CERN*. Available online at: <https://home.cern/science/computing/birth-web>. Accessed on the 25th of January 2021.
17. Language. 2020. Oxford English Dictionary, Language, Available on line at: <https://www.lexico.com/definition/language>, Accessed on the 12th of February 2021.

**Calculul detaliat al entropiei eșantionului de 106 cuvinte în limba română
extras aleator din DEX (2012), în legătură cu mențiunile pe Internet,
cuantificate cu Google Search în 15.03.2021**

Pag.	Col. 1	Col. 2	Cuvant extras aleator	Numar de litere	Mențiuni Google pe Internet	Valori determinate* per cuvânt în data de 15.03.2021		
						15.03.2021	$P_{(x_i)}$	$\ln P_{(x_i)}$
50	4		antispumant	11	15800	1.79443429E-06	-13.23082075	-2.37418E-05
372		39	exponent	8	24900000	0.002827937578	-5.868207604	-0.016594925
488	50		iepuroi	7	33500	3.80465498E-06	-12.47928525	-4.74794E-05
1109		60	tâmpână	7	3	3.40715371E-10	-21.79997368	-7.42759E-09
251		44	criticism	9	225000000	2.55536528E-02	-3.666975005	-0.093704606
92		49	bavură	6	41300	4.69051494E-06	-12.26996819	-5.75525E-05
1038		43	spectacol	9	5390000	6.12151950E-04	-7.398530022	-0.004529025
18	73		afet	4	13100000	1.48779045E-03	-6.510463177	-0.009686205
672	20		monosepal	9	4200	4.77001519E-07	-14.55574616	-6.94311E-06
759	11		ovreică	7	4670	5.30380261E-07	-14.44967161	-7.66382E-06
277	12		decăt	5	55200000	6.26916282E-03	-5.072112454	-0.031797899
118	46		bregmă	6	497	5.64451798E-08	-16.68999594	-9.4207E-07
487		8	idiotism	8	120000	1.36286148E-05	-11.20333894	-0.000152686
160		35	căruș	7	44200	5.01987313E-06	-12.2021059	-6.1253E-05
861		50	prăvălie	8	153000	1.73764839E-05	-10.96039276	-0.000190453
559		1	învălu	7	207000	2.35093606E-05	-10.65811189	-0.000250565
1195	63		vesteji	7	37600	4.27029932E-06	-12.36382664	-5.27972E-05
503	70		individualizare	15	123000	1.39693302E-05	-11.17864633	-0.000156158
676		37	moșneguț	8	13600	1.54457635E-06	-13.38076089	-2.06676E-05
172	53		cerdac	6	120000	1.36286148E-05	-11.20333894	-0.000152686
870		14	prestigiu	9	1180000	1.34014713E-04	-8.917560969	-0.001195084
38	55		amplexiune	10	692	7.85916789E-08	-16.35900001	-1.28568E-06
541	3		încețare	10	55600	6.31459154E-06	-11.97264749	-7.56024E-05
964	13		rotocol	7	528000	5.99659053E-05	-9.721734403	-0.000582973
29	64		alicuantă	9	760	8.63145606E-08	-16.26526753	-1.40393E-06
433		59	geobotanic	10	497000	5.64451798E-05	-9.78224066	-0.00055216
98	18		bazăi	5	784000	8.90402836E-05	-9.326421666	-0.000830427
928		10	recepționar	11	693000	7.87052507E-05	-9.449800687	-0.000743749
489	40		iezuitic	8	2130	2.41907913E-07	-15.23470871	-3.6854E-06
263		63	carpen	6	263000	2.98693808E-05	-10.41867665	-0.000311199
976	37		sarcosporidioză	15	580	6.58716384E-08	-16.53555786	-1.08922E-06
1159	27		țacănit	7	65900	7.48438098E-06	-11.80269224	-8.83358E-05
1052		42	stigmă	6	56200000	6.38273461E-03	-5.05415865	-0.032259353
684		47	muslim	6	575000000	6.53037794E-02	-2.728705367	-0.178194773
329		57	drag	4	817000000	9.27881527E-02	-2.377436312	-0.220597924
399	21		fixism	6	24700	2.80522322E-06	-12.78402744	-3.58621E-05
497	7		imunotransfuzie	15	303	3.44122525E-08	-17.18485316	-5.9137E-07
193		61	clasicism	9	78700	8.93809990E-06	-11.62518753	-0.000103907
1005	69		senar	5	13000000	1.47643327E-03	-6.51812605	-0.009623578
262		76	curea	5	10700000	1.21521816E-03	-6.712831666	-0.008157555
608		72	machetist	9	18200	2.06700658E-06	-13.08940909	-2.70559E-05
1065	23		subînțelege	11	195000	2.21464991E-05	-10.71783113	-0.000237362
1043	39		sprânceană	10	2490000	2.82793758E-04	-8.170792697	-0.002310649
378	74		faclă	5	774000	8.79045657E-05	-9.339258813	-0.000820963
779		25	pas	3	2920000000	3.31629628E-01	-1.103736512	-0.366031729
40	2		analitic	8	1100000	1.24928969E-04	-8.987765228	-0.001122832
935		59	regal	5	235000000	2.66893707E-02	-3.623489893	-0.096708665
897	29		pumnișor	8	1890	2.14650684E-07	-15.35425386	-3.2958E-06
1095		20	știre	5	55560000	6.31004867E-03	-5.065611889	-0.031964258
1082		9	sfadă	5	129000	1.46507609E-05	-11.13101828	-0.000163078
714	16		neurolimfă	10	171	1.94207761E-08	-17.75692241	-3.44853E-07

579	66		laterit	7	677000	7.68881020E-05	-9.473159413	-0.000728373
798		41	penetrație	10	3640000	4.13401317E-04	-7.791091726	-0.003220848
735		28	odolean	7	139000	1.57864789E-05	-11.05635675	-0.000174541
23		15	ahiat	6	36800	4.17944188E-06	-12.38533284	-5.17638E-05
442	75		glorie	6	8920000	1.01306037E-03	-6.894779461	-0.006984828
1068		38	succesor	8	564000	6.40544897E-05	-9.655776435	-0.000618496
983	45		sângerare	9	583000	6.62123537E-05	-9.6226435	-0.000637138
225	36		constituție	11	1310000	1.48779045E-04	-8.81304827	-0.001311197
353		67	endometrită	11	52900	6.00794771E-06	-12.02242735	-7.22301E-05
145		56	cardanic	8	458000	5.20158800E-05	-9.863961502	-0.000513083
663	53		moale	5	9050000	1.02782470E-03	-6.88031065	-0.007071753
155		33	căciulă	7	2800000	3.18001013E-04	-8.05345599	-0.002561007
298		49	desfăta	7	131000	1.48779045E-05	-11.11563336	-0.000165377
479	63		homeoterm	9	13300	1.51050481E-06	-13.40306665	-2.02454E-05
829	32		platelaj	8	3200	3.63429729E-07	-14.82767988	-5.38882E-06
845		40	politicesc	10	3940	4.47472854E-07	-14.61964996	-6.5419E-06
216	6		concentra	9	39900000	4.53151443E-03	-5.396699083	-0.02445522
598	12		loc	3	313000000	3.55479704E-02	-3.336872217	-0.118619035
485	76		iaz	3	3300000	3.74786908E-04	-7.889152939	-0.002956751
252		61	chromocistoscopie	16	4690	5.32651697E-07	-14.4453981	-7.69437E-06
25		28	albie	5	3300000	3.74786908E-04	-7.889152939	-0.002956751
636		72	meditație	9	5580000	6.33730590E-04	-7.363886631	-0.00466672
923		57	răuri	5	1420000	1.61271942E-04	-8.732418536	-0.001408294
702	17		nefresce	8	147000	1.66950532E-05	-11.0003981	-0.000183652
273		9	dărmoz	6	9620000	1.09256062E-03	-6.819231143	-0.007450423
1000		4	seimen	6	393000	4.46337136E-05	-10.01702107	-0.000447097
637	75		megafon	7	17000000	1.93072044E-03	-6.249862063	-0.012066736
962	7		rontgen	7	7210000	8.18852608E-04	-7.107606456	-0.005820082
1193		73	verniz	6	48200000	5.47416029E-03	-5.207716386	-0.028507874
237	67		corijent	8	6610	7.50709534E-07	-14.10224703	-1.05867E-05
836		21	pluti	5	842000	9.56274474E-05	-9.255050672	-0.000885037
639	74		melodramă	9	21300000	2.41907913E-03	-6.024368335	-0.014573424
1099	60		tachina	7	572000	6.49630641E-05	-9.641691695	-0.000626354
907		1	radioastronomie	15	646000	7.33673765E-05	-9.520031183	-0.00069846
21		34	agoniza	7	4490000	5.09937338E-04	-7.581222706	-0.003865949
10	3		act	3	2880000000	3.27086756E-01	-1.117529834	-0.365529208
916	10		răgușit	7	134000	1.52186199E-05	-11.09299089	-0.00016882
325		2	documentație	12	8680000	9.85803140E-04	-6.922053879	-0.006823782
1024	41		sistem	6	324000000	3.67972601E-02	-3.302331892	-0.121516765
1066	5		subordonat	10	422000	4.79272955E-05	-9.945825372	-0.000476677
1113	35		teleancheta	11	87	9.88074576E-09	-18.43267785	-1.82129E-07
871		56	preursi	7	1940	2.20329273E-07	-15.32814271	-3.37724E-06
979	37		sălbatic	8	3080000	3.49801114E-04	-7.95814581	-0.002783768
373		70	extenuant	9	144000	1.63543378E-05	-11.02101739	-0.000180241
949	66		retrimitere	11	128000	1.45371892E-05	-11.13880042	-0.000161927
346	36		electricitate	13	4770000	5.41737440E-04	-7.520729102	-0.004074261
99	11		belcanto	8	5780000	6.56444948E-04	-7.328671725	-0.00481087
1218		55	zărgeală	9	1040	1.18114662E-07	-15.95160997	-1.88412E-06
641		30	mercur	6	4500000	5.11073056E-04	-7.578998011	-0.003873422
864		27	predecedat	10	10400	1.18114662E-06	-13.64902488	-1.61215E-05
653		23	mijloc	6	9080000	1.03123186E-03	-6.877001215	-0.007091783
772	51		parafrizare	11	11600	1.31743277E-06	-13.53982559	-1.78378E-05
572	44		kilometru	9	2780000	3.15729577E-04	-8.06062448	-0.002544978
1107		18	tălgerii	8	15900	1.80579147E-06	-13.22451158	-2.38807E-05
14	43		adjunct	7	40300000	4.57694315E-03	-5.386723938	-0.024654729
Total = 106 cuvinte		Total= 837 (7,9litere/ cuvânt)		8805003403	1.00000000E+00	-	-	-1.909348493

*Notă: Determinarea valorii probabilității $p(x_i)$ este de natură frecvențială $p(x_i) = f(x_i) : \Sigma f(x_i)$