Treballs de Sociolingüística Catalana, núm. 33 (2023), p. 141-157 DOI: 10.2436/20.2504.01.207
http://revistes.iec.cat/index.php/TSC ISSN (ed. impresa): 0211-0784 ISSN (ed. electrònica): 2013-9136

Formulació d'un indicador unitari de competència i d'ús de la llengua per a l'avaluació de polítiques lingüístiques

Formulation of a unitary indicator of language competence and use for the evaluation of language policies

Francesc J. Hernàndez
Universitat de València

Data de recepció: 11 d'octubre de 2021
Data d'acceptació: 3 de maig de 2022

Resum

Aquest article formula un indicador unitari per a competències i usos lingüístics (E), que té en compte tant la taxa com el nombre d'individus d'una unitat territorial. En primer lloc, revisa un model simple de relació entre competència i ús, el que obri la possibilitat d'una consideració de l'entropia lingüística i presenta una singularitat susceptible de ser interpretada com una catàstrofe cúspide. Posteriorment, adapta la coneguda fórmula de l'entropia de Boltzmann i resol alguns problemes pràctics del càlcul. Finalment, determina l'indicador per a les dades d'ús a la llar i públic del valencià (2005-2021) i fa palesa la tendència d'increment de l'entropia.

Paraules clau: singularitat cúspide, entropia, ús lingüístic a la llar, ús lingüístic públic, indicador E.

Abstract

This paper formulates a unitary indicator of language skills and uses (E) that takes into account both the rate and the number of individuals in a territorial unit. First, it reviews a simple model of the relationship between competence and use, which opens up the possibility of a consideration of linguistic entropy and presents a singularity that can be interpreted as a cusp catastrophe. It then adapts Boltzmann's well-known entropy formula and solves some practical calculus problems. Lastly, it calculates the indicator for the data on home and public use of Valencian (2005-2021) and shows the trend of increasing entropy.

Keywords: cusp singularity, entropy, home language use, public language use, indicator E.

Correspondència: Francesc J. Hernàndez. Universitat de València. Departament de Sociologia i Antropologia Social. Facultat de Ciències Socials. Campus dels Tarongers. Avinguda dels Tarongers, 4 b. 46021 València. A/e: *francesc.j.hernandez@uv.es*. A/I: *http://www.uv.es/fjhernan*. Tel.: 963 828 454. Fax: 963 828 450.

1. Introducció i revisió del model simple

A Hernàndez (2020*a*) plantejàrem un model de relació entre *competència* (oral activa) i *ús* (públic) de la llengua, que proporcionava aproximacions satisfactòries als resultats de les enquestes de coneixement i ús del valencià de la Generalitat Valenciana (SIES-CEdCEs, 2005, 2010 i 2015). En primer lloc, revisarem aquest model i, a continuació, formularem un indicador unitari per a les competències i els usos de la llengua amb vista a l'avaluació de polítiques lingüístiques.

Considerem una societat amb individus plurilingües, que anomenarem Π_1, formada per P individus (en el cas de la Comunitat Valenciana, uns cinc milions), un subconjunt dels quals (m) està format pels individus plurilingües que són *competents* en la llengua μ (en el nostre cas, el valencià/català). Si anomenem TC_μ la taxa de competència (a Hernàndez, 2020*a*, l'«oral activa») i TU_μ, la taxa d'ús (a Hernàndez, 2020*a*, l'«ús públic»), la relació entre les dues taxes en general consisteix a formular una funció $\mathfrak{F}$, en què:

$$\mathfrak{F}\left(TC_\mu\right) = TU_\mu \qquad [1]$$

Amb [1] expressem l'ús *en funció* de la competència. Quan fan servir una llengua μ, els m individus (parlants) es poden relacionar de dos en dos (de la manera que anomenarem *diàdica*), de tres en tres (*triàdica*), etc., fins a una relació *m-àdica*. Deixem de banda les relacions *monàdiques*, un assumpte d'importància filosòfica o literària, però no sociològica. Si considerem que p_i és la probabilitat de les relacions *i-àdiques* (és a dir, la seua proporció en el total de les relacions *m-àdiques*), podem establir, seguint l'argument exposat a Hernàndez (2020*a*), que:

$$TU_\mu = \sum_{i=2}^{m} p_i TC_\mu^i \qquad [2]$$

Pel que fa a p_i, resulta trivial la primera condició, a saber, que la suma de tots els seus valors és la unitat:

$$\sum_{i=2}^{m} p_i = 1 \qquad [3]$$

Suposarem també una segona condició:

$$p_2 > p_3 > p_4 > \ldots > p_{m-1} > p_m \qquad [4]$$

Aquestes condicions es verifiquen si, com proposàrem a Hernàndez (2020*a*),

$$p_i = \frac{3}{2}\left(\frac{1}{i^2}\right) \qquad [5]$$

Però [5] produeix una desviació xicoteta. Amb m = 100.000, el valor del sumatori de [3] és 0,9673 i no pas 1,0000. Per això, podem proposar una altra definició de p_i que produïsca aproximacions millors. És el cas de la fórmula següent:

$$p_i = \frac{3101}{2000}\left(\frac{1}{i^2}\right), \qquad [6]$$

que permet un valor del sumatori de 0,9999, amb la mateixa quantitat de m esmentada.

Per tant, substituint [6] en [1] i [2], podem formar:

$$\mathfrak{F}\left(TC_\mu\right) = TU_\mu = \sum_{i=2}^{m} \frac{3101}{2000}\left(\frac{1}{i^2}\right) TC_\mu^i \qquad [7]$$

És clar que [7] representa una suma immensa (en el cas del valencià, amb més de dos milions de sumands!), per la qual cosa proposàrem (Hernàndez, 2020*a*) un procediment per substituir aquesta suma per una equació polinòmica més manejable. En resum, el procediment és calcular TU_μ per a cada mil·lèsima (o deu mil·lèsima) entre 0,00 i 1,00 de la TC_μ, representar la corba resultant i obtindre'n (mitjançant un programa informàtic habitual, com ara Excel) l'equació polinòmica de la línia de tendència. Així podem arribar a l'equació:

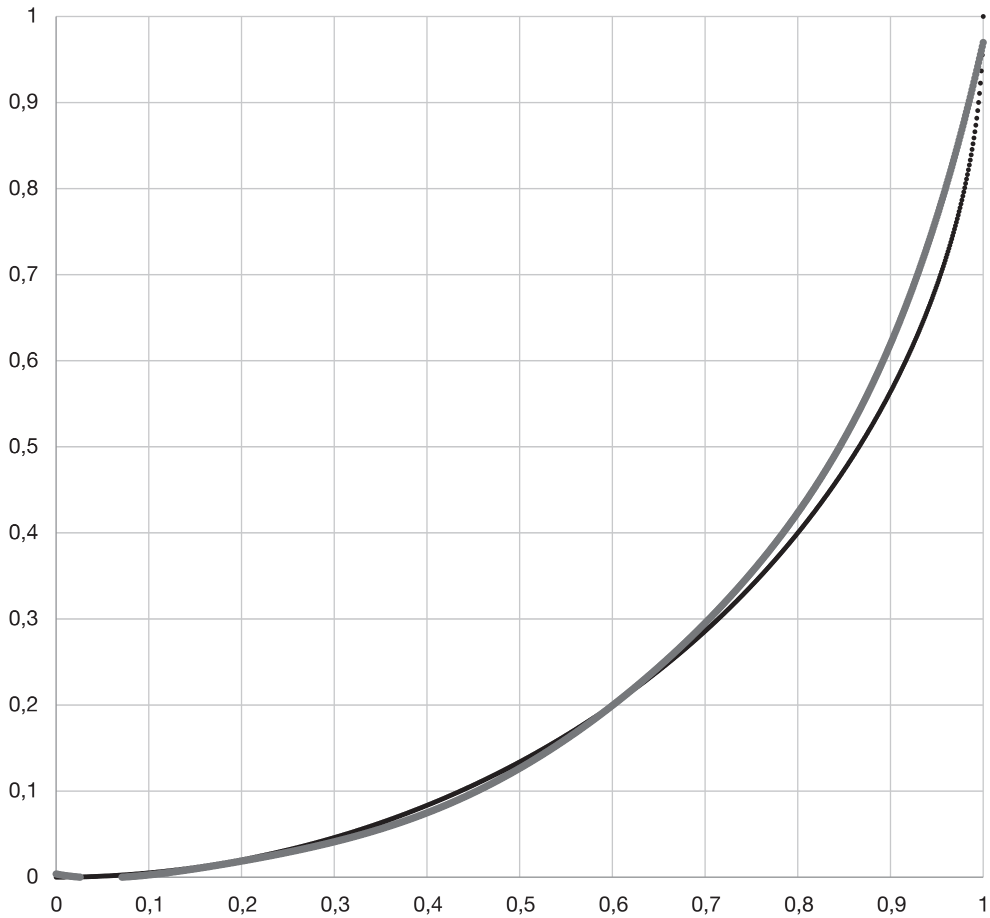
$$\begin{aligned}\mathfrak{F}\left(TC_\mu\right) = TU_\mu &= 10{,}757TC_\mu^6 - 27{,}347TC_\mu^5 + 26{,}809TC_\mu^4 - 11{,}996TC_\mu^3 \\ &+ 2{,}9533TC_\mu^2 - 0{,}2102TC_\mu + 0{,}0037\end{aligned} \qquad [8]$$

Cal fer esment de dues característiques de $\mathfrak{F}$. La primera, totalment inesperada, és que proporciona una corba *molt semblant* a la de la circumferència, que anomenarem $\mathcal{O}$:

$$\mathcal{O}\left(TC_\mu\right) = TU_\mu = 1 - \sqrt{1 - TC_\mu^2} \qquad [9]$$

És el que es pot apreciar al gràfic 1, en què els valors de [8] es representen en gris i els de [9], en negre. Les corbes estan dins del que anomenarem *quadrat unitat*, perquè els valors de TC_μ i TU_μ estan generalment en l'interval (tancat) entre 0 i 1.

Gràfic 1
Representació de [8] i de [9]



Font: Elaboració pròpia.

En segon lloc, cal fer constar que [8] presenta una desviació mínima, ja que, per definició, quan $TC_\mu = 0$, aleshores $TU_\mu = 0$ (si ningú no és competent en una llengua, ningú no la pot usar). Però en [8], si $TC_\mu = 0$, aleshores $TU_\mu = 0{,}0037$. Si eliminem l'últim sumand (realment un residu del procediment d'Excel per establir l'equació polinòmica de la corba), podem formular una funció esmenada, que anomenarem $\mathcal{F}(TC_\mu)$:

$$\begin{aligned}\mathcal{F}\left(TC_\mu\right) = TU_\mu = 10{,}757TC_\mu^6 - 27{,}347TC_\mu^5 + 26{,}809TC_\mu^4 - 11{,}996TC_\mu^3 \\ +2{,}9533TC_\mu^2 - 0{,}2102TC_\mu\end{aligned} \qquad [10]$$

Aquesta funció $\mathcal{F}(TC_\mu)$ és molt interessant perquè es pot factoritzar fàcilment, i obtenim:

$$\mathcal{F}\left(TC_\mu\right)=\mathcal{G}\left(TC_\mu\right)\cdot\mathcal{H}\left(TC_\mu\right), \tag{11}$$

cosa que s'acompleix en el cas que $\mathcal{G}(TC_\mu)$ i $\mathcal{H}(TC_\mu)$ siguen, respectivament:

$$\mathcal{G}\left(TC_\mu\right)\Rightarrow TU_\mu=TC_\mu \tag{12}$$

$$\begin{gathered}\mathcal{H}\left(TC_\mu\right)\Rightarrow TU_\mu=10{,}757TC_\mu^5-27{,}347TC_\mu^4+26{,}809TC_\mu^3-11{,}996TC_\mu^2\\+2{,}9533TC_\mu-0{,}2102\end{gathered} \tag{13}$$

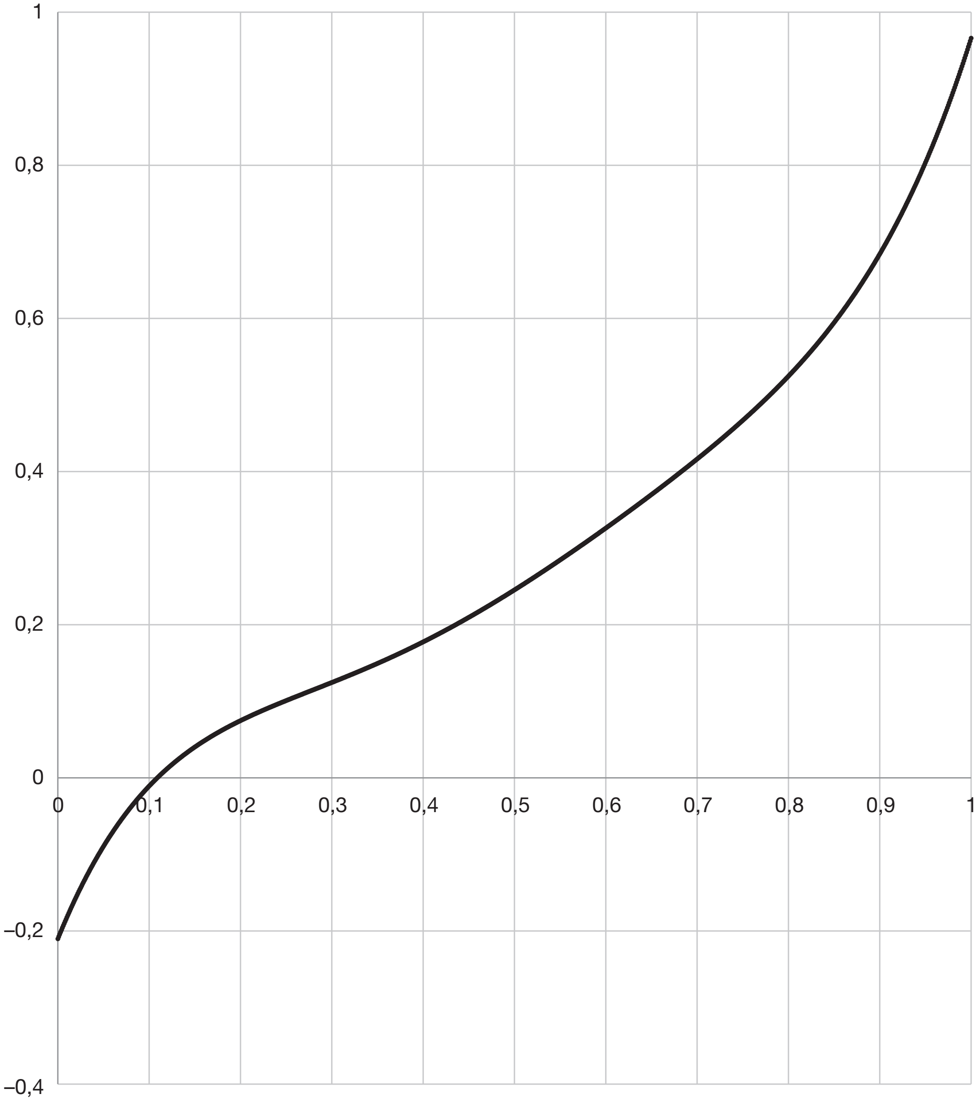
Més concretament, el que hem fet és traure TC_μ com a factor comú a la suma de [10]. Comentem aquestes dues funcions a continuació.

La corba definida per la funció $\mathcal{G}(TC_\mu)$ seria la línia recta que uneix els punts (0, 0) i (1, 1), és a dir, la diagonal que anomenarem *creixent del quadrat unitat*, perquè cada valor de TU_μ és igual que cada valor de TC_μ. Aquesta representació correspon a una situació d'igualtat segons l'índex de Gini.

L'economista Corrado Gini formulà un índex (coeficient, si el seu valor es multiplica per 100) per mesurar la desigualtat socioeconòmica. Partim del fet que, si representem el percentatge acumulat de població en les abscisses (com una mostra o un univers ordenat) i el percentatge de riquesa acumulada en les ordenades, en una situació d'igualtat plena, la corba resultant serà la diagonal creixent (perquè, per exemple, el 10 % de la població acumula el 10 % de la riquesa, el 20 % disposa del 20 %, etc.). En les situacions socioeconòmiques reals, però, la corba resultant serà una línia per sota de la diagonal creixent, amb la qual compartirà els punts (0, 0) i (1, 1) (perquè, de tota manera, el 0 % de la població acumula el 0 % dels recursos i el 100 % dels recursos són acumulats pel 100 % de la població). Per això, l'índex de Gini es defineix com la proporció entre la superfície que hi ha entre la corba resultant i la diagonal creixent i la superfície per sota d'aquesta diagonal (que òbviament mesura la meitat de l'àrea quadrat unitat, és a dir, 1/2).

Retornem ara a $\mathcal{G}(TC_\mu)$, que representa l'ús lingüístic en funció de la competència. Quina interpretació sociolingüística en podem donar? El cas sociolingüístic més simple seria el de la societat Π_2, formada per Q individus, un subconjunt dels quals, r, *només* és competent i usa la llengua ϱ; i el subconjunt complementari, diguem-ne s, *només* és competent i usa la llengua σ. Per tant, s'acompleix [12] per a cada subconjunt i llengua. Més endavant caracteritzarem aquesta situació com d'*entropia (lingüística) zero*. D'aquesta manera, ja podem intuir que una corba per sota de la diagonal creixent definirà situacions de multilingüisme. Ara bé, si continuem amb el cas de Π_1 i representem $\mathcal{H}(TC_\mu)$, arribem a una corba interessant, que es mostra en el gràfic 2.

Gràfic 2
Representació de $\mathcal{H}(TC_\mu)$



Font: Elaboració pròpia.

Com es pot veure al gràfic, la corba presenta un punt d'inflexió, que anomenarem Я (l'última lletra de l'alfabet rus —que es pronuncia [ja] o [^{j}a], igual que el seu pronom de primera persona singular), tot i que tal vegada seria més adequat anomenar-lo Ω, si bé aquesta lletra ja té una significació entròpica que comentarem més endavant. Podem establir aquest punt d'inflexió mitjançant el procediment habitual en matemàtiques:

a) Calculem la derivada primera de la funció $\mathcal{H}(TC_\mu)$:

$$\begin{aligned}\mathcal{H}' = (5)(10{,}757)TC_\mu^4 - (4)27{,}347TC_\mu^3 + (3)(26{,}809)TC_\mu^2 \\ -(2)(11{,}996)TC_\mu + 2{,}9533\end{aligned} \qquad [14]$$

b) Calculem la derivada segona:

$$\begin{aligned}\mathcal{H}'' = (4)(5)(10{,}757)TC_\mu^3 - (3)(4)27{,}347TC_\mu^2 + (2)(3)(26{,}809)TC_\mu \\ -(2)(11{,}996) = 215{,}14TC_\mu^3 - 328{,}164TC_\mu^2 + 160{,}854TC_\mu - 23{,}992\end{aligned} \qquad [15]$$

c) Els valors que fan que $\mathcal{H}'' = 0$ són:

$$\begin{aligned}TC_\mu &= 0{,}27850 \\ TC_\mu &= 0{,}62342 - 0{,}10846\,i \\ TC_\mu &= 0{,}62342 + 0{,}10846\,i\end{aligned} \qquad [16]$$

Tanmateix, només el primer valor és un nombre real. Per tant, el punt Я correspon a les ordenades (0,27850, 0,11446). Ara bé, aquest mateix punt d'inflexió l'haurem de *suposar* per al subconjunt *n* competent en la llengua *v*, per la qual cosa la corba del gràfic 2 hauria de ser modificada, amb el postulat *d'un altre punt* d'inflexió, que correspondria al que podríem anomenar d'una manera tal vegada imprecisa *simètric* de Я. Si es tractara d'una corba parabòlica, el simètric del punt (x, y) seria el punt $(1 - y, 1 - x)$; en conseqüència, el simètric, que podríem representar com a $Я^{-1}$, seria (0,88553, 0,72110), però no hem d'acceptar necessàriament el supòsit de la corba parabòlica.

Aquest punt d'inflexió sobtat obri la possibilitat d'una *singularitat*, a l'estil de la catàstrofe «cúspide», emprada en ciències socials (cf. Guts, Frolova i Páutova, 2013), que ja fou plantejada en la nostra sociolingüística per Querol (1997). En realitat, les diverses singularitats enunciades per la teoria de les catàstrofes semblen reduir-se a la catàstrofe «cúspide» (cf. Boss, 2005: 108-110 i, sobretot, Boss, 2009: 122), però aquest assumpte ultrapassa la pretensió de l'article.

Què significa exactament una *multiplicació de funcions*, com és el cas de $\mathcal{G}(TC_\mu) \cdot \mathcal{H}(TC_\mu)$? O podem formular la pregunta inversa: què vol dir la factorització que hem fet de la funció $\mathcal{F}(TC_\mu)$? És difícil aportar en aquest punt de l'argument una interpretació sociolingüística d'una multiplicació de funcions; per això, bandejarem ara aquesta qüestió i la reprendrem més endavant.

Fins ací hem considerat una corba polinòmica (una corba que es pot expressar mitjançant una equació polinòmica de grau sisé), però també és possible construir un model amb una corba parabòlica o una corba exponencial. En tots dos casos, només es requereix un punt, que podem expressar com (TC_{μ_i}, TU_{μ_i}), ja que hem de suposar els punts (0, 0) i (1, 1). Es pot trobar un desenvolupament d'aquestes consideracions a Hernàndez (2020*b*).

A continuació portarem aquestes consideracions matemàtiques a l'àmbit del valencià, que ara representarem per la lletra grega μ, mentre que emprarem la lletra grega v per al castellà. Suposarem que:

$$TC_v = 1 \quad [17]$$
$$TU_\mu + TU_v = 1 \quad [18]$$

Hem escrit *suposarem*, perquè és clar que [17] pot tindre un valor més petit que la unitat i [18] deixa de banda llengües minoritàries. Aquests postulats permeten, tanmateix, simplificar els càlculs posteriors. El problema metodològic és que [18] presenta l'ús com una variable *dicotòmica*, quan el fet és que les enquestes habituals ofereixen diverses opcions per registrar-lo (per a un àmbit determinat, es pregunta si *vosté parla sempre valencià*, *generalment valencià*, *més valencià que castellà*, *indistintament*, *més castellà que valencià*, *generalment castellà*, *sempre castellà*, *altres llengües*, *NS/NC*?), per la qual cosa més endavant farem servir un procediment per dicotomitzar TU_μ i TU_v.

Després d'aquestes consideracions estem en disposició de plantejar l'assumpte central d'aquest article.

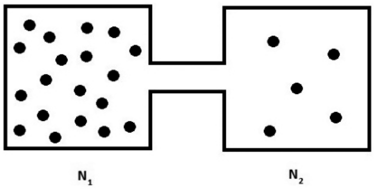
2. El problema d'un indicador unitari i l'adaptació de l'equació de Boltzmann

En l'àmbit de la sociolingüística, la competència o l'ús habitualment es formulen mitjançant variables relatives, com ara les taxes que hem vist en l'apartat anterior. D'aquesta manera es bandegen variables absolutes, com ara la grandària de la població d'una unitat territorial. Aquesta, però, esdevé lingüísticament important (per exemple, una dimensió determinada permetrà el manteniment d'una indústria editorial, la viabilitat de mitjans de comunicació o el sorgiment de figures literàries). Com podem combinar variables relatives i absolutes, en el nostre cas, una taxa de competència o d'ús i la grandària de la població d'una unitat territorial o el nombre de persones competents o usuàries d'una llengua? (I això com un pas previ per permetre elaborar un únic *baròmetre* que puga orientar la revitalització —cf. Costa i Petit Cahill, 2021— de la nostra llengua).

Aquest problema no és nou en la ciència. Ja fou plantejat en el camp de la física matemàtica fa més d'un segle per Ludwig Boltzmann, cosa que el portà a formular la seua equació de l'entropia (per a altres usos d'aquesta noció, cf. Georgescu-Roegen, 1971).

Suposem dos dipòsits amb un gas, interconnectats per un conducte, com mostra el gràfic 3. Cal dir que aquest gràfic està summament simplificat; en realitat, una mostra xicoteta d'un gas té milions de molècules —per exemple, un gram d'hidrogen té una quantitat de molècules que ultrapassa un 3 i 24 zeros darrere, o un centímetre cúbic d'aire té 10 i 20 zeros al darrere. Per imaginar la immensa quantitat de molècules que hi ha en una mostra de matèria podem fer servir una altra comparació: hi ha més molècules d'aigua en un got d'aigua que gots d'aigua podem emplenar amb tota l'aigua de la Terra (exemple manllevat de Penrose, 2006). Amb totes les precaucions sobre la seua simplificació, podem imaginar la situació descrita en el gràfic 3.

Gràfic 3
Model simplificat de l'entropia segons Boltzmann



Font: Elaboració pròpia.

Boltzmann definí l'entropia (S) per mitjà d'aquestes dues fórmules (segons el resum de Prigogine, 2019: 24):

$$S = k \ln \Omega \quad [19]$$

$$\Omega = \frac{(N_1 + N_2)!}{N_1! N_2!} \quad [20]$$

A [19], k és una constant i *ln*, l'abreviatura de *logaritme natural*. El valor de Ω es mostra d'una manera simplificada en [20], on es considera que N_1 i N_2 són el nombre de molècules del gas en cadascun dels dos dipòsits i el signe *!* representa el factorial d'un nombre natural o enter.

S és una mesura del desordre —o, inversament, de l'absència d'ordre— dels estats que es poden produir. Es considera que la situació en la qual totes les molècules estan en un dipòsit seria d'ordre absolut i, per tant, d'entropia *zero* (ja que o N_1 o N_2 serien zero, el seu factorial 1 i, aleshores, la fracció [20] tindria com a resultat 1 i el seu *ln* donaria com a resultat 0). Així mateix, el valor més alt d'entropia s'aconseguiria quan $N_1 = N_2$ (això coincideix amb la variància d'una variable nominal: la distribució binomial amb més variància és la de $\mathbb{P} = \mathbb{Q} = 0{,}5$).

Per tal de traslladar aquesta fórmula a la sociolingüística hem de resoldre un problema de càlcul no menor. Si considerem P el nombre d'individus i U, aquells que, per exemple, fan ús d'una llengua, és fàcil adaptar [20] de la manera següent:

$$\Omega = \frac{P!}{U!(P-U)!} \quad [21]$$

Cal dir, també, que aquesta és la fórmula de les combinacions de *P* elements agafats d'U en U.

Podríem procedir d'una manera anàloga amb la competència; conseqüentment, l'entropia lingüística, que anomenarem $\mathcal{S}$, serà:

$$\mathcal{S} = \ln \frac{P!}{U!\left(P - U\right)!} \qquad [22]$$

Com es pot veure, hem eliminat *k*. Ara abordarem el problema del càlcul de [21], perquè suposa treballar amb factorials de nombres molt elevats, la qual cosa de vegades no és possible amb calculadores ordinàries.[1] Aquesta dificultat es pot resoldre de dues maneres:

a) Podem fer servir programes matemàtics especialitzats. N'és un exemple la plataforma *wolframalpha.com*. N'hi ha prou d'escriure en la línia d'instruccions *ln ((P! / (U! * (P – U)!)* i substituir *P* i *U* pels valors corresponents per obtindre el resultat.

b) La representació en uns eixos cartesians de diversos valors de *P* i el seus *ln*, segons [22], proporciona una línia recta quan es conserva la proporció entre *P* i *P* – *U*. Aleshores, resulta fàcil trobar valors baixos que acomplisquen la proporció, esbrinar l'equació de la línia recta resultant de la representació en els eixos cartesians i resoldre el càlcul. Naturalment, el procediment *a* és el recomanable.

A continuació tractem un altre problema. Per a una població d'uns cinc milions de persones, $\mathcal{S}$ pot variar entre 1 i 35 milions i, si considerem tot el domini —uns 16 milions de persones—, $\mathcal{S}$ pot arribar a un màxim de més de 120 milions, que són quantitats molt elevades per a un indicador o *baròmetre*. Per tal de treballar amb quantitats més manejables, podem aplicar un logaritme a [22]. Arribaríem així a formular el que anomenem *indicador* $\mathfrak{E}$, que seria:

$$\mathfrak{E} = \ln\left[\ln \frac{P!}{U!\left(P - U\right)!}\right] \qquad [23]$$

D'aquesta manera, $\mathfrak{E}$ pot adoptar un màxim que no ultrapassa el valor 20, tant en el cas del valencià com en el de tot el domini. Una altra possibilitat és fer servir un logaritme decimal, que proporciona valors màxims encara més petits. En les dues possibilitats caldria deixar de banda el cas en què $U = 0$ o $P - U = 0$, perquè aleshores la fracció val 1, el primer $ln = 0$, però el segon perd el seu sentit. De tota manera, és una situació impossible per definició en unitats territorials multilingües.

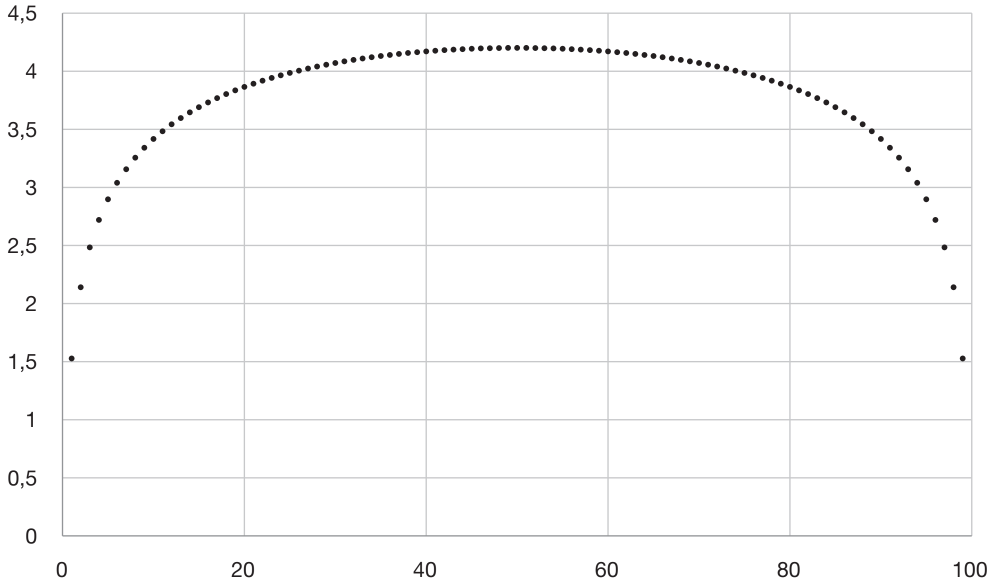
Ara bé, com hem indicat, $\mathfrak{E}$ adopta el mateix valor si intercanviem usuaris per no

1. La calculadora instal·lada a l'ordinador amb el qual redactem aquest article mostra el rètol *desbordament* amb 3.249! i quantitats superiors.

usuaris (és a dir, U i $P - U$). Farem una representació gràfica d'aquest problema i de la solució que hi proposem.

D'una manera simplificada, suposem una població de cent persones. Si representem les persones usuàries d'una llengua determinada en les abscisses (persones o *TU*, ací tant se val) i el valor corresponent de $\mathfrak{E}$ en les ordenades, tindrem el gràfic 4.

Gràfic 4
Representació de $\mathfrak{E}$ si P = 100 i la TU varia de 0 a 100 %



Font: Elaboració pròpia.

Per això, redefinirem l'indicador, que ara anomenarem E (èpsilon). N'hi ha prou d'establir-ne el valor d'aquesta manera:

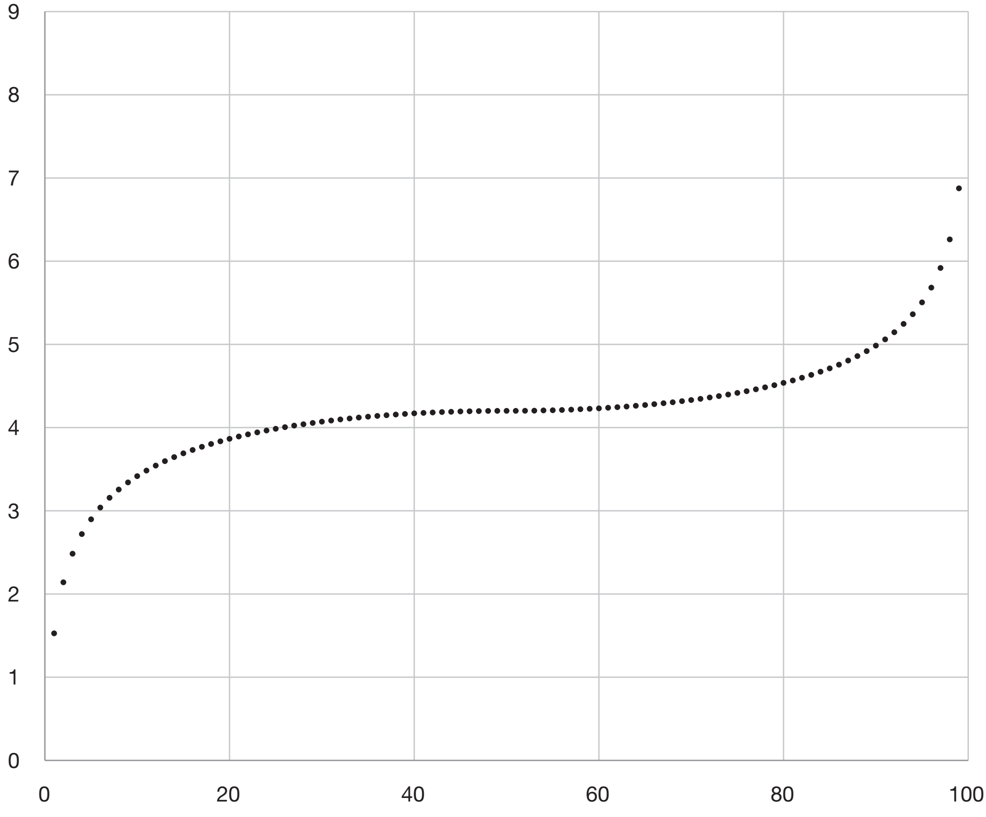
$$\mathrm{E} = \ln\left[\ln\frac{P!}{U!\left(P-U\right)!}\right] + \theta \qquad [24]$$

I afegir la clàusula següent per a θ:

$$\theta = \begin{cases} Si\, U > \left(P-U\right) \Rightarrow \theta = 0 \\ Si\, U < \left(P-U\right) \Rightarrow \theta = 2\left\{\ln\left[\ln\dfrac{P!}{\frac{P}{2}!\frac{P}{2}!}\right] - \ln\left[\ln\dfrac{P!}{U!\left(P-U\right)!}\right]\right\} \end{cases} \qquad [25]$$

És a dir, si $U > (P - U)$ —o sigui, si els usuaris són més que els no usuaris—, emprarem la fórmula [24], però si $U < (P - U)$ —és a dir, si els usuaris són menys que els no usuaris—, hi afegirem θ, que es defineix com el doble de la diferència entre el valor concret de E [24] i el seu valor més alt possible, que és E quan U i $P - U$ són iguals, o sigui, quan els dos presenten el valor $\frac{U}{2}$. D'aquesta manera, el gràfic 4 es converteix en el gràfic 5 (noteu que la meitat de la dreta de la corba adopta ara la posició simètrica respecte de l'ordenada amb el valor màxim del gràfic anterior).

Gràfic 5
Representació de E si P = 100 i la TU varia de 0 a 100 %



Font: Elaboració pròpia.

Com que considerem U i $P - U$ quantitats complementàries, podríem dir que l'entropia que afecta U té a veure amb la TU, però més exactament amb la $T(P - U)$. Altrament dit, quan s'incrementa la taxa de no ús $T(P - U)$, augmenta l'*entropia* de U.

3. Resultats per a l'ús a la llar i per a l'ús al carrer

Comencem per l'ús de la llengua a la llar. Explicarem el càlcul de E en un cas i estalviarem al lector els llargs càlculs en els altres.

En les enquestes (SIES-CEdCEs, 2005, 2010, 2015 i 2021), es distingeixen cinc regions sociolingüístiques en la zona valencianoparlant: Alacant (província d'Alacant valencianoparlant, menys les comarques centrals), Alcoi-Gandia (les anomenades *comarques centrals*), València i AM (ciutat de València i la seua àrea metropolitana), València (província de València, menys les poblacions incloses en les zones Alcoi-Gandia i València i AM) i Castelló (província de Castelló valencianoparlant). Per explicar el càlcul, partim dels resultats de, per exemple, la regió sociolingüística d'Alacant en el cas de l'enquesta del 2005 (SIES-CEdCEs, 2005), pel que fa a l'ús a la llar, que es recullen a la taula 1, segona columna. Hi hem afegit (columnes tercera i quarta) uns factors de ponderació, que corresponen a intervals d'1/6. Òbviament, aquesta atribució és convencional, però la considerem prou ajustada, i preferible a considerar usuàries les persones que contesten *sempre* i *generalment* (o afegir-hi *més… que…*). També proposem una ponderació de la competència amb els factors 4/4: *coneix perfectament…*, 3/4: *coneix bastant bé…*, 1/4: *coneix un poc…* i 0/4: *no coneix gens…* Així mateix, hem hagut de considerar, per manca de dades, que *N* és constant en totes les enquestes de la sèrie, la qual cosa no és cap suposició excessiva, atenent les projeccions demogràfiques de l'Institut Nacional d'Estadística.

Taula 1
Ús del valencià a la llar, regió d'Alacant (zona valencianoparlant) (2005)

	Enquesta	*Ponderació (ús valencià)*	*Ponderació (ús castellà)*	*Estimació (U)*	*Estimació (P-U)*	*Total (P)*
Sempre valencià	10,9 %	6/6	0/6	116.437	0	
Generalment valencià	4,4 %	5/6	1/6	39.168	7.834	
Més valencià que castellà	2,3 %	4/6	2/6	16.380	8.190	
Indistintament	3,9 %	3/6	3/6	20.830	20.830	
Més castellà que valencià	3,4 %	2/6	4/6	12.107	24.213	
Generalment castellà	10,6 %	1/6	5/6	18.872	94.360	
Sempre castellà	61,0 %	0/6	6/6	0	651.620	
Altres llengües	1,9 %					
NS/NC	1,9 %					
				223.794	807.047	1.030.841
N = 1.068.229					50 % total	515.420

Font: Elaboració pròpia a partir de SIES-CEdCEs (2005).

Com que $U < (P - U)$, hi apliquem les fórmules [24] i [25], que podem escriure conjuntament d'aquesta manera:

$$\mathrm{E} = \ln\left[\ln\frac{P!}{U!(P-U)!}\right] + 2\left\{\ln\left[\ln\frac{P!}{\frac{P}{2}!\frac{P}{2}!}\right] - \ln\left[\ln\frac{P!}{U!(P-U)!}\right]\right\} \qquad [26]$$

Encara que semble complicada, atés que [26] presenta la forma general $A + 2(B - A)$, i que $A + 2(B - A) = 2B - A$, aleshores [26] es pot simplificar com:

$$\mathrm{E} = 2\left[\ln\left[\ln\frac{P!}{\frac{P}{2}!\frac{P}{2}!}\right]\right] - \ln\left[\ln\frac{P!}{U!(P-U)!}\right] \qquad [27]$$

Per al cas de l'exemple (amb desviacions petites provocades pels decimals de les ponderacions de la taula 1), tenim:

$$\mathrm{E} = 2\left[\ln\left[\ln\frac{1.030.841!}{\frac{1.030.841}{2}!\frac{1.030.841}{2}!}\right]\right] - \ln\left[\ln\frac{1.030.841!}{223.794!(1.030.841-223.794)!}\right] \qquad [28]$$

Si simplifiquem [28], obtenim:

$$\mathrm{E} = 2\left[\ln\left[\ln\frac{1.030.841!}{515.420!515.420!}\right]\right] - \ln\left[\ln\frac{1.030.841!}{223.794!807.047!}\right], \qquad [29]$$

la qual cosa equival a la instrucció següent en la plataforma *wolframalpha.com*:

2Log[Log[1030841!/(515420! 515420!)]] - Log[Log[1030841!/(223794!807047!)]] (en la plataforma, *Log* és l'abreviatura de *ln*).

Que proporciona el resultat (també a partir d'una estimació que fa la plataforma):

E = 13.760659887617773299842038617553174976724512584880321352584720466…

Farem servir només tres decimals; d'aquesta manera podem formar la taula 2.

Taula 2
Valor de E a la llar

	Alacant (A)	*Alcoi-Gandia (AG)*	*València i AM (VAM)*	*València (V)*	*Castelló (C)*	*Total (Tt)*
2005	13,760	12,760	14,382	12,824	12,820	66,546
2010	13,871	12,802	14,471	12,878	12,766	66,788
2015	13,861	12,794	14,357	12,910	12,795	66,717
2021	13,930	12,787	14,582	12,951	12,805	67,055

Font: Elaboració pròpia a partir de SIES-CEdCEs (2005, 2010, 2015 i 2021).

Com que la suma de logaritmes equival al logaritme del producte, també podríem compondre els valors de la columna de la dreta (Tt_i) d'aquesta manera:

$$Tt_i = \ln\left\{ \begin{array}{c} \ln\dfrac{P_A!}{U_A!\left(P_A - U_A\right)!} \cdot \ln\dfrac{P_{AG}!}{U_{AG}!\left(P_{AG} - U_{AG}\right)!} \cdot \ln\dfrac{P_{VAM}!}{U_{VAM}!\left(P_{VAM} - U_{VAM}\right)!} \cdot \\ \ln\dfrac{P_V!}{U_V!\left(P_V - U_V\right)!} \cdot \ln\dfrac{P_C!}{U_C!\left(P_C - U_C\right)!} \end{array} \right\} \quad [30]$$

Atés que el nostre objectiu era compondre un indicador que integrara variables relatives (taxes) i absolutes (població), ara estem en disposició de comparar els resultats de la taula 2, pel que fa a les columnes de les cinc regions sociolingüístiques, les corresponents *TU* i la seua població (cal dir que, encara que hem considerat la població constant, les diferents ponderacions fan que aquesta presente diverses modificacions). Si apliquem el coeficient de Pearson (R) al conjunt dels valors de les cinc regions en els quatre anys ($N = 20$), la correlació amb les *TU* és $R = -0{,}850$ (lògicament, la correlació és inversa perquè amb una taxa d'ús del valencià més alta tenim una entropia més xicoteta, i viceversa) i amb la població és $R = 0{,}938$, els quals es poden considerar resultats molt satisfactoris per a la construcció d'un indicador unitari.

A continuació, a la taula 3, aportem el càlcul de les dades de l'ús de la llengua al carrer, amb persones desconegudes.

Taula 3
Valor de E al carrer

	Alacant	*Alcoi-Gandia*	*València i AM*	*València*	*Castelló*	*Total*
2005	14,115	12,890	14,586	12,991	12,875	67,457
2010	14,110	12,896	14,562	12,998	12,926	67,492
2015	14,046	12,893	14,494	12,984	12,851	67,268
2021	14,386	12,901	14,848	13,059	13,039	68,233

Font: Elaboració pròpia a partir de SIES-CEdCEs (2005, 2010, 2015 i 2021).

En aquest cas, l'indicador també és satisfactori perquè la correlació amb les *TU* és $R = -0{,}845$ (lògicament, també és inversa) i amb la població és $R = 0{,}905$.

Com es pot veure, les taules 2 i 3 mostren increments globals de l'entropia.

4. Conclusions

Hem partit del que podem anomenar un *model simple*. N'hem explicat el perfeccionament i, per mitjà de la factorització de la funció, hem descobert dues tendències latents ben interessants, representades per dues funcions més. Hem deixat plantejada la qüestió del significat de la multiplicació de funcions, però el cas és que la factorització ens ha permés avançar en dues direccions: d'una banda, hem pogut intuir la singularitat del punt d'inflexió Я, que mereix un tractament en un altre article (però que, en tot cas, fa pensar que els processos de desaparició de les llengües presenten bifurcacions amb llindars a partir dels quals s'accelera l'extinció: una *catàstrofe cúspide*); de l'altra, la factorització ens ha introduït en la noció d'entropia. Hem arribat a aquesta noció també des de la necessitat de combinar variables relatives i absolutes. Això ha permés, segons el nostre objectiu, formular un indicador E, que no només aporta correlacions elevades amb les variables relatives o absolutes, sinó que també ens ofereix informacions interessants quan el calculem respecte de l'ús del valencià/català a la llar i al carrer, com ara l'*augment constant de l'entropia lingüística*. D'aquesta manera, es pot transitar a un model més complex i esbrinar els factors neguentròpics. Ara ja podem recuperar la qüestió sociolingüística del significat d'aquella multiplicació de funcions: ens permet intuir que, per sota del que hem anomenat *model simple*, hi ha una tensió entre una funció que apunta a la igualtat ideal de l'índex de Gini, o l'entropia zero, i una altra que més aviat *amaga* una singularitat cúspide i permet registrar, en el cas del valencià, un increment constant de l'entropia (definida ara sense la simetria habitual). Altrament dit, podem avançar cap a un model més dinàmic. Queda pendent formular amb més detall aquest model de relació entre competències i usos a partir del càlcul entròpic, així com desenvolupar l'aplicació de les teories de les singularitats o bifurcacions; també, revisar alguns elements d'aquesta aportació, com ara els coeficients de ponderació emprats, però considerem fonamentat l'ús de l'indicador unitari E.

5. Agraïments

Al professor Rafael Castelló, que em va fer veure el desplaçament temporal de les respostes *sempre en valencià* a *generalment en valencià*; d'aquestes, a *més valencià*, etc., una dinàmica que aquest article vol formular o quantificar. Al professor Natxo Sorolla, pel seu estímul constant a la publicació d'aquest article.

Bibliografia

Boss, V. (2005). *Intuición y matemática*. Moscou: Krasand.

— (2009). *Lecciones de matemáticas*: *Ecuaciones diferenciales*. Moscou: Krasand.

Costa, J.; Petit Cahill, K. (2021). «Revitalisation linguistique». *Langage et société* [en línia], 2021/HS1, p. 305-309. <https://doi.org/10.3917/ls.hs01.0306>.

Georgescu-Roegen, Nicholas (1971). *The entropy law and the economic process*. Cambridge, Mass.: Harvard University Press.

Guts, A. K.; Frolova, Yu. V.; Páutova, L. A. (2013). *Métodos matemáticos en la sociología*. Moscou: Krasand.

Hernàndez, Francesc J. (2020*a*). «La relació entre competència (oral activa) i ús (públic): un model matemàtic». *Treballs de Sociolingüística Catalana* [en línia], núm. 30 (juliol), p. 235-248. <https://raco.cat/index.php/TSC/article/view/374476> [Consulta: 22 octubre 2020].

— (2020*b*). «Adaptació del coeficient de GINI per a la formulació d'un coeficient de desigualtat lingüística». *Arxius de Sociologia* [en línia], núm. 42, p. 235-263. <https://roderic.uv.es/handle/10550/76706> [Consulta: 22 juny 2020].

Penrose, Roger (2006). *El camino a la realidad*: *Una guía completa de las leyes del universo*. Barcelona: Debate: Random House Mondadori.

Prigogine, Ilya (2019). *Las leyes del caos*. Barcelona: Crítica.

Querol, Ernest (1997). «Un nou model per a l'estudi del comportament lingüístic: la teoria de les catàstrofes», *Caplletra*, núm. 21 (primavera), p. 161-184.

SIES-CEdCEs = Servei d'Investigació i Estudis Sociolingüístics. Conselleria d'Educació, Cultura i Esport (2005). *Enquesta 2005. Sobre coneixement i ús social del valencià (síntesi de resultats)* [en línia]. València: Generalitat Valenciana. Conselleria d'Educació, Investigació, Cultura i Esport. <https://ceice.gva.es/va/web/dgplgm/enquestes-situacio-valencia> [Consulta: 11 octubre 2018].

— (2010). *Enquesta 2010. Sobre coneixement i ús social del valencià (síntesi de resultats)* [en línia]. València: Generalitat Valenciana. Conselleria d'Educació, Investigació, Cultura i Esport. <https://ceice.gva.es/va/web/dgplgm/enquestes-situacio-valencia> [Consulta: 11 octubre 2018].

— (2015). *Enquesta 2015. Sobre coneixement i ús social del valencià (síntesi de resultats)* [en línia]. València: Generalitat Valenciana. Conselleria d'Educació, Investigació, Cultura i Esport. <https://ceice.gva.es/va/web/dgplgm/enquestes-situacio-valencia> [Consulta: 11 octubre 2018].

— (2021). *Enquesta 2021. Sobre coneixement i ús social del valencià (síntesi de resultats)* [en premsa]. València: Generalitat Valenciana. Conselleria d'Educació, Investigació, Cultura i Esport.