

Dues notes matemàtiques sobre la llengua: sobre el Punt de No Retorn i sobre la “Massa” de la llengua

Francesc J. Hernández¹

Resum

La comunicació presenta dues aproximacions matemàtiques innovadores aplicades a la sociologia de la llengua catalana.

La primera part se centra en el concepte del Punt de No Retorn. A partir d'una funció esmenada $\mathcal{F}(TC)$ que relaciona la Taxa de Competència (TC) amb la Taxa d'Ús Màxim (TU_{max}), s'extreu una funció més interessant anomenada $\mathcal{H}(TC)$. En representar gràficament $\mathcal{H}(TC)$, s'observa un punt d'inflexió, designat com el Punt de No Retorn. Aquest punt es determina calculant la segona derivada de $\mathcal{H}(TC)$ i igualant-la a zero. El valor real resultant per a TC és 0,27850, que correspon a les coordenades (0,27850, 0,11446). La presència d'un punt d'inflexió sobtat obre la possibilitat d'una singularitat, que l'estudi vincula directament amb la Teoria de les Catàstrofes, en particular la catàstrofe cúspide. Aquesta perspectiva difereix de models que suggereixen una minva progressiva de la llengua, com el d'Abrams i Strogatz, plantejant en canvi la possibilitat d'una desaparició singular.

La segona part de l'estudi explora una manera creativa de mesurar la “massa” del català, definida com una representació simbòlica de la seua presència i ús en diversos territoris. La metodologia utilitza conceptes matemàtics com ara la densitat, el volum i la Funció omega de Boltzmann. S'apliquen càlculs per determinar el radi d'un cercle equivalent a l'extensió geogràfica d'un territori (NUTS2), el volum d'un sòlid de revolució generat per una funció de decaïment, i una adaptació de la funció omega de Boltzmann per representar la “presència” de la llengua en funció del nombre de persones competents o usuàries, tant en termes absoluts com relatius. Finalment, la massa es calcula com el producte del logaritme natural de la funció omega i el volum. Els resultats quantifiquen la massa per a diferents territoris, amb Catalunya mostrant la massa més elevada. L'estudi destaca els notables coeficients de correlació trobats entre les variables de competència/ús i els valors de massa, suggerint que aquesta metodologia teòrica és vàlida per quantificar fenòmens complexos de la sociologia de la llengua i que podria adaptar-se a altres fenòmens socials o culturals.

Paraules clau

Punt de No Retorn / Teoria de les Catàstrofes / Massa del català / Funció omega de Boltzmann / Sociologia de la llengua

¹ València, 1959. Professor jubilat del Departament de Sociologia i Antropologia social de la Universitat de València. Doctorat de sociologia amb la tesi “El valencià a València”. Més informació: <http://www.uv.es/fjhernan>

*Mathematics is not about numbers, equations, computations,
or algorithms: it is about understanding.*

WILLIAM THURSTON

Introducció

La sociolingüística (catalana) hauria de ser un pont entre les dues cultures (segons la noció clàssica de C. P. Snow), la de ciències i la de lletres, però no ho és. Això es demostra per dos fets: d'una banda, són escassos els usos de les matemàtiques que vagen més enllà de l'estadística descriptiva simple (un repertori de percentatges); d'una altra banda, i potser com a conseqüència, o més bé causa, la disciplina no ha fet una tasca, imprescindible per a l'establiment d'una ciència, de "modelització", que ha de tindre necessàriament base matemàtica, i a la qual no es pot arribar per la inducció d'estudis contingents (pel conegut problema lògic de la inducció incompleta).

Si contemplem el cas de la física, que ha servit de model als últims segles per a la constitució de les ciències, i més concretament la física quàntica, observarem un fet important. Moltes teories físiques fan servir desenvolupaments matemàtics que es van formular dècades abans i que no tenien, en principi, una aplicació a la física de les partícules. Per exemple: els grups de Lie, introduïts el 1870, han sigut emprats per la teoria de Garret Lisi, formulada a partir de 2007; la geometria de Langlands, desenvolupada en 1967-70, es fa servir ara en les teories d'Edward Witten i Anton Kapustin, també amb publicacions a partir de 2007; Felix Finster proposa una reinterpretació del vell Espai de Hilbert i Mikko Partanen i Jukka Tulkki tornen sobre les simetries de gauge i la vella geometria de Minkowski (cf. Hernández, García & Hernández 2025). Què volem dir amb això? Que una de les coses més profitoses per al progrés de la ciència sociolingüística potser siga, en lloc de repetir nocions obsoletes² o

² En els principis de la sociolingüística catalana (Aracil i Ninyoles), tingué una importància decisiva la noció de *diglòssia*, que explicaria el retrocés de la llengua. L'alternança de català i castellà als sainets, en seria un argument. A hores d'ara, el DAVL la defineix així: «Situació en què un idioma o un parlar és usat per a funcions formals, predominantment en l'ús escrit, enfront d'un altre idioma o parlar utilitzat per a

acumular recerques contingents, estimular la creativitat matemàtica. Aquest és l'objectiu del text següent que presenta dos *assaigs*, en el sentit propi del terme, dos intents agosarats i independents.

I. Sobre el Punt de No Retorn

Com mor una llengua? S'extingeix paulatinament, amb una corba de persones usuàries (representada a l'eix Y) asimptòtica a l'eix temporal (X)? O n'hi ha un punt d'inflexió a partir del qual es precipita en una dinàmica d'abandonament, de crisis irreversible? Val a dir: podem senyalar el un Punt de No Retorn lingüístic com aquell punt d'inflexió? Sembla que la sociolingüística, sobretot a partir del conegut article d'Abrams i Strogatz (2003), opera, no sabem si de manera inconscient, amb un escenari de desaparició de les llengües que podem dir asimptòtic. En aquest text mostrarem un desenvolupament matemàtic sobre la taxa de competència i el màxim ús possible d'una llengua en contextos multilingües que apunta més bé a un punt d'inflexió, a un Punt de No Retorn, que, fins i tot, podem determinar.

1.- Definicions

Partim d'una societat multilingüe de N individus, dels quals C són competents en la llengua Λ , de manera que: $N > C$.

La qüestió de quantes interaccions lingüístiques es poden fer en la llengua C és el típic problema estadístic de traure dos boletes del mateix color d'una caixa de boletes de

funcions informals, generalment orals». Ara bé, l'àrab és una llengua diglòssica per definició, i no sembla que estiga necessitada d'un procés de revitalització.

diversos colors. En el cas de les relacions lingüístiques d'individus dos a dos, que anomenarem *diàdiques*, les interaccions possibles (IP_2) estan definides per:

$$IP_2 = \frac{N!}{2!(N-2)!}$$

Como hom sap, el factorial de x , que es representa $x!$, és el producte de x pels inferiors fins a la unitat, de manera que:

$$x! = x(x-1)(x-2) \dots [x-(x-1)] = x(x-1)(x-2) \dots (1)$$

Per la qual cosa:

$$IP_2 = \frac{(N)(N-1)(N-2)(N-3) \dots (2)(1)}{(2)(1)(N-2)(N-3) \dots (2)(1)} = \frac{(N)(N-1)}{2}$$

D'altra banda, les interaccions possibles en la llengua Λ seran:

$$I\Lambda_2 = \frac{C!}{2!(C-2)!} = \frac{(C)(C-1)(C-2)(C-3) \dots (2)(1)}{(2)(1)(C-2)(C-3) \dots (2)(1)} = \frac{(C)(C-1)}{2}$$

Per tant, la taxa d'ús màxima possible (TU_{max}) d'interaccions diàdiques en la llengua Λ respecte de les totals és:

$$TU_{max_2} = \frac{I\Lambda_2}{IP_2} = \frac{\frac{C!}{2!(C-2)!}}{\frac{N!}{2!(N-2)!}} = \frac{\frac{(C)(C-1)}{2}}{\frac{(N)(N-1)}{2}} = \frac{C(C-1)}{N(N-1)} \approx \frac{C^2}{N^2} = \left(\frac{C}{N}\right)^2 = TC^2$$

Per a poblacions nombroses, el signe “ $\approx$ ” indica pràcticament una igualtat. Per exemple, si $N = 5.000.000$ i $C = 2.500.000$ individus, aleshores $TC^2 = 0,25^3$ i $\frac{C(C-1)}{N(N-1)} = 0,24999994$, amb una diferència de sis centmilionèsimes.

Però això es refereix a les interaccions diàdiques. També n'hi ha de triàdiques, tetràdiques... fins a C -àdiques⁴. De manera semblant als càlculs anteriors, podem establir que si $TU_{max_2} \approx TC^2$, aleshores: $TU_{max_3} \approx TC^3$, $TU_{max_4} \approx TC^4$, etc., fins a $TU_{max_C} \approx TC^C$. La persona que llig podrà fer fàcilment la demostració. Per tant, si tenim en compte totes les interaccions (diàdiques, triàdiques... fins a C -àdiques),

³ Emprarem indistintament la notació 0,25 i la notació 25,0%.

⁴ Deixem de banda, és clar, el monòleg, la relació monoàdica, ben interessant filosòficament, però no sociolingüísticament.

aleshores:

$$TUmax = \frac{\sum I\Lambda}{\sum IP} = \frac{\sum_{i=2}^C \frac{C!}{i!(C-i)!}}{\sum_{i=2}^N \frac{N!}{i!(N-i)!}}$$

O més exactament: si anomenem p_i a la proporció *de cada tipus* d'interacció en el conjunt, aleshores:

$$TUmax = \sum_{i=2}^C p_i TC^i$$

Fins ací el plantejament característic de la teoria combinatòria (allò de les boletes de colors).

2.- Càlcul

Ara introduïrem dos corol·laris trivials:

Corol·lari [1]: $TC = 0 \Rightarrow TUmax = 0$ (si cap persona és competent, la $TUmax$ és igual a 0). L'invers seria cert també: $TUmax = 1 \Rightarrow TC = 1$, però no ens interessa ací.

Corol·lari [2]: $\sum_{i=2}^C p_i = 1$ (la suma de totes les proporcions és la unitat).

A més, podem establir la hipòtesi [1]: $p_i > p_{i+1}$ (la proporció d'una relació lingüística *i-àdica* és major que la de la relació *(i+1)-àdica*).

Els dos corol·laris i la hipòtesi s'acompleixen per a una quantitat *il·limitada* de valors de p , com ara:

- a) $p_i = \frac{1}{2^{i-1}}$
- b) $p_i = \frac{3}{2} \left(\frac{1}{i^2} \right)$
- c) $p_i = \frac{31}{20} \left(\frac{1}{i^2} \right)$
- d) $p_i = \frac{3101}{2000} \left(\frac{1}{i^2} \right)$

Prendrem el cas *d*), que aporta una millor aproximació a la hipòtesi [1], la qual cosa, fent les substitucions escaients, permet establir la fórmula [1]

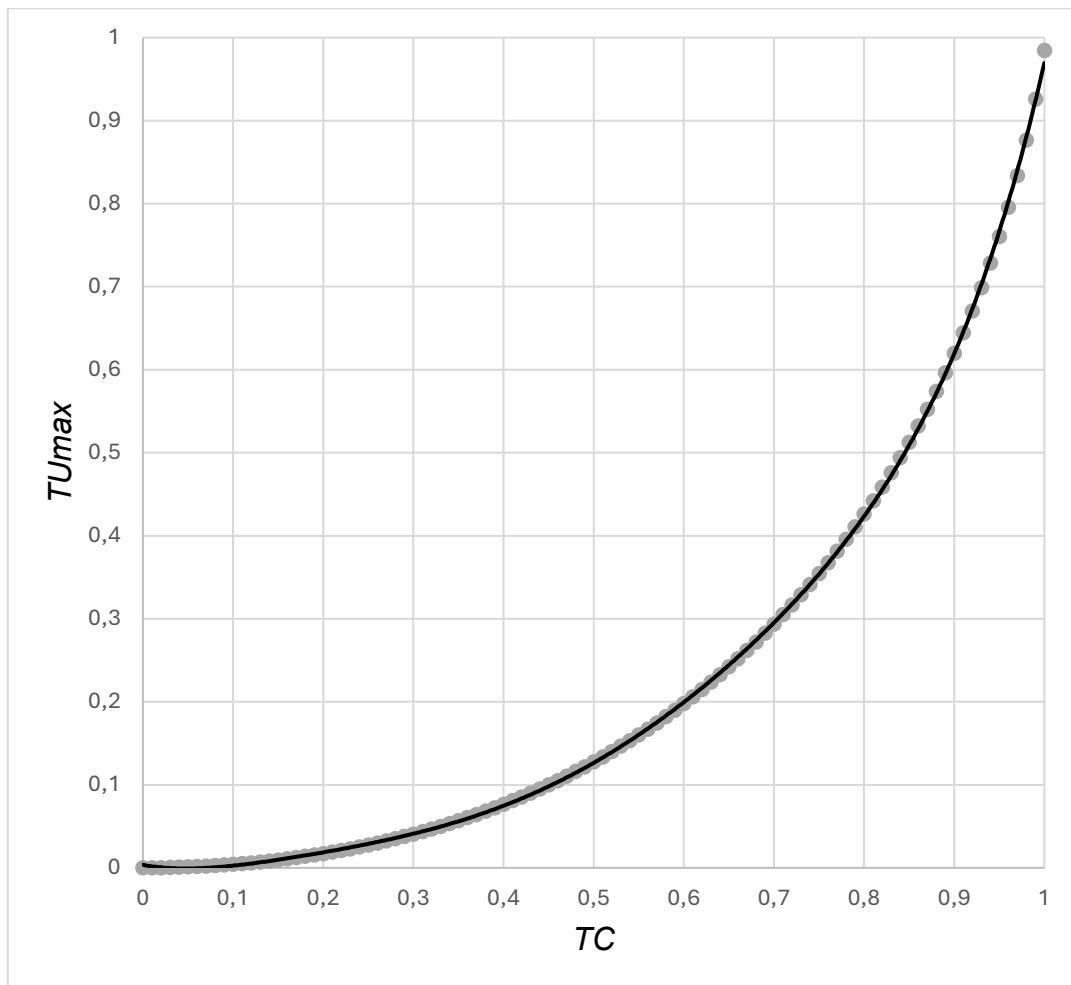
$$TU_{max} \approx \frac{3101}{2000} \left(\frac{1}{2^2} \right) TC^2 + \frac{3101}{2000} \left(\frac{1}{2^3} \right) TC^3 + \frac{3101}{2000} \left(\frac{1}{2^4} \right) TC^4 + \dots$$

$$+ \frac{3101}{2000} \left(\frac{1}{2^{c-2}} \right) TC^{c-2} + \frac{3101}{2000} \left(\frac{1}{2^{c-1}} \right) TC^{c-1} + \frac{3101}{2000} \left(\frac{1}{2^c} \right) TC^c$$

És clar que la suma pot tindre milions de sumands (en té, realment, per al nostre cas, encara que el valor de cada nou sumand minve exponencialment), per la qual cosa procedirem a la seua *simplificació*.

Hi ha prou en assignar valors i representar la corba en els eixos *X* (*TC*) i *Y* (*TU_{max}*), i traçar automàticament una línia de tendència polinòmica, com s'hi pot veure en el gràfic 1.

GRÀFIC 1. REPRESENTACIÓ DE *TU_{max}*



Font: Elaboració pròpia.

L'equació de la línia de tendència polinòmica del gràfic 1 és:

$$y = 10,753x^6 - 27,338x^5 + 26,8x^4 - 11,992x^3 + 2,9524x^2 - 0,2101x + 0,0037$$

El que vol dir que se simplifica l'anterior fórmula [1] (amb milions de sumands) amb la fórmula [2] (amb set sumands), que és concretament:

$$TUMax \approx 10,753TC^6 - 27,338TC^5 + 26,8TC^4 - 11,992TC^3 + 2,9524TC^2 - 0,2101TC + 0,0037$$

La fórmula [2] en realitat representa una *funció*, que anomenarem $\mathcal{F}(TC)$, perquè a cada valor de TC assigna un, i només un, valor de $TUMax$.

Com a conseqüència del procediment de càlcul de l'equació de la línia de tendència que hem fet servir (el del programa Excel), l'anomenat procediment de mínims quadrats, $\mathcal{F}(TC)$ presenta una desviació mínima, que contradiu el corol·lari [1] (si ningú és competent en una llengua, ningú pot usar-la). Però en la fórmula [2], si $TC = 0$, aleshores $TUMax = 0,0037$ (!). Si *eliminem aquest residu*, obtenim una altra funció, que anomenarem $\mathcal{F}(TC)$ (a partir d'ara ja farem servir només el signe d'igualtat):

$$\begin{aligned}\mathcal{F}(TC) &= TUMax \\ &= 10,757TC^6 - 27,347TC^5 + 26,809TC^4 - 11,996TC^3 + 2,9533TC^2 \\ &\quad - 0,2102TC\end{aligned}$$

Aquesta funció $\mathcal{F}(TC)$ és molt interessant perquè podem factoritzar-la fàcilment i obtindre'n:

$$TUMax = \mathcal{F}(TC) = \mathcal{G}(TC) \cdot \mathcal{H}(TC)$$

El que s'acompleix en el cas que $\mathcal{G}(TC)$ i $\mathcal{H}(TC)$ siguin respectivament:

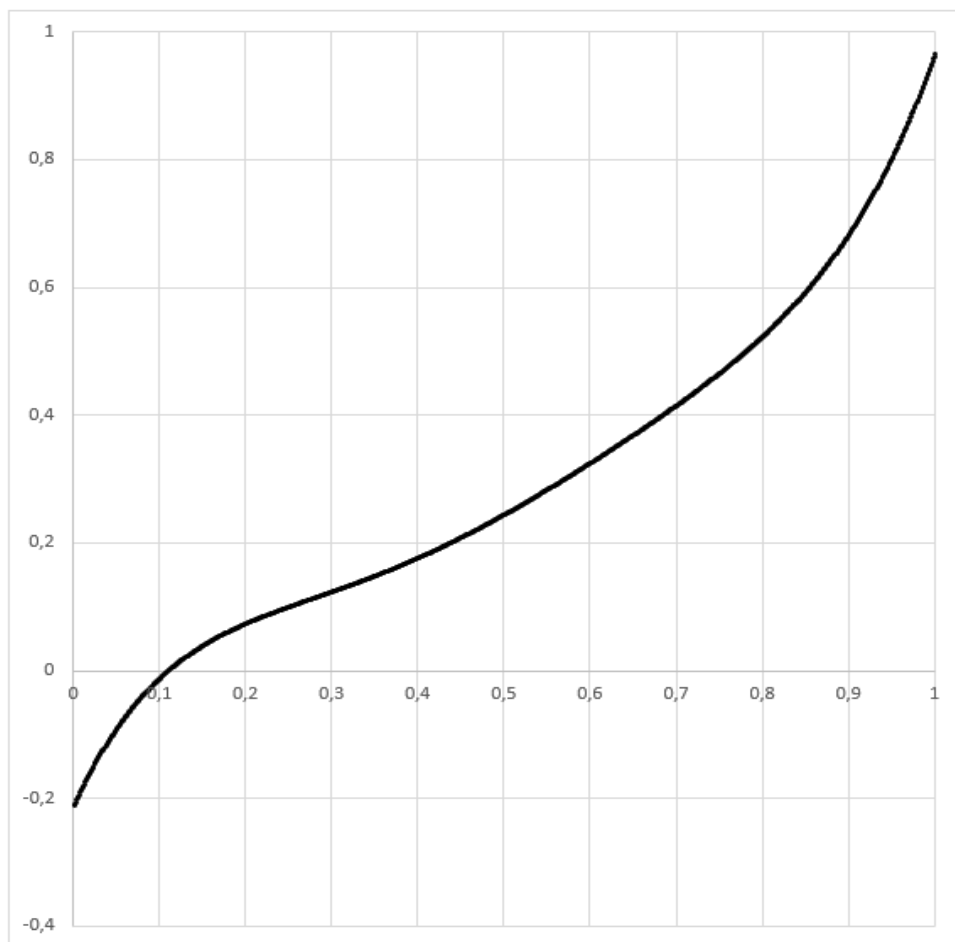
$$\begin{aligned}\mathcal{G}(TC) &\Rightarrow TUMax = TC \\ \mathcal{H}(TC) &\Rightarrow TUMax \\ &= 10,757TC^5 - 27,347TC^4 + 26,809TC^3 - 11,996TC^2 + 2,9533TC \\ &\quad - 0,2102\end{aligned}$$

Comentarem a continuació aquestes dues funcions resultants. No entrarem en la qüestió de què significa una funció que és el resultat d'una multiplicació de funcions.

La corba definida per la funció $\mathcal{G}(TC)$ seria la línia recta que uneix els punts $(0, 0)$ i $(1, 1)$, val a dir, la diagonal que anomenem creixent del quadrat unitat⁵, perquè cada valor de TU_{max} és igual que cada valor de TC . Quina interpretació hauríem de donar a aquesta igualtat segons la sociolingüística? El cas més simple en el que podríem aplicar la funció seria el de la societat 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 això la TU_{max} és igual que cada valor de TC . Altrament dit, un cas de segregació lingüística.

Més interès que la funció $\mathcal{G}(TC)$ presenta, per a la sociolingüística, la que hem anomenat $\mathcal{H}(TC)$. Si fem una representació d'aquesta funció (pel procediment de donar valors a X i calcular Y) arribem a una corba interessant, que s'arplega en el gràfic següent.

GRÀFIC 2. REPRESENTACIÓ DE $\mathcal{H}(TC)$



⁵ Anomenem quadrat unit al que es forma en un pla cartesià i que té com a costat una unitat gràfica. Els vèrtex són els punts $(0, 0)$, $(1, 0)$, $(1, 1)$ i $(0, 1)$.

3. Resultats

Com s'hi pot veure, la corba presenta un *punt d'inflexió*, que anomenarem *Punt de No Retorn* i que representarem mitjançant l'última lletra de l'alfabet rus Я (també podem fer servir l'última lletra grega, ω , però aquesta té un ús en la fórmula de Boltzmann que veurem més endavant). El valor d'aquest punt d'inflexió es pot establir amb el procediment habitual en matemàtiques, que exigeix la introducció de la noció de *derivada*.

Com és sabut, la derivada d'una funció (posem-ne: y) és una altra funció, que podem anomenar més pròpiament derivada *primera* (y'). Com que la derivada primera és una funció, podem calcular també la seua derivada, i aleshores tenim una derivada *segona* de la funció original (que es representa y''), i així successivament.

Si tenim una funció que presenta una corba amb un o més punts d'inflexió, els podem establir a partir d'aquells valors que fan que la derivada segona de la funció original siga equivalent a 0. Farem servir aquest procediment per calcular el punt d'inflexió de la corba del gràfic 2.

En primer lloc, calculem la derivada primera de la funció $\mathcal{H}(TC)$:

$$\begin{aligned}\mathcal{H}' &= (5)(10,757)TC^4 - (4)(27,347)TC^3 + (3)(26,809)TC^2 - (2)(11,996)TC \\ &\quad + 2,9533\end{aligned}$$

Després calculem la derivada segona:

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

A continuació es calculen els valors que fan $\mathcal{H}'' = 0$. Per a esbrinar quin valor de TC resol l'equació:

$$215,14TC^3 - 328,164TC^2 + 160,854TC - 23,992 = 0$$

Podem fer servir diversos procediments. El resultats que se'n poden obtindre són:

$$TC = 0,27850$$

$$TC = 0,62342 - 0,10846 i$$

$$TC = 0,62342 + 0,10846 i$$

On i és la base de números irracionals, $i = \sqrt{-1}$ (no res a veure amb l'ús anterior de la mateixa lletra per a indicar un sumand qualsevol). Només el primer valor, però, és un número real ($TC \in \mathbb{R}^+$, un número real positiu). Com que disposem del valor de TC , podem calcular el valor de $\mathcal{H}$ amb la fórmula anterior.

D'aquesta manera podem esbrinar que el punt $\mathcal{A}$ correspon a les ordenades (0,27850, 0,11446). És a dir: per baix d'una $TC = 27,850\%$ el màxim ús possible cau sobtadament.

Ara bé, aquest mateix punt d'inflexió haurem de *suposar-lo* per al subconjunt complementari (val a dir $N - C$), per la qual cosa la corba del gràfic 2 haurà de ser modificada, amb el postulat d'un *altre punt* d'inflexió, que correspondria al que podríem anomenar de manera imprecisa *simètric* de $\mathcal{A}$ (respecte de la diagonal decreixent del quadrat unitat). Si es tractara d'una corba parabòlica, el simètric del punt (x, y) seria el punt $(1 - y, 1 - x)$, aleshores el simètric, que podríem representar com $\mathcal{A}^{-1}$, seria (0,88553, 0,72140), però no tenim per què acceptar el supòsit de la corba parabòlica (val a dir, de la corba simètrica respecte de la diagonal decreixent).

4. Discussió

Aquest punt d'inflexió sobtat obri la possibilitat d'una *singularitat*. L'estudi de les *singularitats* matemàtiques es va popularitzar amb l'anomenada *teoria de les catàstrofes*,

L'aplicació de la teoria de les catàstrofes a la sociologia de la nostra llengua ja fou proposada per Querol (1997). El que s'aporta ací s'ha de considerar una continuació en aquell esforç i, sobretot, de la concepció subjacent sobre la desaparició *singular* d'una llengua i no una progressiva minva que produiria, si representàrem la taxa de competència

i la taxa d'ús en eixos XY , una corba asimptòtica a l'eix de les abscisses, com en el model d'Abrams i Strogatz.

Abrams i Strogatz (2003) formularen un model per a la “mort” de les llengües i l'aplicaren a nombrosos casos. El model parteix de dues llengües en competència (X, Y) i es pot enunciar amb la fórmula:

$$\frac{dx}{dt} = yP_{yx}(x, s) - xP_{xy}(x, s)$$

Fórmula en la qual t és el temps; y i x , la fracció de la població que parla Y i X (en la nostra terminologia $1 - TU$ i TU), per la qual cosa la seua suma és la unitat. I s és una mesura d'estatus relatiu ($0 \leq s \leq 1$). Els autors formularen P_{yx} i P_{xy} de la manera següent:

$$P_{yx}(x, s) = cx^a s$$

$$P_{xy}(x, s) = c(1 - x)^a(1 - s)$$

On c regula el ritme de canvi. Abrams i Strogatz trobaren “*unexpectedly*” un valor constant de l'exponent a que processos ben diferents culturalment: $a = 1,31 \pm 0,25$ (el que vol dir $\pm$ desviació estàndar).

Miralles (2014), Perucho et al. (2017) han analitzat acuradament l'aplicació d'aquest model al País Valencià i a Catalunya, mitjançant sociologia computacional i una reconstrucció de sèries estadístiques a partir de les enquestes sociolingüístiques disponibles. Les seues conclusions s'arreglen en la taula 1.

TAULA 1. RESULTATS DE L'ADAPTACIÓ DE MIRALLES I PERUCHO ET AL. DEL MODEL D'ABRAMS I STROGATZ

	a	c	s
Resultat	1,00	0,10	0,38
Resultat amb la troballa d'Abrams i Strogatz	1,31	0,06	0,27
Amb dades SIES 1992 (model blau)	1,00	0,80	0,24
Id. (model verd)	1,30	0,12	0,30
Id. (model roig)	1,50	0,15	0,33

Font: Perucho et al. (2017)

El treball de Miralles (2014), Perucho et al. (2017) sens dubte és una fita notable en la sociologia de la llengua aplicada al valencià. La persona que llig advertirà fàcilment algunes diferències entre la seua orientació i la plantejada aquí, on la *language death* s'entén com una qüestió, si es permet dir-ho així, *asimptòtica* o *discontínua*, fonamentada en la catàstrofe cúspide.

5. Conclusió

Hem fonamentat un model segons el qual $\mathcal{R} = (0,27850, 0,11446)$, val a dir, per baix d'una *TC* de 27,850%, que permet una *TU_{max}* del 11,446%, la caiguda de la *TU_{max}* es fa més intensa i provoca una desaparició no *asimptòtica* de la llengua. Amb tot, aquesta és una conclusió pendent d'una consideració del sentit de la multiplicació de funcions.

II. Sobre la «massa» del català

1. Introducció

L'escriptor i polític sir Walter Raleigh li assegurarà a la reina Isabel I d'Anglaterra que era capaç de pesar el fum. S'hi esmenta en algunes pel·lícules, com ara *Smoke*, amb guió de Paul Auster. Què pesa una llengua? Potser a la persona que llig li vinguen al cap tones i tones de llibres... No parlem de la literatura, sinó pròpiament de la llengua. Què pesa el

català? Quina és la seua massa? O també, quina massa aporten els diversos territoris del domini lingüístic a la massa total?

En aquest text explorarem una manera creativa de mesurar la “massa” del català, entesa com una representació simbòlica de la presència i l’ús d’aquesta llengua en diferents territoris. Utilitzarem conceptes matemàtics com la densitat, el volum i la funció omega de Boltzmann per fer aquests càlculs. Tot i que aquesta aproximació és teòrica, ens permet entendre com es poden quantificar fenòmens complexos com la difusió d’una llengua.

2. Metodologia

2.1. Dades dels territoris

Partim d’una taula amb els territoris on es parla català o valencià, la seua població total, el nombre de persones competents en la llengua, el nombre de persones que l’usen habitualment i l’extensió geogràfica de cada territori. Aquestes dades s’obtenen considerant l’equivalent a les regions NUTS2 d’Eurostat. NUTS2 és l’abreviatura d’unitats territorials normalitzades, que equival a les *regions* en el context de la Unió Europea i que, per tant, presenten governs i polítiques regionals. En el cas espanyol, es corresponen amb les comunitats autònomes (vegeu Eurostat 2023).

A continuació imaginarem aquestes regions NUTS2 com un sòlid de revolució, la qual cosa implica determinar un radi de la base i una funció de decaïment.

2.2. Radi d’un cercle equivalent

Per a cada territori, calculem el radi (R) d’un cercle que tinguera la mateixa superfície que el territori. La fórmula utilitzada és:

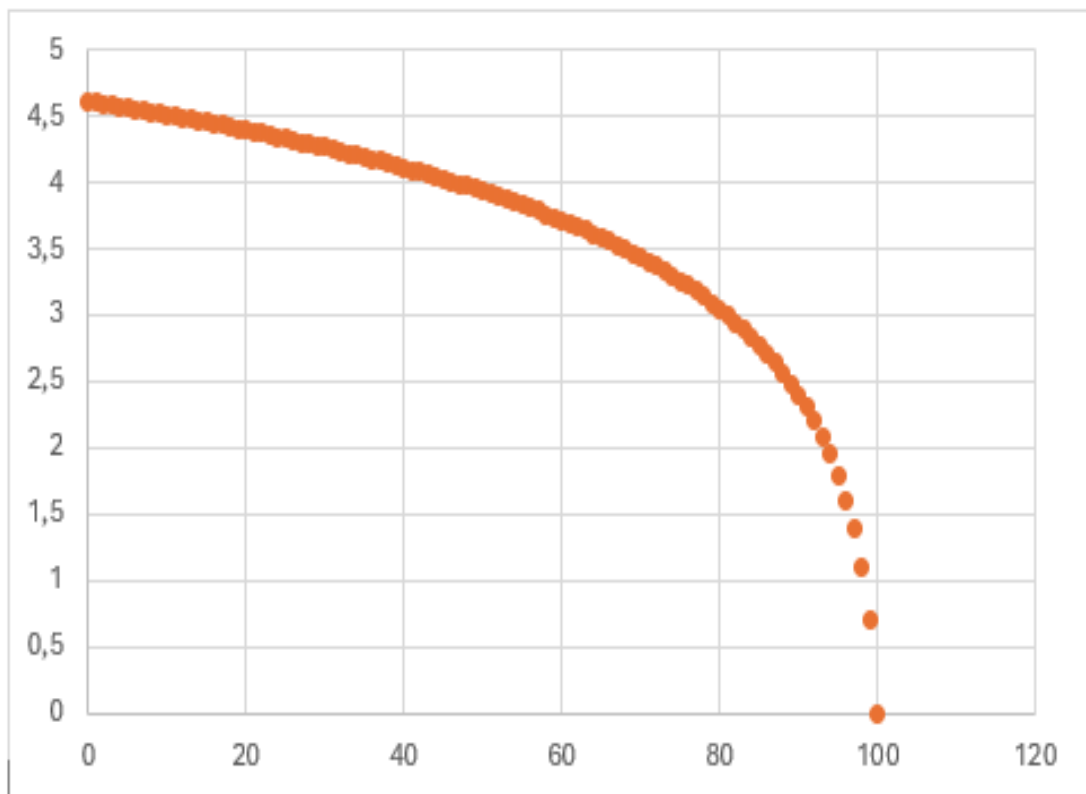
$$R = \sqrt{\frac{A}{\pi}}$$

on A és l’extensió en quilòmetres quadrats.

2.3. Funció de decaïment

Partim del supòsit d'una distribució heterogènia de la massa de la llengua, amb una funció de decaïment. Considerem una funció de decaïment $y = \ln(x)$, que representa com disminueix la presència de la llengua a mesura que ens allunyem del “centre” del cercle que representa el territori. En realitat, considerarem una versió diferent, per posar el valor màxim a l'esquerra, que coincidisca amb una distància 0 (el focus). Vegeu el gràfic 3 on es representa $y = \ln(N) - \ln(x)$, per a $N = 100$.

GRÀFIC 3. EXEMPLE DE FUNCIO DE DECAÏMENT

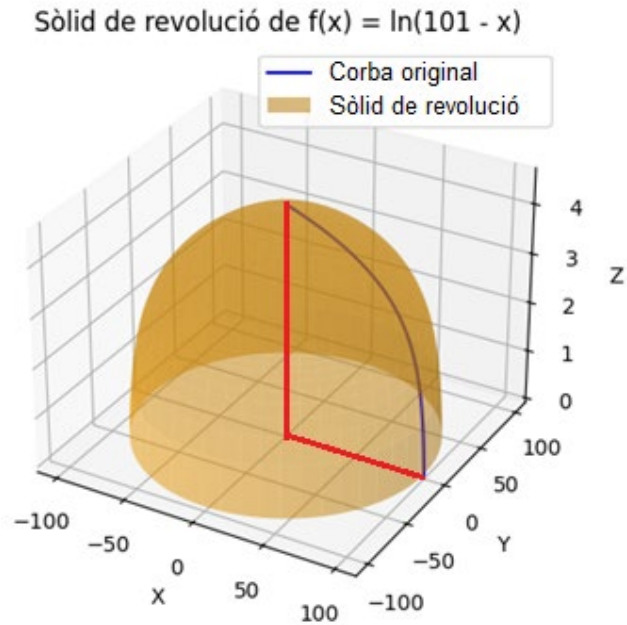


Font: Elaboració pròpia.

2.4. Volum d'un sòlid de revolució

En girar aquesta funció al voltant de l'eix vertical (Y), generem un sòlid de revolució. Vegeu el gràfic 4.

GRÀFIC 4. EXEMPLE DE SÒLID DE REVOLUCIÓ DEL GRÀFIC 3



El volum V d'aquest sòlid es calcula mitjançant la fórmula:

$$V = 2\pi \int_0^R (R - x) \cdot \ln(x) dx$$

2.5. Funció omega de Boltzmann

La funció omega de Boltzmann (ω) ens permet calcular el nombre de microestats accessibles en un sistema. En aquest cas, la adaptem per representar la “presència” de la llengua en funció del nombre de persones competents (A_C) o usuàries (A_U). La fórmula n'és:

$$\omega = \frac{N!}{A! \cdot (N - A)!}$$

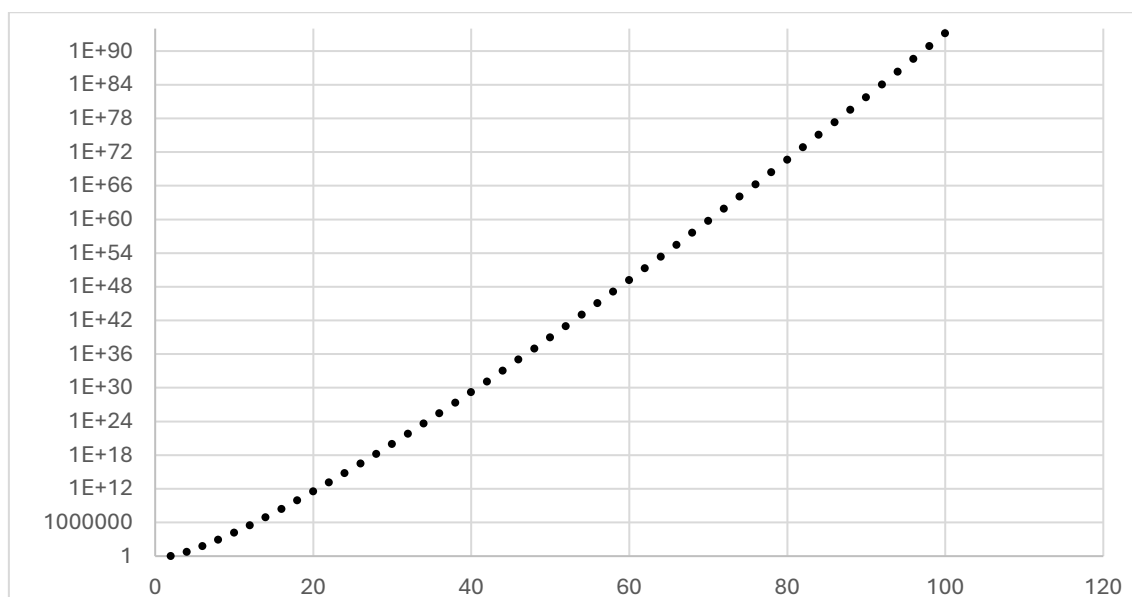
Aquesta fórmula coincideix amb el càlcul combinatori (Stanley, 2011), que representa la quantitat de formes en què A individus poden ser seleccionats d'un total de N (Abramowitz & Stegun, 1964; Graham, Knuth & Patashnik, 1994).

Per què hem d'usar aquesta fórmula i no quantitats absolutes o relatives (percentatges) com es fa habitualment? La raó és senzilla. En sociolingüística són importants les taxes, però també la quantitat total. Per exemple, el mateix percentatge de persones usuàries d'una llengua en una comunitat gran pot permetre el desenvolupament d'una literatura de qualitat o l'impuls a l'edició, el que no esdevé en una comunitat xicoteta. En altres treballs hem mostrat com aquesta fórmula presenta correlacions destacades amb les taxes i les quantitats absolutes (Hernández 2023).

L'ús d'aquesta fórmula planteja alguns problemes, que es comenten i es resolen a continuació.

a) En primer lloc ω pot assolir valors molt elevats. Com es pot deduir fàcilment, per a un determinat valor de N el màxim de ω s'assoleix quan $A = N/2$. Per a donar una idea de la velocitat de creixement de ω , vegeu la gràfica següent on N oscila de 2 a 100, i sempre $A = N/2$. Per a poder visualitzar la corba, els valors en l'eix Y són representats segons una escala logarítmica.

GRÀFIC 5. CREIXEMENT DE ω



Font: Elaboració pròpia. Como és sabut, $1E+60$ significa 1×10^{60} , és a dir 1 i 60 zeros.

Per a solucionar aquest problema del càlcul, farem servir el logaritme natural, és a dir, $\ln(\omega)$.

b) La raó per la qual ω pot adoptar valors tan alts rau en el càlcul de factorials, que pot arribar a valors astronòmics. Per exemple:

$$1000! = 4,023 \times 10^{2567}$$

(Cal pensar que, per exemple, $81!$ ja supera el grandíssim Nombre de Shannon)

Per a poder treballar amb valors tan grans es poden emprar alguns recursos. El primer, d'índole matemàtica, es recórrer a l'*aproximació de Stirling*:

$$\ln(N!) \approx N \ln(N) - N$$

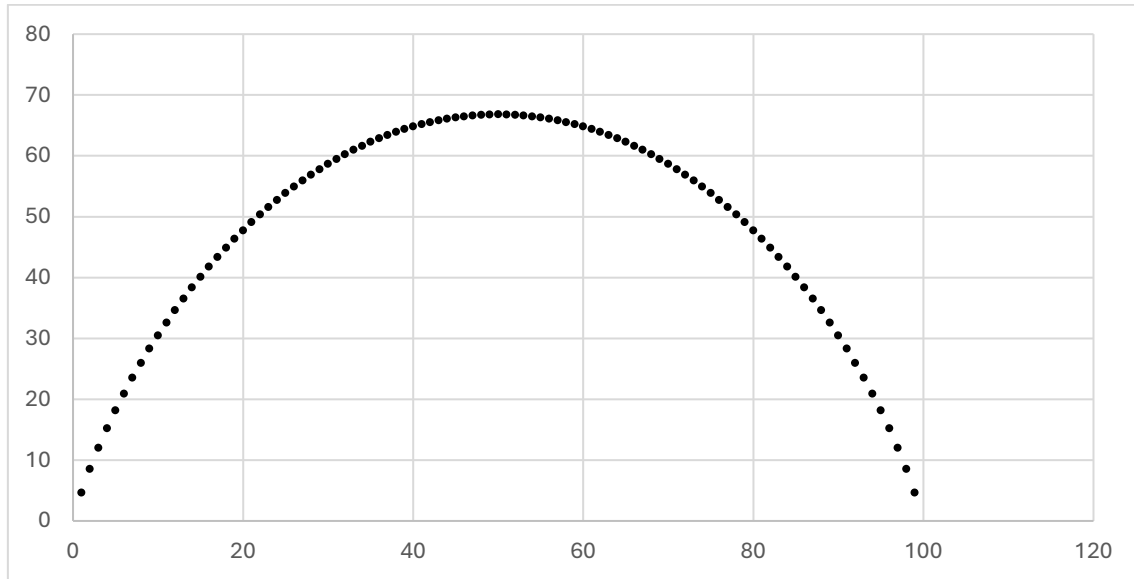
Però tal vegada resulte més pràctic i ràpid fer servir portals matemàtics (com ara <https://wolframalpha.com/>), que tal vegada facen servir l'aproximació de Stirling) o recórrer a la Intel·ligència Artificial Generativa.

c) El tercer problema és que la funció ω és simètrica respecte del valor màxim, ja que, com resulta trivial, tant per al cas d' A com per al del seu complementari $N - A$ s'acompleix que:

$$\omega = \frac{N!}{A! \cdot (N - A)!} = \frac{N!}{(N - A)! \cdot A!}$$

Vegeu la gràfica següent on $N = 100$, en X es representen els valors $1 < A < 99$ i en Y el resultat de $\ln(\omega)$ (naturalment, se eliminen els valors 0 i 100, perquè $\ln(0)$ no hi té sentit).

GRÀFIC 6. SIMETRIA DE ω



Font: Elaboració pròpia

Per a trencar la simetria s'ha d'introduir una correcció, que és la següent:

$$\omega = \begin{cases} Si A \leq \frac{N}{2} \Rightarrow \frac{N!}{A! \cdot (N-A)!} \\ Si A > \frac{N}{2} \Rightarrow 2 \left[\frac{N!}{\left(\frac{N}{2}!\right)^2} \right] - \frac{N!}{A! \cdot (N-A)!} \end{cases}$$

on N és la població total i A és el nombre de persones competents o usuàries.

3. Definicions de densitat i massa

La densitat D es defineix com la relació entre la massa M i el volum V :

$$D = \frac{M}{V}$$

En aquest context, fem equivalent la densitat D i $\ln(\Omega)$. Per tant, la massa es calcula com:

$$M = \ln(\Omega) \cdot V$$

4. Resultats

4.1. Taula resum

A continuació, presentem una taula amb els resultats clau per a cada territori:

Territori (NUTS2)	Extensió (km²)	Radi cercle hipotètic (km)	Població (N)	Persones competents (C)	Persones usuàries (U)
Catalunya	32.114	~101,1	7.750.000	~6.500.000	~5.800.000
Comunitat Valenciana	23.255	~86,0	5.050.000	~3.800.000	~2.500.000
Illes Balears	4.992	~39,8	1.200.000	~900.000	~800.000
Catalunya Nord (França) ¹	4.116	~36,2	480.000	~150.000	~50.000
Andorra	468	~12,2	80.000	~60.000	~50.000
Franja de Ponent (Aragó) ¹	3.600	~33,9	50.000	~30.000	~20.000
Alguer (Sardenya, Itàlia) ¹	224	~8,5	40.000	~20.000	~10.000

(1) Aquests territoris no són pròpiament NUTS2, sinó unitats menors.

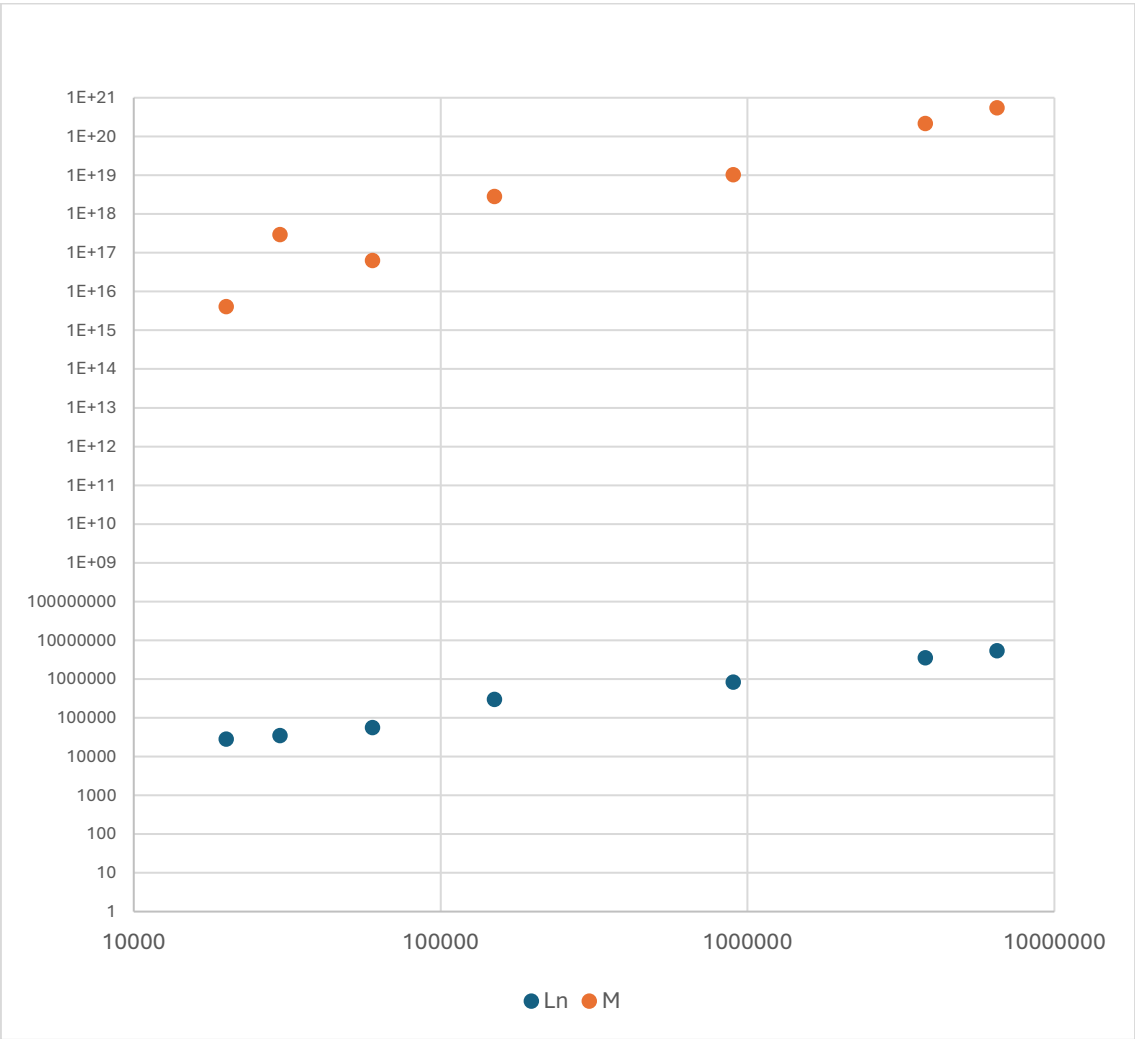
Territori (NUTS2)	C/N	U/N	Volum (V)	ln (Ω_C)	ln (Ω_U)
Catalunya	8,39E-01	7,48E-01	1,02×10 ¹⁴	5,37188E+06	5,37188E+06
Comunitat Valenciana	7,52E-01	4,95E-01	6,15×10 ¹³	3,50039E+06	3,50014E+06
Illes Balears	7,50E-01	6,67E-01	1,23×10 ¹³	8,31770E+05	8,31770E+05
Catalunya Nord (França)	3,13E-01	1,04E-01	9,45×10 ¹²	2,98115E+05	1,60380E+05
Andorra	7,50E-01	6,25E-01	1,12×10 ¹²	5,54466E+04	5,54466E+04
Franja de Ponent (Aragó)	6,00E-01	4,00E-01	8,45×10 ¹²	3,46524E+04	3,36449E+04
Alguer (Sardenya, Itàlia)	5,00E-01	2,50E-01	1,45×10 ¹¹	2,77204E+04	2,24880E+04

Aleshores podem calcular la *massa*, el valor absolut i la proporció respecte del domini:

Territori (NUTS2)	M_C	M_U	% M_C	% M_U
Catalunya	5,47931760E+20	5,47931760E+20	70,554063%	70,6748%
Comunitat Valenciana	2,15273985E+20	2,15258610E+20	27,719609%	27,7650%
Illes Balears	1,02307635E+19	1,02307635E+19	1,317357%	1,3196%
Catalunya Nord (França)	2,81718675E+18	1,51560990E+18	0,362753%	0,1955%
Andorra	6,21001920E+16	6,21001920E+16	0,007996%	0,0080%
Franja de Ponent (Aragó)	2,92812780E+17	2,84299405E+17	0,037704%	0,0367%
Alguer (Sardenya, Itàlia)	4,01945220E+15	3,26076290E+15	0,000518%	0,0004%
Σ	7,76612635E+20	7,75286411E+20		

En representar aquests resultats en un gràfic, podem advertir l'alineació dels valors de competència amb eixos logarítmics. Vegeu el gràfic 7.

GRÀFIC 7. ALINEACIÓ DELS VALORS DE COMPETÈNCIA AMB EIXOS LOGARÍTMICS



Les taules anteriors proporcionen la següent matriu de correlacions:

	C	U	C/N	U/N	$\ln(\Omega_C)$	$\ln(\Omega_U)$	M_C	M_U
C	1	0,98976	0,61508	0,56955	0,99810	0,99833	0,98108	0,98115
U		1	0,61310	0,59718	0,97981	0,98012	0,99341	0,99348
C/N			1	0,96573	0,60173	0,61741	0,56579	0,56734
U/N				1	0,54803	0,56170	0,54388	0,54522
$\ln(\Omega_C)$					1	0,99973	0,96938	0,96941
$\ln(\Omega_U)$						1	0,96904	0,96912
M_C							1	1,00000
M_U								1

En síntesi podem indicar els següents resultats de les correlacions, de major a menor:

- Correlació de C o U amb M_C o $M_U > 0,981$
- Correlació de C o U amb $\ln(\Omega_C)$ o $\ln(\Omega_U) > 0,979$
- Correlació de $\ln(\Omega_C)$ o $\ln(\Omega_U)$ amb M_C o $M_U > 0,969$
- Correlació de C/N o U/N amb $\ln(\Omega_C)$ o $\ln(\Omega_U) > 0,548$
- Correlació de C/N o U/N amb M_C o $M_U > 0,543$

Com s’hi pot veure, les correlacions són menors en el cas de les taxes.

5. Conclusió

Aquest estudi ens permet aproximar-nos a una “massa” simbòlica del català en diferents territoris, utilitzant conceptes matemàtics com la densitat, el volum i la funció omega de Boltzmann. Tot i que els càlculs són teòrics, il·lustren com es poden quantificar fenòmens complexos com la presència i l’ús d’una llengua.

Els coeficients de correlació trobats són notables.

Aquesta metodologia també podria adaptar-se per analitzar altres fenòmens socials o culturals.

Referències

- ABRAMOWITZ, M., & STEGUN, I. A. (1964). *Handbook of Mathematical Functions*. Dover Publications.
- ABRAMS, Daniel M.; STROGATZ, Steven (2003): «Modelling the dynamics of language death». *Nature*, 424: 900.
<https://www.math.uh.edu/~zpkilpat/teaching/math4309/project/nature03_abrams.pdf>
- EUROSTAT (2023). *NUTS - Nomenclature of Territorial Units for Statistics*. Disponible en: <https://ec.europa.eu/eurostat/web/nuts>
- GRAHAM, R. L., KNUTH, D. E., & PATASHNIK, O. (1994). *Concrete Mathematics*. Addison-Wesley.
- HERNÁNDEZ, Francesc J. (2023): «Formulació d'un indicador unitari de competència i d'ús de la llengua per a l'avaluació de polítiques lingüístiques». *Treballs de sociolingüística catalana*, núm. 33, pp. 141-157 [Disponible en: <https://www.raco.cat/index.php/TSC/article/download/424721/519274>]
- HERNÁNDEZ, C.; GARCIA, V. & HERNÁNDEZ, F. (2025): «La ignorancia del todo. Orientaciones de la física cuántica hacia una teoría del todo y su desconsideración curricular». *Studia Humanitatis Journal*, 2025, vol. 5, núm. 2, 21-54. <https://doi.org/10.33732/shj.v5i2.140>
- MIRALLES VILA, Clara (2014): *Models dinàmics de competició entre llengües*, Treball de Grau. <https://www.uv.es/perucho/TESI/TGF-CMiralles.pdf>
- PERUCHO I PLA, Manuel; GRIMALDO MORENO, Francisco; MIRALLES VILA, Clara; QUEROL PUIG, Ernest; LÓPEZ IÑESTA, Emilia (2017): «La modelització matemàtica dels processos de competició entre llengües: estudi comparatiu de l'evolució dels usos a Catalunya i al País Valencià», *Llengua, Societat i Comunicació*, 15 <<http://revistes.ub/index.php/LSC>>.

Ús de programes especialitzats:

Per a les dades generals, els càlculs volumètrics i el codi 3D: deepseek.com

Per als càlculs del punt d'inflexió i ω : wolphramalfa.com

Per als gràfics 1, 2 3, 5 i 6 (2D) i la matriu de correlacions: excel

Per al gràfic 4 (3D): mycompiler.io/es/new/python

16/09/25