

La terminologia de la intel·ligència artificial



Lingüística quantitativa i lleis lingüístiques: de la lingüística a la intel·ligència artificial i la tecnòtica

ANTONI HERNÁNDEZ FERNÁNDEZ
Escola d'Art i Disseny de Terrassa
Universitat Politècnica de Catalunya
ORCID: 0000-0002-9466-2704
antonio.hernandez@upc.edu

RAMON FERRER I CANCHO
Universitat Politècnica de Catalunya
ORCID: 0000-0002-7820-923X
ramon.ferrer@upc.edu

Físic, lingüista i doctor en Ciència Cognitiva i Llenguatge per la Universitat de Barcelona. Professor a l'Escola d'Art i Disseny de Terrassa i a l'Institut de Ciències de l'Educació de la Universitat Politècnica de Catalunya, on també forma part del Comitè d'Ètica.



Enginyer informàtic, doctor en Informàtica i diploma d'estudis avançats en Física Aplicada per la Universitat Politècnica de Catalunya, on és professor agregat, investigador i director del grup de recerca LQMC (Lingüística Quantitativa, Matemàtica i Computacional).

Resum

Tot i que el recorregut terminològic arrela en els mateixos orígens de la computació, l'emergència recent de les aplicacions basades en la intel·ligència artificial ha popularitzat molts ètims ja antics, a la vegada que en projecta uns altres de nous, dels quals potser sentirem a parlar més en un futur proper, alguns d'aquests, encara anglicismes sense traducció acceptada. En aquest succint article transitarem des de la tradició centenària de la lingüística quantitativa —la part de la lingüística que estudia el llenguatge i els sistemes de comunicació sota una perspectiva matemàtica— fins a la intel·ligència artificial actual —popularitzada pels avenços recents en les aplicacions generatives de text i d'imatge, que han cristal·litzat en eines informàtiques molt versàtils i efectives, que plantegen tot un repte social, especialment, a l'educació. Ho farem seguint un reguitzell de mots i de definicions que, a hores d'ara, o bé ja han entrat als nostres diccionaris,amentals i oficials, o bé truquen a la porta per fer-ho.

PARAULES CLAU: lingüística quantitativa; lleis lingüístiques; intel·ligència artificial; xatbot; prompting; tecnòtica

Abstract

Quantitative linguistics and linguistic laws: From linguistics to artificial intelligence and technoethics

Although the path that terminology has followed stems from the same origins as computing, the recent emergence of applications based on artificial intelligence has popularized many old etyma while projecting new words that we may be hearing more about in the near future, some of which are still Anglicisms lacking an accepted translation into Catalan. In this short article, we will travel from the centuries-old tradition of quantitative linguistics, the sub-discipline of linguistics that studies language and communication systems from a mathematical perspective to current artificial intelligence, popularized by recent advances in generative applications of text and images, which have crystallized into very versatile and effective software tools, posing a veritable social challenge, especially in education. We will do this by tracing a series of words and definitions that have by now either entered or are knocking on the door of our mental and official dictionaries. **KEYWORDS:** quantitative linguistics; linguistic laws; artificial intelligence; chatbot; prompting; technoethics

Introducció

La lingüística quantitativa és una subdisciplina de la lingüística que busca establir lleis generals, potencialment universals, que permetin explicar com funcionen les llengües: les anomenades lleis lingüístiques. Una llei lingüística és una regularitat estadística la qual sorgeix a escales i nivells diferents (fonemes, síl·labes, mots, frases, textos...) en les llengües humanes que es pot formular matemàticament i contrastar experimentalment (Torre et al., 2019). Es tracta, per tant, de lleis científiques de base estocàstica, semblants a les que es donen en camps com la física estadística. Així com no es pot dir amb exactitud què li succeirà a una partícula en el si d'un gas, però sí que podem fer prediccions quan tenim el nombre d'Avogadro partícules, semblantment passa amb les dades lingüístiques: tot i que no sempre, sovint cal un nombre elevat d'elements lingüístics per poder establir lleis empíriques i fer inferències.

D'altra banda, la lingüística computacional és un domini indisciplinari entre la lingüística i la informàtica en què se cerquen models del llenguatge natural que permetin processar, entendre i produir (amb generadors textuais o sistemes de síntesi de veu, per exemple) les llengües humanes. Certament, han estat els mètodes computacionals els que han empès la lingüística quantitativa, com ha succeït amb d'altres disciplines científiques. Per això, i tenint en compte el desconeixement previ de la disciplina, és comprensible que se l'hagi assimilat amb la lingüística computacional. De fet, als orígens de la lingüística quantitativa moderna, a finals de segle XIX i inicis del XX, era habitual el processament manual de quantitats de dades gegants (Zipf, 1949).

No obstant això, la lingüística quantitativa no tracta només de fer formulacions matemàtiques precises o programes informàtics eficaços: donat el seu el caràcter científic, cerca explicar el perquè de les lleis lingüístiques i els principis dels quals es deriven aquestes lleis —més enllà de les meres dades—, així com fonamentar el cos teòric de la lingüística i establir connexions epistemològiques amb d'altres ciències (Hernández-Fernández i Ferrer-i-Cancho, 2019, p. 10-11):

La lingüística quantitativa sorgeix com una disciplina empírica, efímera i ambiciosa. [...] actualment aprofita l'enorme potencial dels nostres ordinadors i la gran disponibilitat de dades que, com veurem, ja no són només textos i corpus lingüístics, sinó també genomes o senyals d'origen animal o desconegut. Per això la lingüística quantitativa és ambiciosa, com dèiem, perquè els seus mètodes i els seus conceptes lingüístics s'estan obrint pas en altres camps transdisciplinaris; però també ho és per un altre motiu: pretén transformar-se en el nucli teòric de la lingüística. Ras i curt. Sense cognoms. Això converteix la lingüística quantitativa en una ambiciosa disciplina de

trànsit, fundacional, com ho han estat altres als diferents processos d'establiment de les ciències.

Arraconada en el nostre context acadèmic, sovint desconeguda o, com a molt, diluïda dins la lingüística computacional, la lingüística quantitativa va néixer abans de l'era informàtica amb ànim d'abordar empíricament el problema del llenguatge i l'estudi de les llengües humanes (Best i Rottmann, 2017). A d'altres contrades, com a l'est d'Europa, en què el divorci educatiu entre ciències i lletres era menor i la tercera cultura s'ha cultivat més en aquest àmbit, és on la lingüística quantitativa s'ha investigat amb més profusió (Köhler, Altmann i Piotrowski, 2008).

L'estudi de les lleis lingüístiques en la biologia s'ha anat estenent a d'altres sistemes de comunicació que es donen a la natura (acústics, gestuals, químics...), a la genòmica i la proteòmica, i a l'anàlisi de les interaccions ecològiques (vegeu la revisió de Semple, Ferrer-i-Cancho i Gustison, 2022). En paral·lel, les formulacions matemàtiques i teòriques s'han relacionat de manera fructífera amb la física de sistemes complexos i la teoria de la informació, han assolit solidesa i han plantejat nous reptes, per integrar la lingüística en el nucli de les ciències (Debowski, 2020).

Lleis lingüístiques

Entre 2018 i 2021, des de la Secció de Ciències i Tecnologia de l'IEC es va desenvolupar el primer projecte de recerca sobre lingüística quantitativa en el català.¹ En aquest projecte es van investigar les lleis lingüístiques més conegudes, llistades a la taula 1, tant en corpus orals com escrits, comparant-les amb d'altres llengües com el castellà i l'anglès (Hernández-Fernández et al., 2019), entre les quals podem destacar:

- La llei de Zipf determina que la freqüència de les paraules d'un corpus es relaciona inversament, per mitjà d'una llei de potències, amb el seu rang (la paraula més freqüent, amb més aparicions al corpus, té rang 1; la segona més freqüent té rang 2, i així successivament).
- La llei de Herdan-Heaps relaciona matemàticament el creixement del vocabulari d'un corpus (el nombre de paraules diferents) amb la seva llargària, en nombre de paraules en corpus escrits o en durada temporal en el cas de corpus orals.
- La llei de brevetat estableix que els mots més freqüents a les llengües tendeixen a ser més curts, de manera que la major part d'articles, preposicions o paraules funcionals més emprades solen ser breus, tant acústicament com en la seva representació escrita en caràcters.
- La llei de Menzerath-Altmann relaciona la mida total d'una construcció lingüística amb la mida de les

seves parts o dels seus constituents, de manera que les construccions lingüístiques més llargues tendeixen a tenir constituents més breus (així, les paraules més llargues tendeixen a tenir síl·labes més curtes).

- La llei de mida-rang relaciona el rang d'una paraula, o ordenació segons la freqüència de la paraula en un corpus, amb la seva durada o longitud en nombre de caràcters (segons si analitzem textos escrits o veu).
- La llei de log-normalitat estableix la forma de les distribucions estadístiques dels elements lingüístics (fonemes, paraules, sintagmes, frases...) de les llengües, que segueixen aproximadament una funció log-normal.

La hipòtesi de treball de la lingüística quantitativa és que les lleis lingüístiques són manifestacions de principis (Zipf, 1949) que normalment es defineixen com a funcions que cal optimitzar. Així, per exemple, la llei de brevetat és conseqüència del principi de compressió, el principi de la teoria de la informació segons el qual es minimitza la llargària esperada d'un codi lingüístic (Ferrer-i-Cancho et al., 2013), i les restriccions sobre els encreuaments de dependències sintàctiques es deriven del principi de minimització de la distància entre paraules relacionades sintàcticament (Gómez-Rodríguez i Ferrer-i-Cancho, 2017).

De fet, tot i que hom pot recórrer al diccionari a cercar ràpidament les definicions que s'hi recullen, encara en alguns casos és tot un repte definir en la recerca lingüística les unitats d'estudi, com pot ser el

concepte de paraula. Així, ens quedem amb la paraula ortogràfica, definida com els caràcters que trobem entre espais en blanc? Serveix per estudiar les lleis lingüístiques en l'oralitat, en què les pauses entre mots sovint no existeixen o no són clares?

Investigant, per exemple, la llei de Menzerath-Altmann en l'oralitat, trobem que es relaciona amb les capacitats fisiològiques del nostre organisme: en paraules més llargues, n'escurcem els constituents (síl·labes), ja que ens anem quedant sense aire, progressivament. En aquest sentit, es defineix el grup respiratori (traducció literal de l'anglès *breath group*) com el conjunt de paraules que pronunciem acústicament entre les breus pauses respiratòries que efectuem normalment en la parla per agafar aire (Torre et al., 2019). S'ha determinat així que la llei de Menzerath-Altmann té una base física, fisiològica, que fa que en l'oralitat la llei sigui més robusta que en l'escriptura, en què també es troba, segurament en la mesura que l'escriptura reflecteix l'oralitat, amb les limitacions que esmentàvem de la definició de les unitats d'estudi.

I quina aplicació té tot plegat? Com en tota recerca bàsica, sovint no és immediata. Però, en aquest cas, l'anàlisi de les lleis lingüístiques en el català, a escales diferents, de l'àmbit prefonèmic als grups respiratoris (Hernández-Fernández et al., 2022), ens va permetre determinar en un treball posterior que l'anàlisi de la llei de log-normalitat en les pauses durant la parla de malalts amb demències de tipus Alzheimer, bilingües català-castellà, distingia els pacients amb dos

	Fórmula matemàtica	Variables
Llei de Zipf	$f(r) \sim r^{-\alpha}$	f: freqüència r: rang α : paràmetre
Llei de Herdan-Heaps	$V \sim L^\beta$	L: mida del text/durada V: vocabulari β : paràmetre
Llei de brevetat	$f \sim \exp(-\lambda \ell)$, $\lambda > 0$	f: freqüència ℓ : mida (durada/caràcters) λ : paràmetre
Llei de mida-rang	$\ell \sim \theta \log(r)$, $\theta = \frac{\alpha}{\lambda}$	ℓ : mida (durada/caràcters) r: rang θ : paràmetre
Llei de Menzerath-Altmann	$y(n) = an^b \exp(-cn)$	n: mida total y: mida de les parts a, b, c: paràmetres
Llei de log-normalitat	$p(t; \mu, \sigma) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\ln(t)-\mu)^2}{2\sigma^2}}$	t: durada temporal σ , μ : paràmetres

TAULA 1. Lleis lingüístiques analitzades en el català, amb les fórmules matemàtiques i les variables corresponents, seguint Hernández-Fernández et al. (2019)

subtipus de deteriorament cognitiu lleu (*mild cognitive impairment*) dels pacients amb una demència de tipus Alzheimer i dels subjectes sans o grup de control (Pastoriza-Domínguez *et al.*, 2022), de manera que s'obtenia un marcador no invasiu —derivat de l'anàlisi de la parla— per a la detecció precoç de la malaltia d'Alzheimer.

Amb la mateixa filosofia, aquest resultat ens encoratja per prosseguir la recerca clínica aplicada, en aquest cas, a un segon projecte,² a fi d'explorar les lleis lingüístiques relacionades amb el significat de les paraules i cercar si hi ha diferències en els paràmetres entre els subjectes sans i aquells que pateixen afàsia. *Grosso modo*, en afasiologia es diferencien els pacients que tenen preservada la sintaxi i els que no. Específicament, en un treball previ es van caracteritzar per al català (oral i escrit) les denominades *lleis de Zipf* per que fa al significat i la freqüència de les paraules, que indiquen estadísticament la tendència de les paraules més freqüents a tenir més significats potencials, entesos, en aquest cas, com més entrades i definicions en el diccionari normatiu, seguint una formulació matemàtica precisa (Català *et al.*, 2021). Enguany tot just iniciem el reclutament de pacients amb afàsia, mentre que en paral·lel aprofundim en els models teòrics sobre la semàntica, tot un repte encara des de la lingüística, precisament per explicar la diferenciació tradicional entre *paraules funcionals* —paraules molt freqüents en les llengües, fonamentals per a la sintaxi (articles, preposicions...)— i *paraules de contingut*, aquelles amb menys rellevància sintàctica, però que tendeixen a aportar més càrrega de significat en els corpus, noms i verbs principalment.

Hi ha filòlegs i lingüistes que encara creuen que això de l'estadística no permet parlar de «lleis lingüístiques», que hi ha massa variabilitat en les produccions lingüístiques, creatives, de les persones o les màquines per fer cap inferència científica. Però recordem el que dèiem a l'inici, quan fèiem el paral·lelisme de la física estadística: necessitem quantitats importants de dades, anàlogament a quan s'estudien sistemes físics amb moltes partícules. És cert, no obstant això, que la realitat dels parlants és molt diversa; per exemple, quina és la pronúncia «bona» del manlleu *podcaster*?³ De «bones», ho són totes, si ens entenem; ara bé, la pronúncia normativa és una altra cosa. Trobarem variacions importants entre parlants,⁴ moltes considerades «no normatives». Quan s'estudien les llengües empíricament, i més en l'oralitat, cal diferenciar la realitat física, estocàstica, de la norma.

I si algú tenia dubtes del potencial de l'estudi estadístic del llenguatge, què hi ha de la irrupció dels models generatius de llenguatge basats en (IA)?

Lingüística quantitativa i IA?

Un grup d'experts de la Comissió Europea va definir la IA de la manera següent (AI-HLEG, 2019: 6, la traducció i les cursives són nostres):

Els sistemes d'intel·ligència artificial (IA) són sistemes de programari (i possiblement també de maquinari) dissenyats per humans que, atès un objectiu complex, actuen en la dimensió física o digital percebent el seu entorn mitjançant l'adquisició de dades, interpretant les dades recollides —ja siguin estructurades o no—, raonant sobre el coneixement o processant la informació derivada d'aquestes dades, i decidint la millor acció (o accions) que cal realitzar per assolir un objectiu donat. Els sistemes d'IA poden utilitzar regles simbòliques o aprendre models numèrics, i també poden adaptar el seu comportament analitzant com es veu afectat l'entorn per les seves accions prèvies.

Com a disciplina científica, la IA inclou enfocaments i tècniques diferents, com ara l'*aprenentatge automàtic* (del qual l'*aprenentatge profund* i l'*aprenentatge per reforç* són exemples específics), el *raonament automàtic* (que inclou la planificació, la programació, la representació del coneixement i el raonament, la cerca i l'optimització), i la robòtica (que inclou el control, la percepció, els sensors i els actuadors, així com la integració de totes les altres tècniques en sistemes ciberfísics).

Per començar, de bon principi la definició anterior exclou els sistemes d'IA dissenyats per màquines, en un antropocentrisme, si més no, rar en aquest context: una IA no pot dissenyar d'altres IA? Entre línies hi ha encara, no obstant això, la idea d'una IA «feble», resolutiva en tasques específiques separades i concretes (jugar a escacs o a go, programar, generar imatges o text...), tot i que allunyada del que implica una intel·ligència general com la biològica, una IA «forta» similar a la humana, objectiu últim de la recerca en IA (López de Mántaras i Meseguer, 2017).

Al segon paràgraf de la definició, d'altra banda, a més de la robòtica, es mencionen termes habitualment emprats a l'acadèmia en anglès, corresponents a tècniques pròpies dels sistemes d'IA, i dels quals fem el calc:⁵ *machine learning* (aprenentatge automàtic), *deep learning* (aprenentatge profund) i *reinforcement learning* (aprenentatge per reforç). L'augment recent d'aplicacions basades en IA ha estat exponencial,⁶ i l'impacte d'algunes d'aquestes, especialment en l'educació, encara s'està paint.

Seguirem ara amb manlleus i els calcs de l'anglès corresponents. Els avenços significatius en aprenentatge automàtic i profund, així com en l'estudi de les xarxes neuronals, han implicat un salt qualitatiu en

les aplicacions lingüístiques basades en IA; la irrupció primer de l'arquitectura del transformador (*transformer*) per part dels equips de recerca de Google, i la posterior evolució dels transformadors generatius pre-entrenats (*generative pre-trained transformer*, GPT) —feta per l'equip de recerca d'OpenAI, que a hores d'ara va per la versió GPT-4—, entre d'altres passos notables, ha suposat un canvi disruptiu en aquestes tecnologies. Gràcies a l'entrenament dels algorismes amb una gran quantitat de dades, els anomenats *grans models de llenguatge* (*large language models*, LLM) són models estadístics basats en probabilitats entre elements lingüístics que no només poden processar el llenguatge natural en temps real, sinó que a més poden produir-lo, generar textos o parla, traduir a múltiples llengües humanes i a llenguatges informàtics (la qual cosa implica programar), interpretar i resoldre enunciats de problemes científics o —gràcies a una breu sol·licitud (*prompting*, vegeu l'apartat següent)— transformar les nostres paraules en text, imatges o vídeos. Les oportunitats, els reptes i els riscos de tot plegat just es comencen a plantejar (Bommasani *et al.*, 2022), potser massa tard?

Aquests sistemes lingüístics basats en IA es fonamenten en potència computacional, en càlculs de probabilitats entre elements i en l'accés i l'entrenament amb grans quantitats de dades, amb una consegüent elevada despesa energètica. La seva efectivitat empírica és òbvia, tot i els problemes de fiabilitat o incoherència que poden tenir en les produccions lingüístiques, que d'altra banda també tenim els humans. I en aquests moments hi ha un (encès) debat sobre si d'aquestes tecnologies basades en IA es podran extreure teories i models lingüístics aplicables als sistemes biològics o no.

Tant l'aprenentatge automàtic com l'aprenentatge profund consisteixen a optimitzar funcions de cost (anomenades en aquest context *loss functions*), de manera que l'aprenentatge automàtic i els LLM impliquen minimitzar, explícitament o implícitament, funcions de cost. Com hem vist, la lingüística quantitativa cerca els principis que hi ha darrere de les lleis lingüístiques naturals, que també optimitzen el cost i que normalment requereixen moltes menys dades per a ser observades. Per tant, simplificant molt, la relació principal entre els models d'aprenentatge automàtic i la lingüística quantitativa és l'optimització del cost. No és sorprenent que els LLM aconseguixin imitar el llenguatge humà tan fidelment, atès que les llengües també es configuren seguint principis d'optimització. La lingüística quantitativa pot ajudar a reduir la complexitat dels LLM, i, en conseqüència, el temps, l'energia i l'impacte ambiental que cal per a entrenar-los. Que potser algú pensarà que no cal conèixer les lleis de la natura si podem fer prediccions computacionals?

Estem disposats a soterrar el coneixement humanístic en la caixa negra dels LLM i els gegants tecnològics

ques que els desenvolupen, o anem més enllà obrint el coneixement profund als mateixos humans, és a dir, en pro de la ciència i la saviesa? Vivim, sens dubte, un moment apassionant, de trànsit, en la recerca lingüística. Caldrà no perdre pas el fil de la indagació, de la investigació en un sentit pur (*in vestigium*).

Prompting i xatbots: les paraules creadores del gólem

Les paraules s'han tornat més creadores i creatives que mai amb la IA. De la creació literària, filosòfica o científica, les paraules han fet el salt a la creació visual, de vídeos i imatges, i, amb una impressora en 3D, ara fins i tot es poden construir escultures i volums, partint dels mots! Segons el *Cambridge dictionary*, *prompting* és «the act of trying to make someone say something»,⁷ és a dir, «l'acte d'intentar que algú digui quelcom». Com a verb (*prompt*) es pot traduir com a «induir, persuadir, animar o provocar a algú a fer alguna cosa»; com a nom és una sol·licitud o un requeriment, si seguim el diccionari d'Oxford.⁸ Però no és només això. Demanem paraules a un humà, però —i això ho hauran de considerar també els lexicògrafs britànics— ara les exigim a una computadora, a un programa que encara no podem dir que sigui «algú». Induïm o persuadim la màquina o simplement estem donant una ordre, la programem?

El substantiu *prompt* és també un símbol informàtic que als que tenim una edat ens transporta a les pantalles verd monocrom. El verb *prompt* deixa molt de marge: màquina, fes que passi alguna cosa amb aquestes instruccions, inspira't amb aquestes paraules. El *prompting* és tan rellevant en la nova era de la IA que el seu disseny i la seva generació —en realitat, les instruccions, les ordres que donem a la màquina que obeeix— ja formen part de tota una saga de programes, els *prompt builders*:⁹ són els constructors de *prompts*, els creadors de la provocació, de la persuasió, de les paraules que són les noves guspises creadores de les IA generatives.

Amb els xatbots tot és més fàcil. O no. De nou, des del *Cambridge* ens diuen que un *chatbot* és un programa informàtic dissenyat per mantenir una conversa amb un ésser humà.¹⁰ No tenen sentit les màquines parlant entre elles o amb animals? Potser ens podem estalviar indicar a la definició el tipus d'interlocutor. Per la part que ens toca, afegim el sufix *-bot* a *xat*, obtenim un mot que, com *robot*, ja tenim al *Diccionari de la llengua catalana* de l'IEC, i llestos. Entrada fàcil al diccionari i neologisme de l'any 2023? Un *xatbot* és un programa informàtic dissenyat per mantenir una conversa. Ho tenim. O no?

Potser no. Ara resulta que xatbots com el ChatGPT no només xerren amb nosaltres, sinó que sembla que re-

solen problemes matemàtics o científics si els donem l'enunciat, fan programes informàtics, anàlisi sintàctica, o escriuen sobre tot allò que demanem, sempre amb el corresponent *prompt*, ja sigui literatura, receptes de cuina o reflexions filosòfiques, impostades i simulades, però sovint per sobre de la mitjana d'altres interlocutors humans, depenent del tema. El nostre coneixement ens situa per sobre o per sota de la tecnologia. Ells són el resultat de la seva arquitectura computacional i de l'entrenament que han rebut. Poden fer, doncs, moltes tasques: escriure, programar, calcular..., i bona part dels deures i les activitats que els docents posen als seus alumnes, i dels informes burocràtics que els superiors demanen als subordinats. Aquestes IA generatives, gòlems creadors obedients, tanmateix, poden esdevenir també eines destructores que canviïn moltes tasques i professions, començant per l'educació.¹¹

Tecnoètica en els usos de la IA

Ja ens havien avisat del perill d'aquests sistemes lingüístics massa grans, dels anomenats «lloros estocàstics» (Bender *et al.*, 2021) que xampurregen i concatenen paraules però, sense considerar la veracitat del que diuen, sense cap consciència de les conseqüències socials de les seves emissions (Bommasani *et al.*, 2022). Però les aplicacions basades en la IA han entrat sense control a la societat, a les aules i a moltes professions en els darrers mesos, fomentades pels interessos econòmics de les empreses tecnològiques que hi ha al darrere, ignorant els principis de la Declaració de Barcelona (2017) per a un ús adequat de la IA:¹² prudència, fiabilitat, rendiment de comptes, responsabilitat, autonomia limitada i rol humà. Ignorem la procedència de les dades que ha rebut la IA i el destí de les dades que li donem? I què hi ha del seu impacte social i mediambiental?

El físic i filòsof Mario Bunge, en una conferència a Haifa el 1974, va encunyar el terme *tecnoètica*, una disciplina llavors emergent, com la part de l'ètica que

analitza les problemàtiques inherents de la tecnologia en les societats humanes i reflexiona sobre aquesta qüestió (Bunge, 2019). Bunge apuntà que la tecnologia dona poder sobre les coses i els éssers humans, i que no tot el poder és bo per a tothom. A més, Bunge traslladà als tècnics la responsabilitat moral dels seus actes professionals, ja que aquests, lluny de ser espontanis, són el resultat de decisions deliberades i racionals que segueixen algun codi moral, o bé en prescindeixen (Bunge, 2019, p. 133): «Els instruments són moralment inerts i socialment irresponsables. Per tant, quan actuen com a instruments, el científic, l'enginyer o l'administrador, refusen assumir responsabilitats [...]».

Tecnologia i llenguatge són dos grans trets inherents a l'ésser humà. Però potser no era fàcil preveure l'impacte tecnològic global del desenvolupament dels sistemes lingüístics computacionals, especialment d'aquells basats en IA. Les conseqüències a mitjà termini, així doncs, són difícils de predir, però com en tota tecnologia disruptiva hi haurà ineludiblement repercussions econòmiques, socials i de canvi (atur?) en moltes professions. Algú assumirà responsabilitats, mentre grans empreses s'enriqueixen i una part de la població s'empobreix, o actuarem com a instruments?

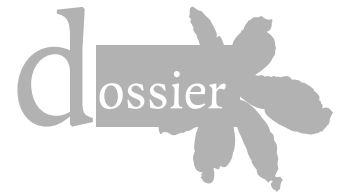
Entretant, comprendre el llenguatge continua sent el repte de la lingüística quantitativa, un objectiu compartit amb d'altres ciències i disciplines afins (Hernández-Fernández i Ferrer-i-Cancho, 2019), un repte convergent en què el ser i el saber estan per sobre de l'aparença, de la impostació tecnològica. Caldrà explicar els paral·lelismes que hi ha entre la comunicació dels sistemes biològics i la que generem amb els artefactes amb els quals imitem la natura. Des de la lingüística quantitativa s'establiran de manera natural ponts entre les ciències i les tecnologies lingüístiques. Però tot això s'hauria de fer sense descuidar els aspectes ètics i les conseqüències socials que tota recerca comporta. A què volem dedicar els esforços, a l'entrenament sofisticat o a la ciència? *

Bibliografia

- AI-HLEG (2019). «High-level expert group on artificial intelligence». A: Ethics guidelines for trustworthy AI [en línia], 6. <<https://eskillsalliancecms.gov.mt/en/news/Documents/2019/AIDefinition.pdf>> [Consulta: 21 de juny de 2023].
- BENDER, Emily; GEBRU, Timnit; MCMILLAN-MAJOR, Angelina; SHMITCHELL, Shmargaret (2021). «On the dangers of stochastic parrots: can language models be too big?». *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, p. 610-623.
- BEST, Karl-Heinz; ROTTMANN, Otto (2017). *Quantitative linguistics, an invitation*. Lüdenscheid: RAM Verlag.
- BOMMASANI, Rishi [et al.] (2022). «On the opportunities and risks of foundation models». [en línia]. <<https://arxiv.org/abs/2108.07258>> [Consulta: 21 de juny de 2023].
- BUNGE, Mario (2019). *Filosofia de la tecnología*. Barcelona: IEC: Edicions UPC.
- CATALÀ, Neus; BAIXERIES, Jaume; FERRER-I-CANCHO, Ramon; PADRÓ, Lluís; HERNÁNDEZ-FERNÁNDEZ, Antoni (2021). «Zipf's laws of meaning in Catalan». *Plos ONE*, 16 (12), p. e0260849.
- DEBOWSKI, Lukasz (2020). *Information theory meets power laws: Stochastic processes and language models*. Hoboken, NJ: John Wiley & Sons.
- FERRER-I-CANCHO, Ramon; HERNÁNDEZ-FERNÁNDEZ, Antoni; LUSSEAU, David; AGORAMOORTHY, Govindasami; HSU, Minna; SEMPLE, Stuart (2013). «Compression as a universal principle of animal behavior». *Cognitive Science*, 37 (8), p. 1565-1578.
- GÓMEZ-RODRÍGUEZ, Carlos; FERRER-I-CANCHO, Ramon (2017). «Scarcity of crossing dependencies: A direct outcome of a specific constraint?». *Physical Review E*, 96 (6), p. 062304.
- HERNÁNDEZ FERNÁNDEZ, Antoni; FERRER I CANCHO, Ramon (2019). *Lingüística quantitativa: La estadística de las palabras*. Barcelona: EMSE Ediciones: Prisanoticias.
- HERNÁNDEZ-FERNÁNDEZ, Antoni; GARRIDO, Juan Manuel; LUQUE, Bartolo; TORRE, Iván G. (2022). «Linguistic laws in Catalan». *Quantitative Approaches to Universality and Individuality in Language*, 75, p. 49-62.
- HERNÁNDEZ-FERNÁNDEZ, Antoni; TORRE, Iván G.; GARRIDO, Juan-María; LACASA, Lucas (2019). «Linguistic laws in speech: The case of Catalan and Spanish», *Entropy* [en línia], 21 (12), p. 1153. <<https://doi.org/10.3390/e21121153>> [Consulta: 21 de juny de 2023].
- KÖHLER, Reinhard; ALTMANN, Gabriel; PIOTROWSKI, Rajmund G. (ed.) (2008). *Quantitative Linguistik / Quantitative linguistics: Ein internationales Handbuch / An international handbook*. Berlín: Walter de Gruyter.
- LÓPEZ DE MÁNTARAS, Ramon; MESEGUER, Pedro (2017). *Inteligencia artificial*. Madrid: CSIC: Libros de la Catarata. (¿Qué Sabemos de?)
- PASTORIZA-DOMÍNGUEZ, Patricia; TORRE, Iván G.; DIÉGUEZ-VIDE, Faustino; GOMEZ-RUIZ, Isabel; GELADÓ, Sandra; BELLO-LÓPEZ, Joan; ÁVILA-RIVERA, Asunción; MATÍAS-GUIU, Jordi A.; PYTEL, Vanesa; HERNÁNDEZ-FERNÁNDEZ, Antoni (2022). «Speech pause distribution as an early marker for Alzheimer's disease». *Speech Communication* [en línia], 136, p. 107-117. <<https://doi.org/10.1016/j.specom.2021.11.009>> [Consulta: 21 de juny de 2023].
- SEMPLE, Stuart; FERRER-I-CANCHO, Ramon; GUSTISON, Morgan L. (2022). «Linguistic laws in biology». *Trends in Ecology & Evolution*, 37 (1), p. 53-66.
- TORRE, Iván G.; LUQUE, Bartolo; LACASA, Lucas; KELLO, Christopher T.; HERNÁNDEZ-FERNÁNDEZ, Antoni (2019). «On the physical origin of linguistic laws and lognormality in speech». *Royal Society Open Science* [en línia], 6, p. 191023. <<https://doi.org/10.1098/rsos.191023>> [Consulta: 21 de juny de 2023].
- ZIPF, George Kingsley (1949). *Human behaviour and the principle of least effort: An introduction to human ecology*. Cambridge, MA: Addison-Wesley.

Notes:

1. Síntesi, resultats i publicacions disponibles a <https://www.iec.cat/recerca/projecte1.asp?codi=PRO2018-So3-HERNANDEZ>.
2. Síntesi, resultats i publicacions disponibles a <https://www.iec.cat/recerca/projecte1.asp?codi=PRO2022-So3-HERNANDEZ>.
3. Vegeu <https://www.termcat.cat/ca/divulgacio/la-bona-pronuncia-dels-podcasters-i-les-podcasteres>.
4. Així ho mostra el projecte Corpus Oral de la Llengua Catalana (COLC) de l'Institut d'Estudis Catalans (IEC), que té com a objectiu la recollida, el tractament i l'anàlisi de mostres orals significatives de totes les varietats i els dialectes majors de la llengua catalana. Per a més detalls, consulteu <https://www.iec.cat/recerca/projecte1.asp?codi=PR2017-So4-DOLS>.
5. Vegeu <https://www.termcat.cat/ca/divulgacio/com-aprenen-les-maquines-alguns-termes-la-intelligencia-artificial-recercat>.
6. Vegeu, per exemple, el cercador d'aplicacions basades en IA <https://aifindy.com/>.



7. Vegeu, al *Cambridge dictionary*, les entrades de *prompting* (<https://dictionary.cambridge.org/dictionary/english/prompting>) i *prompt* (<https://dictionary.cambridge.org/dictionary/english/prompt>).
8. *Oxford learner's dictionaries* ens defineix *prompting* com «an act of persuading somebody to do something» (<https://www.oxfordlearnersdictionaries.com/us/definition/english/prompting>).
9. Vegeu, per exemple, <https://promptomania.com/>.
10. Vegeu <https://dictionary.cambridge.org/dictionary/english/chatbot>.
11. Vegeu <https://aihub.csic.es/inteligencia-artificial-en-educacion-golem-creativo-o-destructor/>.
12. Vegeu <https://www.iiia.csic.es/barcelonadeclaration/>.